# Two-Sided Cultural Fit: The Differing Behavioral Consequences of Cultural Congruence Based on Values Versus Perceptions

Richard Lu
University of California, Berkeley

Jennifer A. Chatman
University of California, Berkeley

Amir Goldberg
Stanford University

Sameer B. Srivastava
University of California, Berkeley

Forthcoming in *Organization Science*

How do people establish and maintain cultural fit with an organization? Prior research has offered two perspectives that have heretofore been conceptually disconnected: One focuses on personal values, while another emphasizes perceptions of the cultural code. We develop a theoretical account that integrates these approaches by linking them to distinct mechanisms and behavioral consequences of cultural fit. We propose that *value congruence*—the match between one's values and those that prevail in an organization—relates to the mechanism of group attachment and shapes behavior when one periodically steps back from day-to-day interactions, assesses one's identification with an organization, and determines whether to stay or voluntarily depart. In contrast, we argue that *perceptual congruence*—the degree to which one implicitly understands an organization's prevailing values and norms—relates to the mechanism of interpersonal coordination and influences behavior when one engages in routine peer interactions. Accordingly, we theorize that these two forms of cultural fit relate to distinct behaviors, *voluntary exit* and *linguistic conformity* with peers, respectively. Drawing on email and survey data from a mid-sized technology firm, we find support for our theory and discuss implications of our findings for research on person-culture fit, dual-process models of culture and cognition, and the pairing of surveys with digital trace data.

## Introduction

Whether assimilating to a country or adapting to a new school, people typically seek to fit in culturally with their social groups. The benefits of conformity, as well as the sanctions and penalties that come with failed cultural integration, are particularly stark in contemporary organizations. Indeed, prior work has consistently demonstrated that high levels of individual cultural fit are associated with increased productivity, stronger commitment, and less turnover (Kristof-Brown et al. 2005, Chatman and O'Reilly 2016). Moreover, employers are increasingly screening and selecting new hires based on their anticipated cultural fit rather than just their skills (Chatman 1991, Meyer et al. 2010, Rivera 2012). At the same time, as the average tenure in organizations has declined (Hall 1996), workers must frequently retool themselves culturally as they move from one organization to the next. Yet people vary considerably in how well they fit into and adapt to a given organization (Chatman 1989, Srivastava et al. 2018). How do people establish and maintain cultural fit in an organization and what are its behavioral consequences?

Existing research offers two perspectives that have heretofore been conceptually disconnected. The first focuses on values. This line of work, echoing a long tradition in psychology and sociology, sees the locus of culture in the alignment between the values espoused by the individual—typically defined as relatively enduring beliefs about preferred modes of conduct (Rokeach 1973)—and those that prevail in a group. Fitting in therefore implies having personal values that match those of other members. Indeed, a robust literature has demonstrated that the alignment between a person's values and those that are pervasive in a social group predicts a variety of individual and organizational outcomes (Chatman 1991, Edwards 2008).

A second explanation, which is rooted in cultural sociology, largely rejects the notion that personal values affect behavior, positing instead that culture shapes action through perceived situational cues. This approach argues that behaviors are primarily driven by perceptions of cultural scripts that are invoked through interactions with others. An employee's decision to use polite language in a meeting, for example, often reveals little about her underlying preference for civil discourse but instead reflects the values and norms she perceives in the behavior of other meeting participants. Indeed, people pursue action for which their "cultural equipment is well suited" (Swidler 1986, p. 277), suggesting that those who fit in are those whose readings of the group's cultural code lead them to behave in appropriate ways.

These two perspectives emphasize distinct mechanisms of cultural fit. The former suggests that cultural fit is the result of internalizing and embracing prevailing values and norms. Fitting in thus relates especially to the mechanism of group attachment. The latter views cultural fit as the product of correctly deciphering the group's dominant values and norms. Fitting in therefore relates primarily to the mechanism of interpersonal coordination.[1] By conceptually integrating these two approaches, we develop a more complete account of what cultural fit means and how it manifests in the form of distinct organizational behaviors.

We first propose that *value congruence*—the match between one's personal values and those that prevail in an organization (Chatman 1989, Alba and Nee 2009)—involves more deliberative choices about whether and how strongly to attach oneself to the group and thus shapes behavior when one periodically steps back from day-to-day interactions, assesses one's identification with an organization, and determines whether to stay or voluntarily depart. In contrast, because perceptions of multilayered situations in organizational life are difficult to systematize and convey to others, we argue that *perceptual congruence*—the degree to which one implicitly understands an organization's prevailing values and norms—involves more automatic choices that enable smooth interpersonal coordination and influences behavior when one engages in routine interactions with peers. Thus, we anticipate that these two forms of cultural fit will relate to distinct behaviors: Value congruence will be negatively associated with the likelihood of voluntary exit, while perceptual congruence will be positively tied to *linguistic conformity* with peers.

To evaluate these ideas, we employ a multi-method empirical strategy that draws on survey data, eight years of internal email data, and personnel records from a mid-sized technology firm. We use different features of the Organizational Culture Profile (OCP) (Chatman 1991), a validated culture assessment, to separately measure value congruence and perceptual congruence. We measure linguistic conformity by applying the interactional language-use model to a corpus of internal email messages (Srivastava et al. 2018, Goldberg et al. 2016).

We begin by reporting cross-sectional results that are consistent with our hypotheses. Yet, recognizing that cultural fit is likely to play out over time and that prior studies of cultural fit have focused on measures collected only once or a handful of times, we also employ a novel machine learning-based method to impute value congruence and perceptual congruence for individuals over time. Although this method still needs to be validated in other empirical settings in which researchers have access to multiple waves of survey data and can therefore assess the degree to which the relationship between survey responses and communication behavior is stable over time, it nevertheless gives us a preliminary window into within-person changes in value congruence and perceptual congruence and thus allows us to estimate longitudinal models. Results from our longitudinal analyses lend further support for our expectation that value congruence relates to voluntary exit, while perceptual congruence is associated with linguistic conformity. We conclude by discussing the implications of our findings for research on person-culture fit, dual-process models of culture and cognition, and the pairing of surveys with digital trace data.

## Theory and Hypotheses

### Cultural Fit Based on Personal Values Versus Perceptions of Prevailing Values

Values—"enduring beliefs that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite mode of conduct or end-state of existence" (Rokeach 1973, p. 5)—feature prominently in scholarship on organizational culture and the process by which individuals fit into, and are conversely shaped by, their work environments (Lofquist and Dawis 1978, Dawis 2005, Chatman and O'Reilly 2016). Research in this vein has tended to conceptualize individual cultural fit through the prism of *value congruence*: the match between a person's values and those that prevail in her social group. People whose ideal preferences are compatible with those prevalent in their organizational environment exhibit higher subjective well-being and enjoy greater attainment (O'Reilly et al. 1991).

Work that focuses on the value congruence dimension of cultural fit emphasizes group attachment as a core mechanism that links values to individual outcomes in organizations. Individuals whose values are compatible with those prevalent in an organization are more likely to self-identify with that organization (Cable and Judge 1996, Judge and Cable 1997). Such identification, in turn, leads to heightened motivation, stronger commitment, and higher productivity (Chatman 1991, Baron et al. 2001).[2]

The notion that values are fundamental drivers of human behavior has a long history in sociology (Parsons 1968) and psychology (Rokeach 1973, Schwartz 1992, Hofstede 2001). This research demonstrates, for example, that values are associated with cross-national and regional differences in economic growth (Inglehart and Baker 2000) and violence (Nisbett and Cohen 1996), as well as with individual lifestyle (Miles 2015), financial (Keister 2008), and occupational (Alesina et al. 2015) choices. Yet a growing body of research finds that people's stated values are, in many cases, poor predictors of their behavior (Greenwald and Banaji 1995). Economically disadvantaged high school students, for example, tend to express mainstream attitudes on educational achievement and sexual behavior but adopt behaviors that appear to be inconsistent with these ideals (Harding 2007). In organizations, too, people's behaviors are often incongruent with their stated beliefs: Self-reported values on cross-functional collaboration, for example, are largely unrelated to individuals' propensity to build network ties that span functional boundaries (Srivastava and Banaji 2011).

Research in cultural sociology has therefore tended to downplay the role of personal values in shaping behavior. This work often relies on two fundamental and interrelated assumptions. The first is that "people know more culture than they use" (Swidler 1986, p. 277), namely, that they subscribe to multiple, and potentially inconsistent, cultural logics and value systems. Given this multiplicity, the same setting can elicit different interpretations, leading to inconsistent behavioral responses. The second assumption is that people's behavior is situationally driven. Subtle contextual cues in other people's behavior serve as signals about how to interpret a situation and, consequently, what kind of behavior is appropriate. Because these perceived meanings emerge through interaction (Childress and Friedkin 2012, Gibson 2011), value assignment often occurs retroactively (Boltanski and Thévenot 2006).

This constructivist understanding of culture shifts focus from what people value to how they interpret their experiences of the world and produce meaning through interaction. Culture, according to this approach, systematically shapes behavior through what Eliasoph and Lichterman (2003) call "group styles:" idiosyncratic cultural codes that connect symbols, actions, and vocabularies to meaningful categories. Consider, for example, the perennially disgruntled employees in Weeks' (2004) ethnography of a British bank. To an outsider observing people habitually complaining, it may have seemed that these employees were fundamentally rejecting the organization and its culture. As Weeks artfully demonstrates, however, employees were instead partaking in rituals intended at reaffirming their bonds and their commitment to the bank.

Thus, beyond having values that align with those of other organizational members, fitting into an organizational culture also depends on possessing the tacit and layered understanding necessary for deciphering the organization's intricate cultural code. We refer to this ability as *perceptual congruence* and argue that it arises from two underlying processes.[3] The first relates to the person's construal of a situation, by which we mean the mental representation that she conjures when making sense of others' behaviors (DiMaggio and Goldberg 2018). A colleague's cynical joke in a meeting, for example, can be interpreted as a friendly attempt to establish rapport or as a derogatory comment aimed at undercutting others. An observer's capacity to correctly construe the meeting as friendly or adversarial depends on the compatibility between her and others' interpretations of participants' behaviors. Second, the person's reading of the values and norms that are prevalent in the organization shapes what behaviors she deems appropriate in light of her construal. Having an implicit understanding of prevailing values and norms thus eases interpersonal interaction and coordination because it allows people to develop mutually compatible behavioral expectations. For example, recognizing that detail-orientation is a prevailing norm might lead a manager to check in less frequently with subordinates and instead assume that they will deliver high-quality outputs without needing to be prompted. This, in turn, will reduce the likelihood of interpersonal frictions arising from subordinates viewing the manager as "micro-managing."

**Two-Sided Cultural Fit**

To integrate these two approaches—one focused on the compatibility of personal values with those of group members and the other on the accuracy of perceptions about group culture—we draw on the insight that different facets of cultural understanding about an organization are associated with different kinds of thinking and associated action. The aspects of cultural understanding that are readily distilled and articulated involve more deliberative and conscious choices, whereas those that are more nuanced and difficult to express in simple terms shape habitual and less reflective behavior (Vaisey 2009, Srivastava and Banaji 2011, Lizardo 2017, Leschziner 2019). Recognizing that different processes are involved in habitual versus reflective behavior requires rethinking cultural fit as a two-sided construct rather than one that is determined solely by either personal values (e.g., Chatman 1991) or perceptions of culture (e.g., Swidler 1986, Srivastava et al. 2018). Accordingly, we argue that personal values are more consequential for reflective decisions that are less influenced by normative cues given off by others. In contrast, perceptions matter more for routine behavior that is guided in part by observing others and inferring what is appropriate.

Most activities in organizations occur routinely, in settings that provide high situational clarity given people's familiarity with the setting and the availability of habituated behavioral responses within it (Davis-Blake and Pfeffer 1989, Sørensen 2002, Michel 2011). We therefore posit that perceptual congruence will be consequential for an individuals' ability to exhibit culturally compliant behavior. To productively participate in ritualistic complaining, for example, the employees in Weeks' (2004) ethnography of *BritArm Bank* had to complain at the appropriate level: not too much so as to avoid rocking the boat but enough to signal membership and belonging with the group.

We refer to the linguistic expressions of such compliance with normative expectations as *linguistic conformity*.[4] We focus on linguistic conformity as a consequential outcome given prior work showing that, when people communicate with group members in ways that match the group's normative expectations, it enables all parties to come to a common understanding of the opportunities and challenges they collectively face and facilitates their ability to coordinate their responses—behaviors that are fundamental to group effectiveness (Lazear 1999, Lewis 2008, Weber and Camerer 2003, Cremer et al. 2007). Of course, perceptual congruence is likely related to other behavioral outcomes, from dressing appropriately to correctly reading expectations about after-hours work (Rafaeli and Pratt 1993). Language is, however, the most central medium through which employees coordinate their activities in contemporary organizations. Thus, linguistic conformity represents one of the most significant manifestations of an individual's grasp of the group's cultural code (Lazear 1999).

We further argue that value congruence will, in contrast, be less consequential for a person's capacity to conform to her group's routine normative expectations. Although people whose values are more congruent with their organization's may be motivated to behave in normatively compliant ways (e.g., Chatman 1991), they may still lack the implicit understanding of prevailing norms needed to do so. It is one thing to prefer, for example, a cooperative work environment and another to understand which behaviors signal cooperativeness in a specific cultural context (Chatman and Barsade 1995). Moreover, values are held at the individual level and are often private (Rokeach 1973). Thus, we would not expect them to play a significant role in behavior that is more collective and public in nature.

Consider, for example, a person whose personal values are about behaving ethically but who works in an organization where the dominant values and norms are about ignoring ethical guidelines. Now imagine two behaviors: whether to cheat on one's own expense statement and whether to improperly recognize revenue for a product line one oversees. The former is a mostly private behavior: It will only be known to others if one is caught. The latter is more public in that the relevant parties are more likely to know that revenue that should have been recognized for the next quarter is being inappropriately counted toward the current quarter's goals. We would expect the individual in question to eschew cheating on her personal expense statement because doing so would be consistent with her personal values and not reveal to peers the deviation from prevailing values and norms. Conversely, assuming she correctly perceives the organization's dominant values and norms, we would expect her to be more likely to engage in improper revenue recognition even if such behavior violates her personal values.

Building on this logic, we expect that value congruence will be consequential for group attachment, predicting behavior when people periodically assess their place in an organization and contemplate whether they want to stay or instead exit (e.g., O'Reilly and Chatman 1986). Indeed, organizational researchers have long considered voluntary exit, the choice to remain in the group or voluntarily depart, as perhaps the most tangible and visible way in which one expresses one's attachment (or lack thereof) to a group (e.g., Michaels and Spector 1982, Harman et al. 2007). When people make such decisions, they respond less to what types of appropriate behaviors the situation activates and more to their beliefs about what is desirable. Moreover, such reflection often occurs in private contexts in which colleagues' behavioral cues and normative expectations are not on display and thus less salient. Although the announcement of departure from an organization is public, the deliberations that lead to this decision typically occur privately or not at all given that people seek to keep their options open until they have reached a clear conclusion (Lee 2014).

Just as we do not anticipate value congruence to matter for linguistic conformity with peers, so we argue that perceptual congruence will not factor significantly in people's choices to voluntarily exit an organization. One may, for example, be cognizant of the values and norms that are widespread in an organization and have the capacity to behave in normatively compliant ways but still not feel motivated to continue doing so if the prevailing values are at odds with one's personal values.[5] Together, these arguments lead us to formulate the following two hypotheses:

***Hypothesis 1.*** *Perceptual congruence is positively related to linguistic conformity in routine interactions.*

***Hypothesis 2.*** *Value congruence is negatively related to voluntary exit from the organization.*

# Method

## Overview

Previous work on cultural fit in organizations has, by and large, relied exclusively on self-reports to assess both cultural and behavioral variables. This approach has two major limitations (Gerald and George 2010). First, self-reports predominantly elicit deliberative cognition (e.g., subjective well-being or retroactive behavioral accounts). They are less well-suited to detecting behavior that arises more automatic modes of cognition. Second, it is usually impractical or too costly to collect self-reports on a frequent basis. Consequently, they are not well-suited to measuring subtle changes on a granular timescale.

To address these limitations, we employ a multi-method approach that draws on both survey and email communication data. We begin by testing our hypotheses using cross-sectional data. We then use a machine learning technique to impute time-varying measures from cross-sectional data and estimate longitudinal models with individual fixed effects that account for time-invariant unobserved heterogeneity.

## Data

Our empirical setting is a mid-sized technology firm that broadly operated in the energy sector. During the observation period, the firm experienced periods of rapid growth, as well as some industry-level shocks that necessitated budget cuts and layoffs. We collected from this firm three types of data:

**Personnel Records.** We received monthly extracts from the firm's human resource information system. These extracts included demographic information such as age and gender, organizational status such as departmental affiliation and start date, and information about individual outcomes such as monthly bonus received.

**Email Data.** We collected eight years of email data from the organization, including not only metadata (i.e., who sent messages to whom and when) but also raw message content. Given our focus on cultural dynamics within the organization, we excluded emails exchanged between employees and the outside world. We also eliminated automatically generated messages and, per instructions from the company's in-house lawyers, messages sent from or to members of the (small) legal department. The resulting data set included over five million unique emails.

**Organizational Culture Profile.** All employees were invited to complete an Organizational Culture Profile (OCP) (Chatman et al. 2014) assessment about the organization's current culture. We also asked a randomly selected half of employees to complete the assessment based on their own personally desired cultural characteristics.[6] As described below, our measure of perceptual congruence is based only on the assessment about the current culture, which 440 individuals completed. Our measure of value congruence entails a comparison of others' reports about the current culture with an individual's own preferences. Value congruence is therefore defined for the 238 people who completed the assessment about their personally desired culture. Once we matched the raw email data to personnel records and removed identifying information, the resulting data set consisted of 29,255 person-month observations, spanning the period from 2008 to 2016.[7]

**Dependent Variables**

**Linguistic Conformity.** Hypothesis 1 predicts a positive relationship between perceptual congruence and linguistic conformity. We operationalized linguistic conformity as the similarity between an individual's language and her reference group's, using the Interactional Language Use Model (ILUM) (Goldberg et al. 2016, Srivastava et al. 2018). Although language is not the only means through which culture is enacted—for example, culture also manifests in dress and various forms of nonverbal communication—it is a dominant medium through which cultural information is exchanged (Lazear 1999). Given that linguistic similarity can sometimes reflect alignment for non-cultural reasons—for example, two people coordinating on a shared task might use similar language even when they are culturally incompatible—we focus on the similarity of linguistic style between an individual and her reference group. Drawing on previous sociological work on culture (Bail et al. 2017, Doyle et al. 2017), ILUM uses the well-established and widely used Linguistic Inquiry and Word Count (LIWC) lexicon (Pennebaker et al. 2007) to measure linguistic style. LIWC is a semantic dictionary that maps words into 64 high-level emotional, cognitive, and structural categories. A comprehensive body of work demonstrates that the linguistic units identified by LIWC relate to a wide and universal array of meaningful psychological categories (Tausczik and Pennebaker 2010).

Using LIWC allows us to focus on expressions that are inherently cultural, while downplaying linguistic exchange that is organization- or context-specific or primarily related to functional coordination between organizational members.[8] Imagine, for example, an organization with an aggressive and competitive culture. Such a culture might manifest linguistically in expressions of certainty, negation, and the use of swear words and other forms of non-deferential language. Contrast such a normative environment with one characterized by politeness and the use of tentative and inclusive language, indicating a collaborative and non-confrontational culture. LIWC is specifically designed to capture such culturally meaningful dimensions.

To derive our measure of linguistic conformity, we first translated raw emails into LIWC category counts. We then aggregated each individual's incoming and outgoing emails into monthly time periods and represented each person-month observation as two probability distributions of outgoing and incoming communication over LIWC categories. We used the Jensen-Shannon divergence metric (inverse and log-transformed) between these two probability distributions as the measure of linguistic conformity.

Intuitively, when the outgoing and incoming distributions are nearly identical, the divergence approaches zero, suggesting high linguistic conformity; conversely, greater deviation between the probabilities of usage of LIWC categories translates to greater divergence and thus implies lower linguistic conformity. Thus, the more an employee's use of cognitive, emotional, and structural terms in sent emails matches the use of those terms in received emails, the greater her linguistic conformity in a given month.

We discuss the technical details of this measure in Appendix A, which also includes a table with the LIWC categories and reports the results of two validation checks. The first compares LIWC and OCP categories to demonstrate that our language-based measure reflects culturally meaningful content. The second reports the results of a simulation analysis, which reveals that our measure is robust to the exclusion of arbitrary sets of LIWC categories. In other words, even if we assume that given sets of LIWC categories are culturally meaningless, their exclusion would have a negligible effect on the resulting measure.

**Voluntary Exit.** In Hypothesis 2, we predict that value congruence will be negatively related to a person's chances of departing voluntarily. We identified voluntary exit based an employee's departure date. We used company records to distinguish between voluntary and involuntary exit.[9] (We do not theorize about the relationship between value congruence or perceptual congruence and involuntary exit because involuntary exit is often influenced by circumstances beyond the focal person's control—for example, a financial downturn that triggers a round of layoffs.)

**Work Performance.** To help validate our measures of value congruence, perceptual congruence, and linguistic conformity, we report below results of models in which we examine their relationship to individual work performance. We used monthly bonus payments as the measure of individual work performance. For people in job roles such as sales or operations in which productivity could be objectively assessed, the company established a formula that linked specific productivity indicators—for example, a sales person's conversion of leads into revenue—to monthly bonus payments. Given that the distribution of bonuses was skewed, we logged this measure in the analyses reported below.

**Independent Variables**

**Perceptual Congruence.** We used two facets of the OCP to derive our measure of perceptual congruence. The OCP consists of 54 value statements (e.g., fast moving, being precise) that emerged from a review of academic and practitioner-oriented writings on culture (O'Reilly et al. 1991). Using the Q-sort methodology (Block 1961), respondents are asked to array these 54 statements into nine categories, with a specified number of statements in each category. The required distribution of statements across categories is 2-4-6-9-12-9-6-4-2, so that, for example, respondents rating the current culture of their organization would place two value statements each in the "most characteristic" and "most uncharacteristic" categories, respectively, four value statements each in the "quite characteristic" and "quite uncharacteristic" categories respectively, and 6 statements each in the "fairly characteristic" and "fairly uncharacteristic" categories respectively, and so on, until all 54 value statements were categorized. Unlike a Likert-format scoring scheme in which many or all items can be rated as high or low, or a ranking process, which, with 54 value statements to rank, would be unwieldy for human raters, this semi-idiographic approach forces respondents to choose cultural value statements that are most and least characteristic of their organization.

To derive our measure of perceptual congruence, we focused on an OCP question that was asked of all respondents: "To what extent do the value statements characterize the organization as a whole?" We defined perceptual congruence as the match between an individual's current culture profile and those of a reference group of peers. To make this measure comparable to our measure of linguistic conformity, we chose the same reference group—that is, the set of colleagues a person had email contact with in a given month weighted by communication volume.

**Value Congruence.** For value congruence, we focused on two different facets of the OCP. First, we considered participants' responses to the question: "To what extent do the value statements characterize your personally desired values, that is, the values you desire in an organization?" Next, we defined value congruence as the correspondence between an individual's personal culture profile (the values she herself prefers) and the reference group's current culture profile (the values that group members report as being dominant in the organization). For consistency, we chose the same reference group for value congruence as we did for perceptual congruence and linguistic conformity.

### Imputing Perceptual Congruence and Value Congruence Over Time

The procedure above yields cross-sectional measures of perceptual congruence and value congruence. Models based on such measures cannot account for time-invariant, unobserved heterogeneity—for example, stable personality traits and dispositions that might be related to our outcomes of interest.

We therefore undertook a procedure to transform our cross-sectional measures of value congruence and perceptual congruence into longitudinal measures. Taking inspiration from Salganik's (2017) notion of *amplified asking*—that is, combining surveys with digital trace data to infer responses for people who cannot be feasibly surveyed or whose responses are missing—we undertook a procedure based on machine learning techniques to identify from raw email content (rather than the higher-level LIWC categories used to derive our measure of linguistic conformity) the "linguistic signature" of perceptual congruence and value congruence.

We assumed that, if language reflects internal processes of cognition (Pinker 2007), then there should be an identifiable relationship between email communication and the two dimensions of perceptual congruence and value congruence. Specifically, we used a random forest model to help uncover this underlying link between language and cognition (Ho 1995, Friedman et al. 2001). Random forest models have several beneficial characteristics for this task: they can detect arbitrary, nonlinear relationships; they typically require fewer observations than do other machine learning methods to produce comparable results; and they are inherently robust to overfitting, or incorrectly inferring signal from idiosyncratic noise in the data. Figure 1 provides a conceptual overview of this procedure. Further procedural details are provided in the Appendix B.

Because we only had access to a single administration of the OCP assessment, we were unable to validate a core assumption underpinning this imputation approach: that the relationship between survey responses and communication behavior remains relatively stable over time. Analyses based on our imputed measures should therefore be interpreted with some caution—pending validation of this assumption in future research. But even without this validation, we see value in reporting results based on imputed measures for two reasons. First, it provides a demonstration of the technique, which we hope will spur further work that brings together survey-based measures with measures based on digital trace data. Second, it allows us to estimate models with both individual and period fixed effects, thereby controlling for a host of time-invariant, unobserved factors that could potentially be related to our outcomes of interest.

### Control Variables

We estimated both within-person and between-person models for our analyses. In within-person models, which were based on our imputed measures of value congruence and perceptual congruence, we included two time-varying controls that prior research suggests are relevant to the study of cultural conformity. First, we included (lagged) managerial status since employees may be more likely to accommodate the behaviors, and specifically the language use, of interlocutors who possess greater structural power (Mayer et al. 2009). Next, we included departmental dummies since departments vary in relative centrality and power, which may in turn influence the degree to which their members are motivated to conform to behavioral norms (Thompson 1967, Salancik and Pfeffer 1974).[10] For our between-person models, we included additional control variables for age, age-squared, and gender.

[FIGURE 1 ABOUT HERE]

### Analytical Approach

We tested Hypothesis 1, which suggests that perceptual congruence will be positively related to linguistic conformity, by estimating OLS regressions on cross-sectional data, as well as fixed effect regressions based on longitudinal data (including imputed measures of perceptual congruence and value congruence). In our longitudinal models, we also included year fixed effects to account for unobserved differences across time such as stages in the company's evolution (e.g., a start-up with relatively loose structures and processes versus a more established firm with relatively tight structures and processes) or the quality of outside opportunities, which could have in turn influenced people's motivations to fit in. Hypothesis 2, which predicts that value congruence will be negatively related to voluntary exit, was estimated using Cox proportional hazard models. Because it is not possible to include period fixed effects in Cox proportional hazard models, we included a control for firm size (based on the number of employees) to account for different stages in the company's evolution. We standardized all variables in the regression models reported below. We use lagged predictors in longitudinal models to address (though not fully resolve) reverse causality.

## Results

### Main Results

Correlations between the main variables of interest are shown in Table 1. Linguistic conformity is positively and significantly related to perceptual congruence and manager status but not to value congruence. Perceptual congruence and value congruence are positively related to each other—although this relationship is modest and only marginally significant. Perceptual congruence is also positively and significantly related to manager status. This finding is consistent with prior work showing a positive relationship between organizational rank and the accuracy of perceptions about social interactions in the workplace (Casciaro 1998). In contrast, value congruence is not significantly associated with manager status. Perhaps reflecting a culture that is less hospitable for women in many technology firms (Cheryan et al. 2017), women's value congruence is significantly lower than that of men. There is also a positive link between value congruence and age; however, this relationship is also only marginally significant.

[TABLE 1 ABOUT HERE.]

Table 2 provides a test of Hypothesis 1. Models 1 to 3 report results from cross-sectional data, with linguistic conformity averaged over three months preceding the administration of the OCP. In support of Hypothesis 1, perceptual congruence is significantly related to linguistic conformity, while value congruence is not; moreover, these patterns hold whether the value congruence and perceptual congruence are modeled separately (Models 1 and 2) or jointly (Model 3). As shown in Figure 2, a one standard deviation increase in perceptual congruence is associated with a 0.15 standard deviation increase in linguistic conformity. Of the control variables used in the cross-sectional models, only managerial status is significantly related to linguistic conformity.

[TABLE 2 ABOUT HERE.]

[FIGURE 2 ABOUT HERE]

Table 2, Models 4 to 6, echo the results from the cross-sectional analyses in longitudinal specifications that include individual, department, and year fixed effects. The longitudinal results provide further support for Hypothesis 1 given that perceptual congruence is significantly related to linguistic conformity, while value congruence is not. As individuals' perceptual congruence increases, their linguistic conformity correspondingly increases. Changes in value congruence, in contrast, are unrelated to changes in linguistic conformity.[11]

Changes in managerial status are also significantly related to linguistic conformity. We conjecture that managers exhibit greater linguistic conformity than do individual contributors either because their general tendency toward cultural congruity was conducive to their past promotion into management or because subordinates are more likely to linguistically accommodate their communication style.

Table 3 reports tests of Hypothesis 2. Our competing risks Cox hazard models focus on voluntary exit as a function of value congruence and perceptual congruence (with involuntary exit serving as the competing risk).

As Table 3, Model 1, indicates, perceptual congruence is not significantly related to voluntary exit. In contrast, Model 2 shows that value congruence is significantly and negatively related to voluntary exit. Model 3 demonstrates that the significant relationship between value congruence and voluntary exit remains even when perceptual congruence is added as a covariate. A one standard deviation increase in value congruence is associated with a 13% decrease in the rate of voluntary exit. This association, which is based on an imputed longitudinal measure of value congruence, is consistent with prior work based on a cross-sectional measure that predicted departure from firms up to two years later (Chatman 1991). Among the control variables, female and firm size (measured by the number of employees) are positively and significantly related to voluntary exit, while the association between age and voluntary exit has a curvilinear form.[12]

[TABLE 3 ABOUT HERE.]

**Supplemental Analyses—Evaluating the Variables of Interest**

Here we summarize two additional analyses that sought to evaluate the validity of our measures of value congruence and perceptual congruence, including the imputed versions of these measures, as well as linguistic conformity, which was not imputed. First, given that we theorized that value congruence is relatively stable over time while perceptual congruence is more susceptible to change, we traced the two imputed measures over a person's tenure in the organization. We restricted this analysis to the first 36 months of employment given that only about 10% of employees had tenure exceeding 36 months during our observation period. We separately estimated OLS and fixed effect regressions of value congruence and perceptual congruence using indicators for each month (up to month 36 of employment). These results are depicted in Figure 3. According to both models, when employees first enter the organization, they have relatively high value congruence and relatively low perceptual congruence. Through approximately the first year of employment, however, perceptual congruence increases sharply and continues a more gradual ascent thereafter. In contrast, value congruence increases—albeit not as steeply—in the first four months of employment and then remains mostly stable over the remaining months. These results support the view that value congruence is relatively stable, while perceptual congruence is more malleable.

[FIGURE 3 ABOUT HERE]

Second, in Table 4 we report the results of OLS regressions with individual, department, and year fixed effects, where the dependent variable is bonus (logged) and independent variables—linguistic conformity, perceptual congruence (imputed) and value congruence (imputed)—are lagged. The fixed effects specification with lagged predictors allows us to estimate the effects of within-person changes in the two congruence measures and linguistic conformity on subsequent productivity.

Whether modeled independently or together, all three measures are significantly positively related to productivity. Thus we find, consistent with prior work, that linguistic conformity (Srivastava et al. 2018) and value congruence (Chatman 1991) are positively related to positive job performance—even when we use imputed longitudinal measures of value congruence and perceptual congruence. We also demonstrate that perceptual congruence is related to performance independent of its effects on linguistic conformity. Indeed, the coefficients for linguistic conformity and perceptual congruence are of similar magnitude, and the two variables retain their significance even when included together in Model 4. This suggests that perceptual congruence relates to performance above and beyond its influence on linguistic conformity.

In contrast, the association between value congruence and bonus is more modest. This result is consistent with our expectation that value congruence remains more stable over time. Given that the unwavering component of value congruence is subsumed in the individual fixed effect, it is not surprising that its time-varying component accounts for less of the variance in job performance.

It is also possible that having high value congruence is by itself not a guarantee that a person will have the ability to to behave in normatively compliant ways that are associated with positive performance. Overall, these supplemental analyses help to validate the value congruence and perceptual congruence measures derived from our imputation methodology. Figure 4 depicts the marginal effects of linguistic conformity, perceptual congruence, and value congruence on monthly bonus payments.

[TABLE 4 ABOUT HERE.]

[FIGURE 4 ABOUT HERE]

## Discussion and Conclusion

Adjustments to new and changing cultural environments are requirements in modern life. People's identities in contemporary society typically intersect many social boundaries—including ethnic, religious, political, occupational, and organizational. This crisscrossing of boundaries requires ongoing effort. The contemporary workplace—with its growing emphasis on culture on the one hand and employees' declining average tenure on the other—is a central arena in which these cultural transitions play out. Navigating the cultural heterogeneity across and within organizations involves maintaining multiple and partial commitments to different cultural orders, which in turn requires cultural awareness and adaptability (Friedland and Alford 1991, Morris et al. 2015, DiMaggio and Goldberg 2018).

Prior research has offered divergent explanations for why some people fit into their organizations better than others. One perspective has highlighted the importance of alignment between individual and group values in shaping behavior, while another has emphasized the role of situational cues and the ability to read the group's cultural code. We integrate these different perspectives by developing a two-sided theory of cultural fit that encompasses both values and perceptions. We argue and find empirical support for the notion that value congruence matters for deliberative choices such as whether to voluntarily exit an organization, while perceptual congruence instead shapes routine, habitual behavior—specifically, linguistic conformity with peers.

**Contributions**

Our theoretical framework and concomitant findings offer three core contributions. The first is in advancing person-culture fit theory. Specifically, we demonstrate that the behavioral consequences of cultural fit vary with different modes of cognition. More deliberative forms of cognition are operative when people contemplate how their values match those that prevail in the organization and shape group attachment outcomes such as when a person decides to depart an organization. In contrast, more automatic forms of cognition are at play when people determine how to communicate with their peers to enable effective interpersonal coordination. Together, these insights open the door to further investigations of the role that different modes of cognition play in shaping how people fit into social groups. Next, we demonstrate that both value congruence and perceptual congruence, as well as the behavioral manifestation of the latter, linguistic conformity, enable people to reap positive career rewards. Indeed, all three of our fit measures are positively linked to individual productivity, as indicated by bonus payments. Finally, our study complements the work of Goldberg et al. (2016), who examine the *consequences* of linguistic conformity for positive attainment in the form of favorable performance ratings and negative attainment in the form of involuntary exit. In contrast, the present study highlights the role of perceptual congruence as a key *antecedent* to linguistic conformity.

The conceptual separation of cultural fit into value congruence and perceptual congruence also paves the way for investigations into: (a) how these two dimensions relate to each other dynamically; (b) the degree to which they are influenced by individual-level characteristics versus broader structural factors; and (c) how features of organizational culture as a whole intersect with these individual-level measures of congruence. On the first question, we speculate that value congruence may provide a motivational channel through which a person is more or less vigilant in achieving and maintaining perceptual congruence. We similarly conjecture that people with chronically low value congruence may be able to maintain high perceptual congruence for a finite period of time but that doing so may, over time, adversely affect their identity and sense of self-worth (cf.

Hochschild 2012). Conversely, even if those with high perceptual congruence and low value congruence do not experience intrapsychic conflict, they may still experience the deleterious effects of being judged by others as inauthentic. Alternatively, we speculate that such individuals may—through self-perception and attribution processes (Ross 1977)—begin to experience an increase in value congruence. Examining the interrelationships between value congruence and perceptual congruence over time is a fruitful avenue for further developing theories of person-culture fit.

On the second question, we anticipate that perceptual congruence arises in part through the quality of a person's social network in an organization: Those who are connected to peers who are perceptually accurate are more likely to update their potentially flawed interpretations of the culture and converge to a more accurate understanding than are those whose peers are miscalibrated. We see great potential in future research that takes advantage of exogenous shifts in network structure—stemming, for example, from an unanticipated organizational restructuring that is undertaken for reasons unrelated to internal collaboration patterns—to causally identify the link between the perceptual congruence of peers and that of the focal individual. It remains to be explored the degree to which one's perceptual congruence is more a function of individual-level characteristics such as extraversion (John et al. 1999) or self-monitoring orientation (Snyder 1979) versus the interpersonal transmission of culture between individuals (Herrmann et al. 2013).

On the third question, we conjecture that organizations with strong cultures—ones in which values and norms are widely shared and strongly held (Chatman et al. 2014, Sørensen 2002)—will generally tend to attract and retain employees who exhibit high levels of value congruence. Yet this will not always be the case. Imagine, for example, an organization with a strong but toxic culture that offers compensation packages that are well above the industry average. People may choose to enter such an organization and perhaps feign alignment with the prevailing culture in their outward behavior, even as they exhibit low levels of value congruence. In similar fashion, such organizations may tend to have employees who exhibit generally high levels of perceptual congruence—though we again anticipate individual variation on perceptual congruence based on such factors as the quality of people's network connections and their rank. Future research could fruitfully examine the relationship between such facets of organizational culture as strength and intensity and the twin constructs of value congruence and perceptual congruence.

Next, we contribute to dual-process theories of culture and cognition (Vaisey 2009, Srivastava and Banaji 2011, Miles 2015, Lizardo et al. 2016) in two key ways. First, we make a conceptual link between different modes of cognition—more automatic versus more deliberative (Lizardo 2017)—and the types of behavior a person engages in within organizational settings. Whereas previous work in this tradition has thought about the link between values and behavior in binary terms—i.e., values either do or do not shape behavior—we develop a more nuanced account of the relationship. Our results indicate that values matter for some kinds of behavior (voluntary exit) but not others (linguistic conformity). This insight paves the way for exploring more generally how values matter for a broader range of behavior, especially when there is variation in the social reference groups that people perceive to be relevant (Diehl and McFarland 2010). Second, although dual-process theories of culture in action have proliferated, the empirical evidence in support of their link to concrete behaviors remains scant. We add to this evidence base by establishing a clear link between cultural fit constructs that are tied to automatic versus deliberative cognition and consequential behaviors such as how people communicate with their colleagues, their choice of voluntary exit, and their level of work productivity (as reflected in bonus payments). Extending the logic we develop, future work can examine the relationship between value congruence and other behaviors beyond voluntary exit that signal group attachment—for example, extra role behaviors such as the choice to work beyond normal work hours or invest in mentoring other group members (e.g., Ferris et al. 2018)—as well as the link between perceptual congruence and other subtle behaviors beyond linguistic conformity that reflect normative conformity—for example, knowing how much personal banter to engage in before diving into the substance of a meeting or whether positive deferrals (i.e., statements deferring to another member's status or expertise) are normative in group discussions (e.g., Chatman et al. 2008).

Finally, through this work, we make two methodological contributions. First, we bring together an established culture assessment approach (i.e., the OCP) with an unobtrusive data source (i.e., internal email communication) to open a window into complementary facets of culture. In doing so, we demonstrate how the integration of disparate approaches to assessing culture can yield richer and more complete insights into the complex cultural dynamics that occur within organizations. Second, building on Salganik's (2017) notion of "amplified asking," we demonstrate an empirical approach that transforms a one-time self-report into a longitudinal data set. Such an approach is of course, selectively appropriate, with requirements that include having a sufficient number of survey observations, access to rich communication content, protocols and safeguards to protect individual privacy and company confidentiality, and significant computational bandwidth. Nevertheless, given the ubiquity of digital trace data, the increasing difficulty of collecting survey data (particularly over time and from a large number of organizations), the widespread dissemination of off-the-shelf machine learning tools, and the declining cost of processing capacity, we anticipate that the pairing of self-reports and digital trace data will become increasingly common in social science research (Evans and Aceves 2016, McFarland et al. 2016, Lazer and Radford 2017).

**Limitations and Future Directions**

Although we develop a novel theoretical account of cultural fit and bring together disparate forms of data and analytical methods, we also acknowledge that the study has certain limitations. First, it is based on data about individuals working in a single organization with a strong culture that experienced a particular growth trajectory, which raises questions about the extent to which the findings would generalize to individuals working in other industries (e.g., those with different systemic turnover rates than what we observe in our setting) and types of organizations (e.g., those with relatively weak cultures or that experienced uniformly positive growth). As digital trace data of internal employee communications become increasingly accessible, future research can help establish the scope conditions of our findings by compiling comparable data from multiple organizational settings. We also leave to future research the task of investigating the robustness of our findings to alternative outcome measures. For example, there may be less variance in linguistic conformity in employees' written exchanges through the company email system than might be observed in their spoken communication. Similarly, even if people choose not to voluntarily depart from an organization, they may simply decide to work less hard if they experience a disconnect between their personal values and those that prevail in the organization. Given that the present study focused on more extreme behaviors such as diverging from prevailing norms in formal, written communications and voluntarily exiting from an organization, we conjecture that our results represent conservative estimates of the theorized relationships.

Next, our imputation models rely on the assumption that the relationship between language use and the relevant cultural fit variables is stable over time. Our results based on imputed measures should therefore be treated with caution, pending future studies that include multiple administrations of the OCP and thus allow this assumption to be empirically validated. Finally, even with the inclusion of individual fixed effects in our longitudinal models, we acknowledge that our estimates are correlational and do not identify causal relationships.

One possibility for pinning down a causal relationship between perceptual congruence and the outcomes of linguistic conformity and individual performance would be to implement a field experiment in which employees take an OCP, with a treatment group receiving feedback about how their perceptions differed from the actual perceptions of their interlocutors and a control group receiving no such feedback. Assuming such an intervention resulted in an increase in perceptual congruence in the treatment group but not in the control group, researchers could then examine whether it led to subsequent increases in the treatment group's linguistic conformity or performance relative to the control group's. In similar fashion, random assignment to formal socialization programs (e.g., assigning new hires to a veteran employee who can serve as a mentor or sponsor) might induce differences in value congruence between newcomers to an organization and influence retention over time.

## Conclusion

In sum, this study underscores that cultural fit is a two-sided construct that encompasses both values and perceptions. By considering both sides simultaneously, we can gain a more complete understanding of how the two facets of cultural fit independently and jointly shape organizational behavior.

## Endnotes

[1]We recognize that values and perceptions, although analytically distinct, may not be completely orthogonal since a person's interpretation of a stimulus, part of the perception process, is typically subject to her prior experience, motives, and expectations (e.g., Balcetis and Dunning 2006). The main distinction we are drawing is that values are much more focused on evaluations of what a person cares about—what is considered good or bad—while perceptions are closer to the less processed detection of stimuli (e.g., Gibson 1979, 2002). Thus, perceptions of an entity or object (such as an organization's culture) are significantly less subject to evaluation because they occur earlier in the information processing chain. In contrast, values develop even after attitude formation, which requires that sensation and perception have already occurred; however, values are even more ingrained, permanent, and stable in nature. They are also more general and less tied to any specific referent than is the case with many attitudes (England and Lee 1974, p. 412). Despite the extensive history of establishing discriminant validity among the information processing stages (e.g. Mesulam 1998), cognitive scientists generally acknowledge that there may be feedback loops from one step in the information processing sequence to the next.

[2]Although cultural fit is generally beneficial for individuals, there are cases in which high levels of cultural fit can be detrimental. For example, individuals who are ensconced in networks characterized by high levels of density may benefit from standing out from the crowd culturally (Goldberg et al. 2016). In a similar fashion, cultural homogeneity at the organizational level may be beneficial for interpersonal coordination but come at the cost of creativity and innovation (Flynn and Chatman 2001, Corritore et al. 2020).

[3]We distinguish perceptual congruence from two seemingly related constructs: cultural intelligence and self-monitoring. Whereas cultural intelligence and self-monitoring are stable psychological traits, perceptual congruence is a state that can ebb and flow across a person's tenure in an organization. Cultural intelligence is defined as "an individual's capability to function and manage effectively in culturally diverse settings" (Ang et al. 2007, p. 336). It is a relatively stable individual

difference that focuses on an individual's efficacy "in situations arising from differences in race, ethnicity, and nationality" (Ang et al. 2007, p. 336). Although some researchers have applied it conceptually to organizational culture (Earley and Mosakowski 2004), the construct is "uniquely relevant to intercultural contexts rather than monocultural contexts" (Van Dyne et al. 2019, p. 1). Similarly, high self-monitors (Snyder 1979) are consistently responsive to social cues of situational appropriateness (Snyder 1979, Kilduff and Day 1994, Sasovova et al. 2010). They tend to regulate their behavior given their read of what is expected of them, whereas low self-monitors adhere to their sense of self, irrespective of the situation. Self-monitoring is also related to a persistent capacity for deep-acting, the ability to adapt emotions to organizational expectations, leading to more genuine displays of cultural congruence (Grandey 2000, Scott et al. 2012). Like cultural intelligence, self-monitoring is an individual difference that is often described as a personality characteristic (Ang et al. 2006, Snyder 1979). It is quite likely that individuals who are higher in cultural intelligence or self-monitoring will exhibit, all else equal, greater perceptual congruence. Yet perceptual congruence ultimately reflects an individual's exposure to culturally relevant signals given off by others. Even those with high levels of cultural intelligence or self-monitoring will misinterpret the cultural code if the peers they learn the code from are behavioral misfits. Moreover, the signals that give rise to perceptual congruence are independent of the dimensions that matter for cultural intelligence—chiefly race, ethnicity, and nationality.

[4]Although we follow Goldberg et al. (2016) and Srivastava et al. (2018) in how we operationalize linguistic conformity, we depart from them in how we label this construct. They refer to linguistic conformity with peers as a behavioral measure of "cultural fit." Given that we consider multiple manifestations of cultural fit in this paper, to avoid confusion, we generally refer to the specific constructs of value congruence, perceptual congruence, and linguistic conformity.

[5]Because we see perceptual congruence informing behavior that is more collective in nature and value congruence shaping more private forms of behavior, we also do not theorize a potential interaction effect between the two congruence constructs. In the discussion section below, we identify

potential avenues for future research to examine potential interrelationships between perceptual congruence and value congruence.

[6]The other half completed an assessment of the cultural characteristics needed for the organization to be successful in the future. We shared the results of this latter assessment with organizational leaders as a condition of gaining access to the organization as a research site; however, we do not report these results here because they do not pertain to our theory and hypotheses.

[7]Archived email data and personnel records were collected in multiple batches starting in 2015 and concluding toward the end of 2016. The OCP was administered in October of 2016.

[8]This represents an advantage over alternative techniques such as word embedding models (Mikolov et al. 2013, Kozlowski et al. 2019, Lix et al. 2022), which would reveal semantic differences between an individual's communication and that of her peers. Such a measure would, however, not distinguish between communication related to functional coordination versus to styles of discourse that are normatively reinforced.

[9]The company's human resources information system included a field for whether a person was active or inactive in a given month. It also included a field for exit date for those who were inactive, as well as a field for whether the exit was voluntary or involuntary based on how human resources coded the departure. Employees who were in good standing but chose to leave of their own accord (i.e., they were viewed as"regretted" departures) were coded as voluntary departures. The code of involuntary departure was assigned to employees who were terminated at the company's discretion (e.g., laid off or asked to leave the company for poor performance or other breaches of the employment contract). In our panel data set, a given individual appears in every month from the time of hire to the time of exit. Voluntary exit is an indicator variable set of 0 for all months the person was active in the organization and to 1 for a person's final month (assuming the exit was coded by HR as being a voluntary rather than involuntary departure).

[10]Managerial status and departmental affiliation can be estimated in fixed effect models because some employees get promoted from individual contributor to managerial roles and because some employees move across departments.

[11]In unreported analyses, we also assessed potential interaction effects between perceptual congruence and value congruence and find that the interaction term is not a significant predictor of linguistic conformity.

[12]Neither perceptual congruence nor value congruence significant predicts involuntary exit when we use the same framework with voluntary exit as the competing risk. We also find that the interaction of perceptual congruence and value congruence does not significantly predict voluntary exit.

## References

Alba R, Nee V (2009) *Remaking the American Mainstream: Assimilation and Contemporary Immigration* (Harvard University Press).

Alesina A, Algan Y, Cahuc P, Giuliano P (2015) Family values and the regulation of labor. *Journal of the European Economic Association* 13(4):599–630.

Ang S, Van Dyne L, Koh C (2006) Personality correlates of the four-factor model of cultural intelligence. *Group & Organization Management* 31(1):100–123.

Ang S, Van Dyne L, Koh C, Ng YK, Templer KJ, Tay C, Chandrasekar AN (2007) Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation, and task performance. *Management and Organization Review* 3(3):335–371.

Bail CA, Brown TW, Mann M (2017) Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation. *American Sociological Review* 0003122417733673.

Balcetis E, Dunning D (2006) See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology* 91(4):612.

Baron JN, Hannan MT, Burton MD (2001) Labor pains: Change in organizational models and employee turnover in young, high-tech firms. *American Journal of Sociology* 106(4):960–1012.

Block J (1961) *The Q-sort Method in Personality Assessment and Psychiatric Research* (Thomas Springfield, IL).

Boltanski L, Thévenot (2006) *On Justification: Economies of Worth* (Princeton University Press).

Cable DM, Judge TA (1996) Person–organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes* 67(3):294–311.

Casciaro T (1998) Seeing things clearly: Social structure, personality, and accuracy in social network perception. *Social Networks* 20(4):331–351.

Chatman JA (1989) Improving interactional organizational research: A model of person-organization fit. *Academy of Management Review* 14(3):333–349.

Chatman JA (1991) Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly* 36(3):459.

Chatman JA, Barsade SG (1995) Personality, organizational culture, and cooperation: Evidence from a business simulation. *Administrative Science Quarterly* 423–443.

Chatman JA, Boisnier AD, Spataro SE, Anderson C, Berdahl JL (2008) Being distinctive versus being conspicuous: The effects of numeric status and sex-stereotyped tasks on individual performance in groups. *Organizational Behavior and Human Decision Processes* 107(2):141–160.

Chatman JA, Caldwell DF, O'Reilly CA, Doerr B (2014) Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior* 35(6):785–808.

Chatman JA, O'Reilly CA (2016) Paradigm lost: Reinvigorating the study of organizational culture. *Research in Organizational Behavior* 36:199–224.

Cheryan S, Ziegler SA, Montoya AK, Jiang L (2017) Why are some stem fields more gender balanced than others? *Psychological Bulletin* 143(1):1.

Childress CC, Friedkin NE (2012) Cultural reception and production: The social construction of meaning in book clubs. *American Sociological Review* 77(1):45–68.

Corritore M, Goldberg A, Srivastava SB (2020) Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly* 65(2):359–394.

Cremer J, Garicano L, Prat A (2007) Language and the theory of the firm. *The Quarterly Journal of Economics* 122(1):373–407.

Davis-Blake A, Pfeffer J (1989) Just a mirage: The search for dispositional effects in organizational research. *Academy of Management Review* 14(3):385–400.

Dawis RV (2005) The minnesota theory of work adjustment. Brown SD, Lent RW, eds., *Career Development and Counseling: Putting Theory and Research to Work*, 3–23 (John Wiley & Sons, Inc.).

Diehl D, McFarland D (2010) Toward a historical sociology of social situations. *American Journal of Sociology* 115(6):1713–1752.

DiMaggio P, Goldberg A (2018) Searching for homo economicus: Variation in americans' construals of and attitudes toward markets. *European Journal of Sociology* 1–39.

Doyle G, Goldberg A, Srivastava S, Frank M (2017) Alignment at work: Using language to distinguish the internalization and self-regulation components of cultural fit in organizations. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 603–612.

Earley CP, Mosakowski E (2004) Cultural intelligence. *Harvard Business Review* 82(10):139–146.

Edwards JR (2008) Person-environment fit in organizations. *The Academy of Management Annals* 2(1):167–230.

Eliasoph N, Lichterman P (2003) Culture in interaction. *American Journal of Sociology* 108(4):735–794.

England GW, Lee R (1974) The relationship between managerial values and managerial success in the united states, japan, india, and australia. *Journal of Applied Psychology* 59(4):411.

Evans JA, Aceves P (2016) Machine translation: Mining text for social theory. *Annual Review of Sociology* 42:21–50.

Ferris GR, Bhawuk DP, Fedor DF, Judge TA (2018) Organizational politics and citizenship: Attributions of intentionality and construct definition. *Attribution Theory*, 231–252 (Routledge).

Flynn FJ, Chatman JA (2001) Strong cultures and innovation: Oxymoron or opportunity. *International Handbook of Organizational Culture and Climate* 263–87.

Friedland R, Alford RR (1991) *Bringing Society Back in: Symbols, Practices and Institutional Contradictions* (University of Chicago Press).

Friedman J, Hastie T, Tibshirani R (2001) *The Elements of Statistical Learning*, volume 1 (Springer series in statistics New York).

Gerald JH, George SH (2010) Self-report: Psychology's four-letter word. *American Journal of Psychology* 123(2):181–188.

Gibson DR (2011) Avoiding catastrophe: The interactional production of possibility during the cuban missile crisis. *American Journal of Sociology* 117(2):361–419.

Gibson JJ (1979) *An Ecological Theory of Perception* (Boston, MA: Houghton Miflin).

Gibson JJ (2002) A theory of direct visual perception. *Vision and Mind: Selected Readings in the Philosophy of Perception* 77–90.

Goldberg A, Srivastava SB, Manian GV, Monroe W, Potts C (2016) Fitting in or standing out? the tradeoffs of structural and cultural embeddedness. *American Sociological Review* 81(6):1190–1222.

Grandey AA (2000) Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology* 5(1):95.

Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102(1):4.

Hall DT (1996) *The Career Is Dead–Long Live the Career. A Relational Approach to Careers. The Jossey-Bass Business & Management Series.* (ERIC).

Harding DJ (2007) Cultural context, sexual behavior, and romantic relationships in disadvantaged neighborhoods. *American Sociological Review* 72(3):341–364.

Harman WS, Lee TW, Mitchell TR, Felps W, Owens BP (2007) The psychology of voluntary employee turnover. *Current Directions in Psychological Science* 16(1):51–54.

Herrmann PA, Legare CH, Harris PL, Whitehouse H (2013) Stick to the script: The effect of witnessing multiple actors on children's imitation. *Cognition* 129(3):536–543.

Ho TK (1995) Random decision forests. *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, 278–282 (IEEE).

Hochschild AR (2012) *The Managed Heart: Commercialization of Human Feeling* (Univ of California Press).

Hofstede G (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (SAGE Publications).

Inglehart R, Baker WE (2000) Modernization, cultural change, and the persistence of traditional values. *American Sociological Review* 65(1):19–51.

John OP, Srivastava S, et al. (1999) *The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives*, volume 2 (University of California Berkeley).

Judge TA, Cable DM (1997) Applicant personality, organizational culture, and organization attraction. *Personnel Psychology* 50(2):359–394.

Keister LA (2008) Conservative protestants and wealth: How religion perpetuates asset poverty. *American Journal of Sociology* 113(5):1237–1271.

Kilduff M, Day DV (1994) Do chameleons get ahead? the effects of self-monitoring on managerial careers. *Academy of Management Journal* 37(4):1047–1060.

Kozlowski AC, Taddy M, Evans JA (2019) The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review* 84(5):905–949, ISSN 0003-1224, URL `http://dx.doi.org/10.1177/0003122419877135`.

Kristof-Brown AL, Zimmerman RD, Johnson EC (2005) Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology* 58:281–342.

Kullback S, Leibler RA (1951) On information and sufficiency. *The Annals of Mathematical Statistics* 22(1):79–86.

Lazear EP (1999) Culture and language. *Journal of Political Economy* 107(S6):S95–S126.

Lazer D, Radford J (2017) Data ex machina: Introduction to big data. *Annual Review of Sociology* 43:19—39.

Lee TW (2014) Why employees quit. *Decision Making in the Workplace*, 73–90 (Psychology Press).

Leschziner V (2019) Dual-process models in sociology. *The Oxford Handbook of Cognitive Sociology* 169–191.

Lewis D (2008) *Convention: A Philosophical Study* (John Wiley & Sons).

Lin J (1991) Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory* 37(1):145–151.

Lix K, Goldberg A, Srivastava SB, Valentine MA (2022) Aligning differences: Discursive diversity and team performance. *Management Science* .

Lizardo O (2017) Improving cultural analysis: Considering personal culture in its declarative and nondeclarative modes. *American Sociological Review* 82(1):88–115.

Lizardo O, Mowry R, Sepulvado B, Stoltz DS, Taylor MA, Van Ness J, Wood M (2016) What are dual process models? implications for cultural analysis in sociology. *Sociological Theory* 34(4):287–310.

Lofquist LH, Dawis R (1978) Values as second-order needs in the theory of work adjustment. *Journal of Vocational Behavior* 12(1):12–19.

Mayer DM, Kuenzi M, Greenbaum R, Bardes M, Salvador RB (2009) How low does ethical leadership flow? test of a trickle-down model. *Organizational Behavior and Human Decision Processes* 108(1):1–13.

McFarland DA, Lewis K, Goldberg A (2016) Sociology in the era of big data: The ascent of forensic social science. *The American Sociologist* 47(1):12–35.

Mesulam MM (1998) From sensation to cognition. *Brain: A Journal of Neurology* 121(6):1013–1052.

Meyer JP, Hecht TD, Gill H, Toplonytsky L (2010) Person–organization (culture) fit and employee commitment under conditions of organizational change: A longitudinal study. *Journal of Vocational Behavior* 76(3):458–473.

Michaels CE, Spector PE (1982) Causes of employee turnover: A test of the mobley, griffeth, hand, and meglino model. *Journal of Applied Psychology* 67(1):53.

Michel A (2011) Transcending socialization: A nine-year ethnography of the body's role in organizational control and knowledge workers' transformation. *Administrative Science Quarterly* 56(3):325–368.

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed Representations of Words and Phrases and their Compositionality. Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds., *Advances in Neural Information Processing Systems 26*, 3111–3119 (Curran Associates, Inc.).

Miles A (2015) The (re)genesis of values: Examining the importance of values for action. *American Sociological Review* 80(4):680–704.

Morris MW, Chiu Cy, Liu Z (2015) Polycultural psychology. *Annual Review of Psychology* 66:631–659.

Nisbett RE, Cohen D (1996) *Culture of Honor: The Psychology of Violence in the South* (Westview Press).

O'Reilly CA, Chatman J (1986) Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. *Journal of Applied Psychology* 71(3):492.

O'Reilly CA, Chatman J, Caldwell DF (1991) People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal* 34(3):487–516.

Parsons T (1968) *The Structure of Social Action* (Free press).

Pennebaker JW (2013) *The Secret Life of Pronouns: What Our Words Say About Us* (Bloomsbury USA), ISBN 9781608194964, URL `https://books.google.com/books?id=p9KmCAAAQBAJ`.

Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ (2007) *Linguistic Inquiry and Word Count (LIWC): LIWC2007* (Austin, TX: LIWC.net).

Pinker S (2007) *The Stuff of Thought: Language as a Window Into Human Nature* (New York, NY: Viking).

Potts C (2011) Sentiment-aware tokenizer. *Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License: http://creativecommons.org/licenses/by-nc-sa/3.0/* .

Rafaeli A, Pratt MG (1993) Tailored meanings: On the meaning and impact of organizational dress. *Academy of Management Review* 18(1):32–55.

Rivera LA (2012) Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review* 77(6):999–1022.

Rokeach M (1973) *The Nature of Human Values.* (Free press).

Ross L (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology* 10:173–220.

Salancik GR, Pfeffer J (1974) The bases and use of power in organizational decision making: The case of a university. *Administrative Science Quarterly* 453–473.

Salganik MJ (2017) *Bit by Bit: Social Research in the Digital Age* (Princeton University Press).

Sasovova Z, Mehra A, Borgatti SP, Schippers MC (2010) Network churn: The effects of self-monitoring personality on brokerage dynamics. *Administrative Science Quarterly* 55(4):639–670.

Schwartz SH (1992) Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Zanna MP, ed., *Advances in Experimental Social Psychology*, volume 25, 1 – 65 (Academic Press).

Scott BA, Barnes CM, Wagner DT (2012) Chameleonic or consistent? a multilevel investigation of emotional labor variability and self-monitoring. *Academy of Management Journal* 55(4):905–926.

Snyder M (1979) Self-monitoring processes. *Advances in Experimental Social Psychology* 12:85–128.

Sørensen JB (2002) The strength of corporate culture and the reliability of firm performance. *Administrative Science Quarterly* 47(1):70–91.

Srivastava SB, Banaji MR (2011) Culture, cognition, and collaborative networks in organizations. *American Sociological Review* 76(2):207–233.

Srivastava SB, Goldberg A, Manian GV, Potts C (2018) Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science* 64(3):1348–1364.

Swidler A (1986) Culture in action: Symbols and strategies. *American Sociological Review* 273–286.

Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

Thompson JD (1967) *Organizations in Action: Social Science Bases of Administrative Theory* (Transaction publishers).

Vaisey S (2009) Motivation and justification: A dual-process model of culture in action. *American Journal of Sociology* 114(6):1675–1715.

Van Dyne L, Ang S, Tan ML (2019) Cultural intelligence. *Oxford Bibliographies* (Englewood Cliffs, NJ: Prentice-hall).

Weber RA, Camerer CF (2003) Cultural conflict and merger failure: An experimental approach. *Management Science* 49(4):400–415.

Weeks J (2004) *Unpopular Culture: The Ritual of Complaint in a British Bank* (University of Chicago Press).

# FIGURES

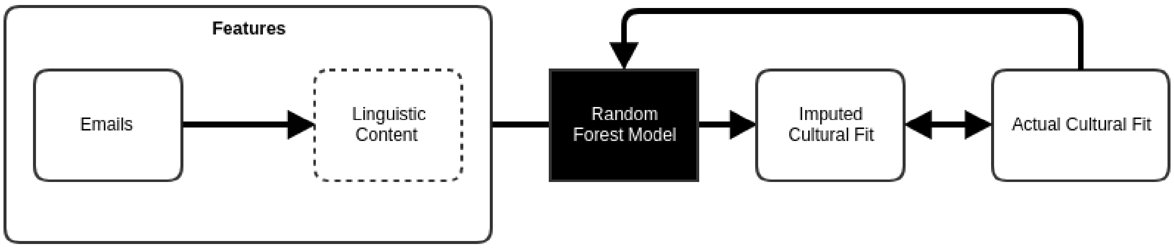


**Figure 1** **Conceptual Overview of the Machine Learning Process**

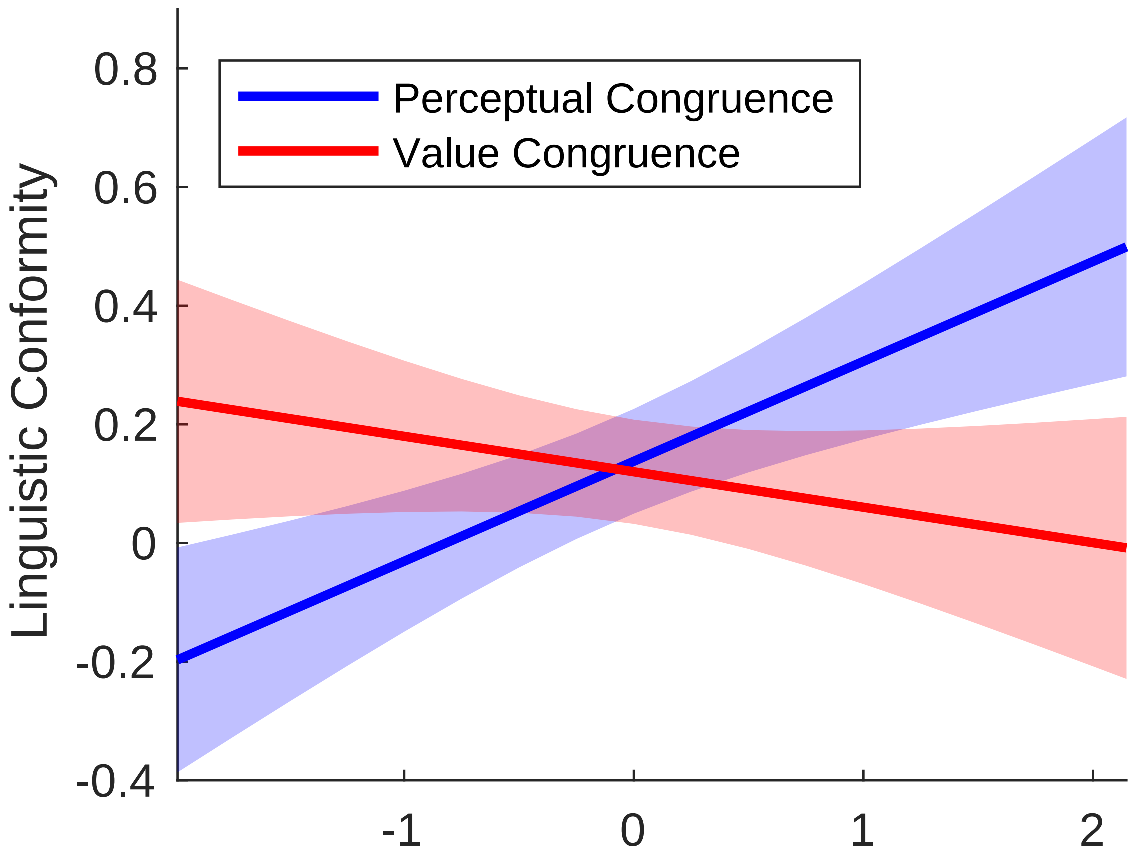


**Figure 2** **Marginal effects of perceptual congruence and value congruence on linguistic conformity based on Table 2, Model 3. Note that linguistic conformity is not significantly related to value congruence, whereas it is positively and significantly related to perceptual congruence. A one standard deviation increase in perceptual congruence is associated with a 0.15 standard deviation increase in linguistic conformity.**

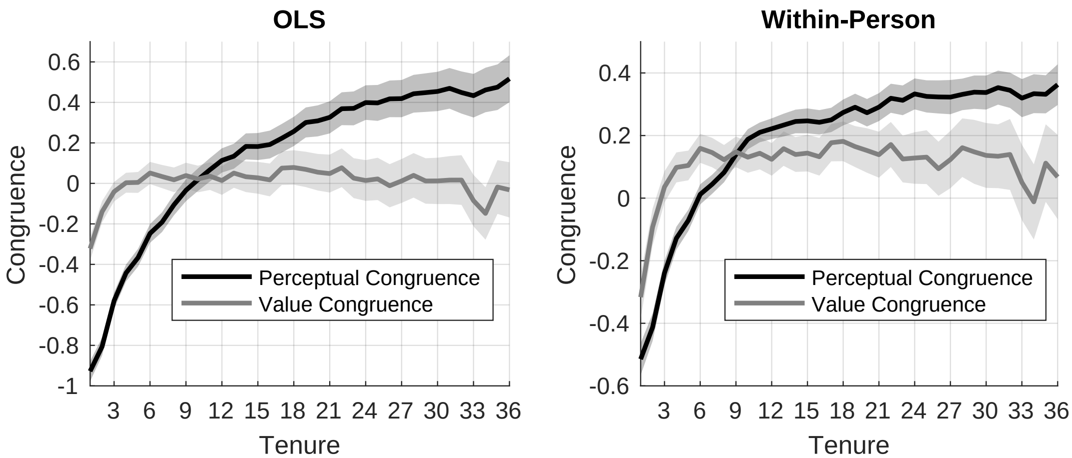


**Figure 3** **OLS and fixed effect regressions of perceptual congruence and value congruence, with indicators for each tenure month up to 36 months in the company.**

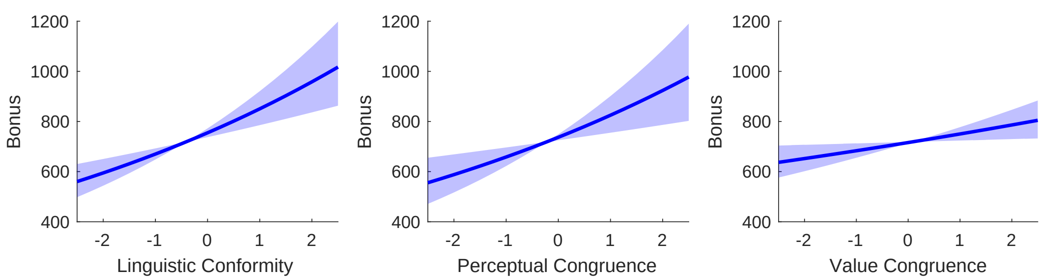


**Figure 4** **Marginal effects of linguistic conformity, perceptual congruence, and value congruence on monthly bonus payments based on Table 4, Model 4.**

# TABLES

**Table 1 Correlation Matrix**

| Variables | Linguistic Conformity | Perceptual Congruence | Value Congruence | Manager | Female | Age |
|---|---|---|---|---|---|---|
| Linguistic Conformity | 1.000 | | | | | |
| | | | | | | |
| Perceptual Congruence | 0.250 | 1.000 | | | | |
| | (0.000) | | | | | |
| Value Congruence | -0.032 | 0.132 | 1.000 | | | |
| | (0.644) | (0.061) | | | | |
| Manager | 0.397 | 0.168 | 0.107 | 1.000 | | |
| | (0.000) | (0.001) | (0.124) | | | |
| Female | 0.049 | 0.064 | -0.178 | -0.026 | 1.000 | |
| | (0.327) | (0.211) | (0.010) | (0.601) | | |
| Age | -0.013 | -0.073 | 0.118 | 0.166 | -0.055 | 1.000 |
| | (0.794) | (0.151) | (0.088) | (0.001) | (0.273) | |

p-values in parentheses

**Table 2 Cross-Sectional and Longitudinal Fixed Effects Regressions of Linguistic Conformity**

| | Cross-Sectional | | | Longitudinal | | |
|---|---|---|---|---|---|---|
| | Model 1$^{\dagger}$ | Model 2$^{\dagger}$ | Model 3$^{\dagger}$ | Model 4$^{\ddagger}$ | Model 5$^{\ddagger}$ | Model 6$^{\ddagger}$ |
| Perceptual Congruence | 0.138$^{***}$ | | 0.168$^{***}$ | 0.049$^{**}$ | | 0.048$^{**}$ |
| | (4.24) | | (3.87) | (3.14) | | (3.11) |
| Value Congruence | | -0.029 | -0.060 | | 0.014 | 0.013 |
| | | (-0.58) | (-1.15) | | (1.40) | (1.33) |
| Manager | 0.664$^{***}$ | 0.711$^{***}$ | 0.639$^{***}$ | 0.299$^{***}$ | 0.303$^{***}$ | 0.298$^{***}$ |
| | (7.67) | (5.24) | (4.99) | (5.39) | (5.44) | (5.37) |
| Female | 0.052 | -0.034 | -0.059 | | | |
| | (0.73) | (-0.33) | (-0.57) | | | |
| Age | 0.023 | 0.011 | 0.028 | | | |
| | (0.85) | (0.29) | (0.74) | | | |
| Age$^2$ | -0.000 | -0.000 | -0.000 | | | |
| | (-1.00) | (-0.32) | (-0.72) | | | |
| Constant | -0.223 | -0.128 | -0.429 | -0.351$^{**}$ | -0.380$^{**}$ | -0.353$^{**}$ |
| | (-0.42) | (-0.17) | (-0.59) | (-3.19) | (-3.26) | (-3.22) |
| Individual FE | No | No | No | Yes | Yes | Yes |
| Department FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | No | No | No | Yes | Yes | Yes |
| Observations | 386 | 209 | 202 | 24215 | 24215 | 24215 |
| R$^2$ | 0.257 | 0.189 | 0.244 | 0.042 | 0.040 | 0.042 |

*t* statistics in parentheses; standard errors clustered by individual when individual fixed effects are used

$^{\dagger}$ Linguistic congruence is averaged over 3 months, $^{\ddagger}$ Imputed and lagged measures in Models 4-6

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

**Table 3 Competing Risks Model of Voluntary Exit**

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Perceptual Congruence | 0.994 | | 0.994 |
| | (-0.09) | | (-0.09) |
| Value Congruence | | $0.874^{*}$ | $0.874^{*}$ |
| | | (-2.30) | (-2.30) |
| Manager | 0.845 | 0.873 | 0.876 |
| | (-0.71) | (-0.58) | (-0.56) |
| Female | $1.387^{*}$ | $1.392^{*}$ | $1.392^{*}$ |
| | (2.54) | (2.56) | (2.57) |
| Age | $0.899^{***}$ | $0.899^{***}$ | $0.899^{***}$ |
| | (-3.32) | (-3.35) | (-3.33) |
| $\text{Age}^2$ | $1.001^{**}$ | $1.001^{***}$ | $1.001^{***}$ |
| | (3.29) | (3.34) | (3.31) |
| Firm Size (Employees) | $1.002^{***}$ | $1.002^{***}$ | $1.002^{***}$ |
| | (9.94) | (10.56) | (9.96) |
| Department FE | Yes | Yes | Yes |
| Observations | 27452 | 27452 | 27452 |
| chi2 | 194.530 | 199.505 | 201.607 |
| ll | -1.3e+03 | -1.3e+03 | -1.3e+03 |

Exponentiated coefficients; $t$ statistics in parentheses
Standard errors clustered by individual; sample weights by tenure
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

**Table 4 Fixed Effect OLS Regressions of Log Bonus on Covariates**

| height | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Linguistic Conformity† | 0.128*** | | | 0.119*** |
| | (4.46) | | | (4.14) |
| Perceptual Congruence† | | 0.132*** | | 0.112** |
| | | (3.98) | | (3.05) |
| Value Congruence† | | | 0.057** | 0.047* |
| | | | (3.17) | (2.36) |
| Manager | -0.194 | 0.025 | 0.063 | -0.180 |
| | (-1.13) | (0.13) | (0.31) | (-1.02) |
| Constant | 5.652*** | 5.420*** | 5.303*** | 5.702*** |
| | (28.20) | (26.70) | (25.70) | (28.61) |
| Individual FE | Yes | Yes | Yes | Yes |
| Department FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 4782 | 6377 | 6377 | 4778 |
| Number of Individuals (Clusters) | 1057 | 1303 | 1303 | 1056 |
| $R^2$ | 0.059 | 0.043 | 0.040 | 0.065 |

$t$ statistics in parentheses; standard errors clustered by individual

† lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

# APPENDIX A: LINGUISTIC CONFORMITY

## The Interactional Language Use Model

We implement the procedure detailed in Goldberg et al. (2016) and Srivastava et al. (2018) to measure behavioral fit. We begin by using LIWC to translate each individual's outgoing and incoming messages in each period $t$ (defined as a calendar month) into probability distributions over the 64 LIWC categories. Table A1 lists the categories, provides illustrative words in each category, and indicates the number of words per category. Specifically, we define $\overrightarrow{m}_{it}$ as each email individual $i$ sends at time $t$ and $\overleftarrow{m}_{it}$ as each email individual $i$ *receives* at time $t$. We then define the set of LIWC categories as $L$ and the set of all times in any given month as $T$. Our procedure iterates over all emails sent and received and produces $\overrightarrow{m}_{it}^{l}$ and $\overleftarrow{m}_{it}^{l}$ for the count of terms in email $\overrightarrow{m}_{it}$ and $\overleftarrow{m}_{it}$ in LIWC category $l \in L$, respectively. Then, by aggregating all individual email counts $\overrightarrow{m}_{it}^{l}$ and $\overleftarrow{m}_{it}^{l}$ for $t \in T$, it produces sent and received LIWC counts in month $T$, $\overrightarrow{m}_{iT}^{l}$ and $\overleftarrow{m}_{iT}^{l}$. We normalize each LIWC count in each month by the total of all LIWC counts in that month to transform the LIWC probability distribution to a standard probability distribution. We use the notation, $O_{iT}^{l}$ to denote the outgoing normalized probability and $I_{iT}^{l}$ to denote the incoming normalized probability.

$$O_{iT}^{l} = \frac{\overrightarrow{m}_{iT}^{l}}{\sum_{l \in L} \overrightarrow{m}_{iT}^{l}} \tag{1}$$

$$I_{iT}^{l} = \frac{\overleftarrow{m}_{iT}^{l}}{\sum_{l \in L} \overleftarrow{m}_{iT}^{l}} \tag{2}$$

[TABLE A1 ABOUT HERE.]

We define an individual $i$'s linguistic conformity in month $T$ as the negative log of the Jensen-Shannon (JS) divergence (Lin 1991) metric between $i$'s outgoing and incoming normalized distributions:

$$BF_{iT} = -log\,(JS(O_{iT} \parallel I_{iT})) \tag{3}$$

where the JS-divergence between two probability distributions is defined as a symmetric measure built by first taking the mean probability distribution between the normalized outgoing and incoming distributions, $M_{iT} = \frac{1}{2}(O_{iT} + I_{iT})$, and summing the Kullback-Leibler (KL) divergence

(Kullback and Leibler 1951) of the outgoing and incoming distributions from that mean probability distribution.

$$JS(O_{iT} \parallel I_{iT}) = \frac{1}{2} KL\left(O_{iT} \parallel M_{iT}\right) + \frac{1}{2} KL\left(I_{iT} \parallel M_{iT}\right) \tag{4}$$

$$KL(D_{iT} \parallel M_{iT}) = \sum_{l \in L} D^{l}_{iT} \log_2 \frac{D^{l}_{iT}}{M^{l}_{iT}} \tag{5}$$

**Validation of Linguistic Conformity**

We have argued above that the LIWC lexicon, on which the linguistic conformity measure is based, is a useful categorization scheme for measuring culturally meaningful behaviors. Indeed, as previous work demonstrates (e.g. Goldberg et al. 2016, Srivastava et al. 2018), this measure of linguistic conformity is effective at predicting individual attainment in an organization. Since this is the first time our measure of linguistic conformity has been related to a validated measure of organizational culture, the OCP, we also sought assurances that the LIWC categories contained face valid connections to the existing OCP dimensions. Therefore, we conducted two types of analyses to further establish the behavioral measure's construct validity.

First, we compared respondents' language use to their responses to the OCP survey. Recall that we asked respondents to describe their desired culture (personal culture survey) and their perception of the organizational culture (current culture survey). We expected there to be a systematic relationship between people's desired and perceived cultures on the one hand and their linguistic behaviors on the other. For example, it would seem plausible that a preference for a people-oriented cultural environment would be reflected in greater use of affective words. Thus, we expected to observe a systematic relationship between people's cultural preferences and perceptions, as reflected in their explicit responses to the OCP and their use of language as captured by LIWC.

To examine this, we compared individuals' rankings of the 54 OCP categories with their LIWC category frequencies in outgoing email communication in a 3-month period close to the OCP survey administration. For the personal culture survey, we found 229 significantly correlated ($p < 0.05$) pairs of OCP and LIWC categories (with sample size of 231 individuals). For the current culture survey, we found 583 significant correlations (for 414 individuals). We found an even greater number of significant OCP/LIWC pair correlations when comparing the current culture survey to respondents' incoming email communication, suggesting that—consistent with our hypotheses—individuals' perceptions of the culture are inherently related to the behaviors they observe. We also compared LIWC frequencies to the eight high-level OCP categories (such as collaborative or detail-oriented, see Chatman et al. (2014) for details). For the personal cultural survey we find that 34% of LIWC categories are correlated with at least one high-level dimension, and that 85% of LIWC categories are correlated with at least one high-level dimension in the current culture survey. Together, these analyses indicate that LIWC use significantly and substantially co-varies with desired and perceived culture.

As illustration, we examine the link between language use and a preference for a people orientated culture. We find that respondents who value people orientation tend to include more affect words (e.g., happy, cry, abandon), perceptual process words (e.g., observe, hear, feel), positive emotion words (e.g., love, nice, sweet), and second-person words (e.g., you, your) in their outgoing communication.[13]. We refrain from substantively interpreting these findings, but we view them as qualitative evidence for the cultural meaningfulness of LIWC use and leave a systematic exploration of the complex relationship between stated beliefs and naturally occurring linguistic behaviors to future work.

In our second test of the construct validity of our linguistic conformity measure, we recognized that LIWC was originally developed as a means to identify the linguistic signatures of psychological, rather than purely cultural categories. Whereas some linguistic categories contained in the LIWC lexicon, such as swearing, are clearly inherently related to culture, others, such as the use of articles, are more ambiguously cultural. Thus, we sought to understand whether our linguistic conformity measure represented a meaningful and relevant set of culturally oriented linguistic categories.

Before discussing these analyses in detail we highlight why we assume that LIWC categories are culturally meaningful. Specifically, while some LIWC categories may initially appear to be unrelated to culture, extensive research by Pennebaker (2013) suggests that the categories are meaningful at both a psychological and sociological level. For example, the use of articles such as *a*, *an* or *the*—each of which seemingly represents a minute technical linguistic decision—actually reflects the speaker's emotional stability, organization, and conservatism (Pennebaker 2013). A group that uses a linguistic style that emphasizes articles might therefore be indicative of a rule-oriented culture that emphasizes attention to detail.

Thus, rather than requiring a typology that distinguishes non-cultural from cultural LIWC categories and that maps the latter to underlying cultural dimensions, we assumed that all LIWC categories are culturally meaningful and that the same category might vary in its cultural meaning across contexts. Our measure of behavioral cultural fit therefore takes all LIWC categories into account and does not privilege certain categories over others.

To test our assumption, we analyzed the measure's robustness to LIWC category inclusion. Let $k < 64$ be the size of a subset of LIWC categories used to generate an alternative measure of linguistic conformity, labeled $BF_k$. We randomly selected $k$ LIWC categories and constructed the measure as we did above (according to equation 3), using only this subset of categories. We repeated this process 1,000 times for each value of $k$ (because $\binom{64}{k}$ is extremely large for most values of $k$, we could not realistically explore all possible subsets). For each $BF_k$ that we generated, we identified its correlation with the original $BF$ measure based on all 64 categories.

We report the average correlation between $BF_k$ and $BF$ for all 1,000 random samples in Figure A1. As the plot clearly indicates, the linguistic conformity measure is robust regardless of whether LIWC categories are removed. The measure remains effectively unchanged even if half of the LIWC categories are removed. We interpret these results as an indication of two properties. First, linguistic conformity is not driven by one or a handful of LIWC categories. It is therefore not merely a reflection of a specific linguistic feature or style. Second, the pattern illustrated in Figure A1 indicates that even if certain LIWC categories are culturally irrelevant in this context, their inclusion in the measure construction does not bias its value. In other words, even if we were to conclude that half of the LIWC categories are non-cultural (a conclusion that, for the reasons stated above, we believe is unwarranted) and decide to remove them from the measure, we would still recover near-identical values.

# APPENDIX B: MACHINE LEARNING PROCEDURE

## Overview

The procedure consisted of five major steps, which are documented at a conceptual level in Figure 1 in the main manuscript and described in greater detail below.

Our first step was to translate the raw email data into a format that is usable by the random forest model. We tokenized and stemmed all words in the body of email messages. Tokenization involves separating the text into distinct terms, for which we used the TwitterTokenizer designed for linguistic analysis Potts (2011). Stemming involves reducing each term to a root form, for which we used the Porter Stemmer from the python nltk package. We removed all characters that could not be encoded into unicode, such as "\x00," and split the text into n-stems, where n is in the set [1,2,3]. Given that language use tends to follow the power law, in which few terms are used frequently and many terms are used infrequently, we then undertook steps to reduce the dimensionality of the data to make it computationally tractable. We retained all n-stems in emails sent from individuals, but only uni-stems in emails sent to individuals. Additionally, we retained only those n-stems that were used by at least 1% of employees in a subsample of emails. Finally, we used principal component analysis (PCA) to further reduce dimensionality, retaining only the top 3,000 PCA components for each type of n-stem. These resulting components served as the feature inputs to our model.

The second step was to transform our measures of cognitive cultural fit into categories that are more conducive to classification given the relatively small number of observations from which we had to fit the model. Recall that perceptual congruence and value congruence were computed as correlations, ranging from 0 to 1. We transformed these continuous measures into three discrete categories–low, medium, and high. Intuitively, this allowed our model to detect distinctive features of belonging to each category, an important characteristic to which we will return when we discuss the testing of our model. For perceptual congruence, we set the cutoffs for low fit at 20% and for high fit at 80%, with everything else considered medium fit. For value congruence, for which we had even fewer observations, we had to set more extreme cutoffs at 10% and 90% to achieve strong model fit.

The third step was to use our feature inputs and their now-discrete mappings to cognitive cultural fit to train a random forest model. The random forest model is an ensemble method, which means it aggregates and blends multiple independent decision trees (Ho 1995, Friedman et al. 2001). After several such decisions according to specific features of the input, all of the inputs are sorted into decision leaves. The random forest model then collects those independent trees and their leaves and predicts results for new observations. New observations get sorted into resultant leaves depending on their own features, and their probabilities of being predicted as a certain class depend on the other data points sorted into that leaf in the trained model. In a simplistic model, imagine that the only decision is that PCA1 $> .5$ and that all observations with PCA1 $> .5$ are high in cultural fit. Then, a new observation whose PCA1 $> .5$ would also get sorted into the same leaf and would then be classified as high cultural fit.

The fourth step was to evaluate the trained model. To do so, we assessed the model's predictions compared to the original continuous values. Random forest models produce, along with the classifications of input, probabilities of observations belonging to each class. Conceptually, this means that if an observation has certain characteristics that correspond to a given class, it will have a higher probability of being in that class. For example, if an individual's email communication has indicators of low, medium, and high cognitive cultural fit, but more indicators of high cultural fit than the others, then his or her output from the random forest model might indicate a 0.2 probability of low fit, a 0.3 probability of medium fit, and a 0.5 probability of high fit. We can then take a weighted sum of these probabilities to generate a measure that is conceptually analogous to the original continuous measure. We used a mix of methods to evaluate the model, including the area under the curve of the receiving operating characteristic curve (ROC AUC), precision-recall, and separation between low and high cognitive cultural fit with respect to the original continuous values. As reported in Appendix C, the final models we used performed well on these evaluations.

The final step was to impute perceptual congruence and value congruence using their corresponding random forest models for all individuals in all time periods for which we had corresponding email data. To do this, we followed the first step above to retrieve the input feature vector for each individual over time and used all the linguistic data for each individual up to a certain month to impute perceptual congruence and value congruence for that individual in that month.

There were a total of over five million unique emails. Each email can be sent from an individual and several other individuals (via the to/cc/bcc lines). We included both messages sent to and received from the focal individual in our final model.

**Dimensionality Reduction of Features** Considering the size of our potential feature vector, we used dimensionality reduction techniques to make our process computational tractable. In particular, we used a discriminative heuristic to determine which n-stems to keep, since there is a tradeoff between keeping frequent and non-frequent terms: frequent terms allow for discrimination to the extent that they are used differently among a large population of people, while non-frequent terms allow for discrimination to the extent that some people use them and others do not. Given this trade-off, we retained those n-stems that were used by at least 99% of all employees, regardless of their objective frequency. To retain as much information from this pared down set of n-stems, we used principal component analysis (PCA). This allowed us to reduce the hundreds of thousands of features to only a few thousand per n-stem, while still retaining a large part of the variance of the original data. Because of the exponential size of the "to" stems compared to the "from" stems, we ended up using the top 3,000 PCA components from the "from" uni-, bi-, and tri-stems, and from the "to" uni-stems.

**Random Forest Model Specification** We selected the random forest model because of several favorable characteristics. First, random forest models allow for nonlinear relationships between input and output. Decision trees in general, of which random forest is a collection, thus allow for arbitrarily complex relationships, which we would assume govern the relationship between linguistic data and cognitive cultural fit. Second, random forests are ensembles of decision trees, which inherently reduce overfitting and increase robustness. Since there is the potential for a link between linguistic data and cognitive cultural fit to be extremely idiosyncratic (e.g., use of a certain phrase or way of communicating), it greatly helps that we use a more robust method. Third, random forest models do not require as much training data as neural networks. Deep neural networks have the same, if not better, ability to pick up complex relationships, but require far more training data, depending on the depth of the model. As a result, random forest models are simpler and tend to require fewer training data for comparable results.

We split the data into the usual training, development, and testing sets, with 56% of the original data in the training set, 14% in the development set, and 30% in the testing set. Because of the way the random forest algorithm is implemented, it is strongly vulnerable to the "class imbalance" problem. Specifically, if the input to the model from the training set were 10% class 0, 80% class 1, and 10% class 2, then the model would err towards predicting most new observations as class 1. To overcome this, we used a bootstrapping procedure that randomly samples with replacement the lesser classes until they reach the amount of the most populated class. This procedure ensured that, on average, input classes were balanced and therefore class prediction depended more on the splits than on the original balance of the input classes. In addition to searching the hyperparameter space, we also tested varying N for bootstrapped samples.

**Test Set Metrics**

Because of the way we constructed our pseudo-continuous imputed cultural fit, we needed to use a set of test metrics that accurately capture what it means to have a "good model." The choice of bounds for the continuous to discrete distribution is forced; it is an educated guess that produces empirically validated results. Therefore, observations that lie just on one side may not differ substantively from observations that lie just on another side. Concretely, observations that are on the high end of the medium cultural fit may be very similar to observations that are on the low end of the high cultural fit, given that we had set the cutoff ourselves. Therefore, our measures should focus less on perfect categorization (i.e., precision, recall), and more on separation of low and high cultural fit and predictive power of imputed results on actual results. As a result, our performance metrics are a mix of the traditional machine learning metrics, as well as novel metrics we developed ourselves.

For the traditional test metrics, we present the pairwise precision and recall measures on the test set. We provide the pairwise precision recall rather than an F score, because we differentially care about the pairwise results. That is, we care the most about the precision recall between the high and the low cultural fits and less about the precision recall between the mid and either high or low cultural fits, as per our previous discussion.

[TABLE B2 ABOUT HERE.]

A better metric might be to directly examine the separation between groups. If we link the original continuous values with the classifications, then we would see a split like in the figure below.

[FIGURE B1 ABOUT HERE.]

We then used the means and standard deviations of each group to see if the classifier successfully split the observations into statistically distinct groups. We find that the models appear to appropriately distinguish between the low and high groups.

[TABLE B3 ABOUT HERE.]

Finally, we used the receiver operating characteristic curve (ROC) that has become popular in machine learning. Since the ROC works with threshold probabilities of classification, mapping the true positive rate versus the false positive rate at different thresholds, it conceptually measures the extent to which the rank-ordering of predicted values is in line with expectations. For a perfect area under the curve (AUC), the rank-ordering would be monotonically increasing such that all actual values of 1 would have higher probabilities of being classified as 1 than all actual values of 0, and vice versa. Since we have three classes versus the regular binary classification, we use the micro-averaged ROC curve, which takes into account this structure. The ROC curves with their AUC's are presented below.

[TABLE B4 ABOUT HERE.]

# APPENDIX FIGURES

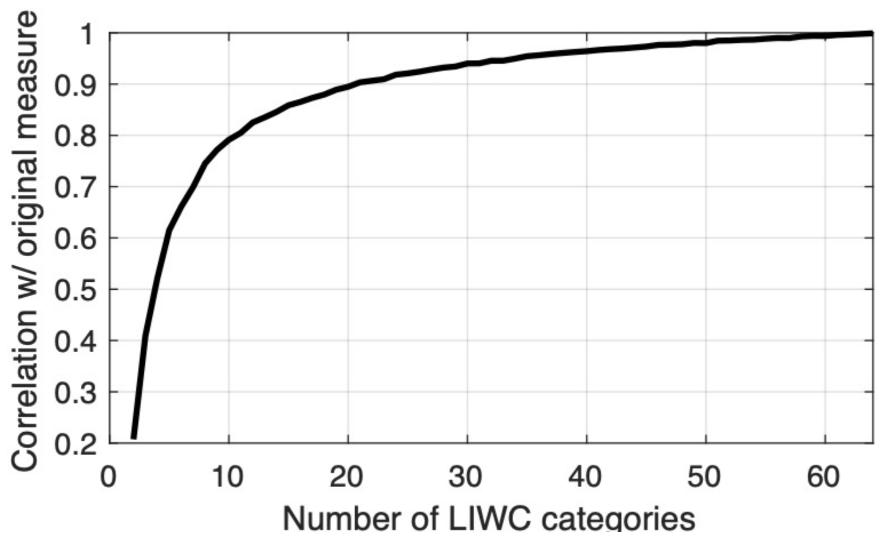


**Figure A1** **Robustness of the linguistic conformity measure to simulated changes in LIWC category composition**

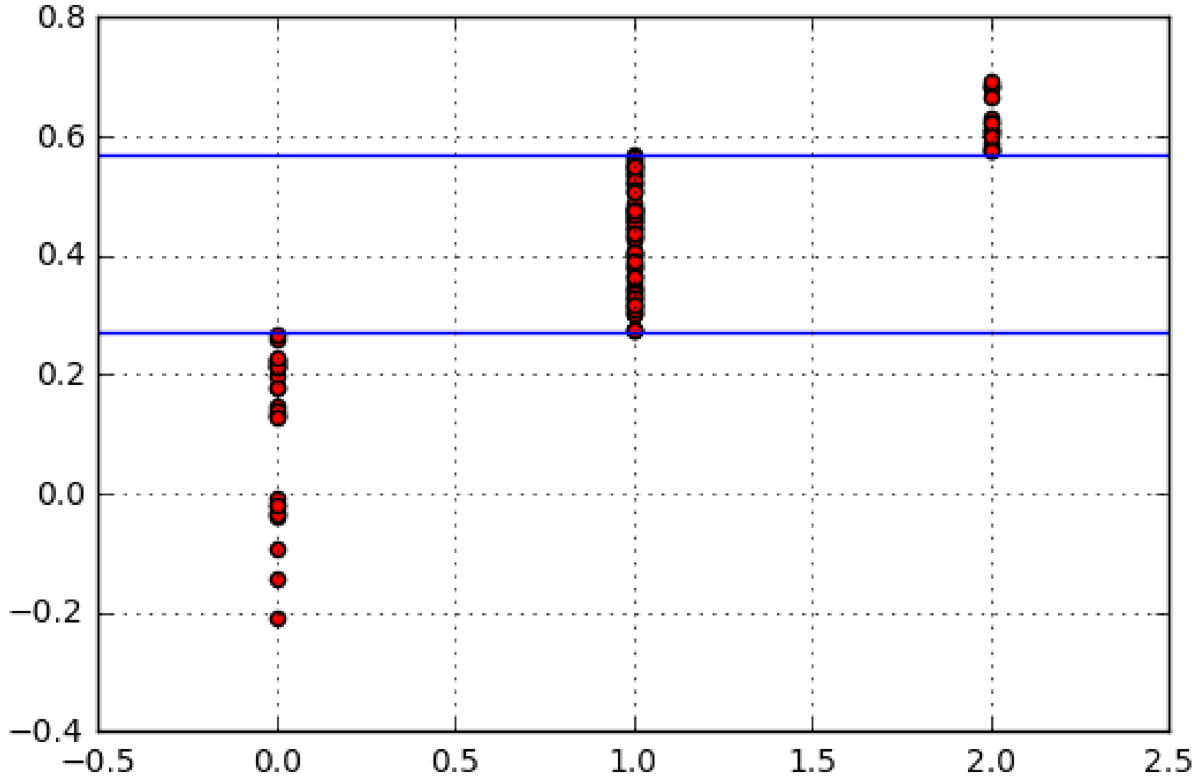


**Figure B1** **Division of Continuous Cultural Fit into Classes**

## APPENDIX TABLES

Table A1: Linguistic Inquiry and Word Count (LIWC)

| Category | Examples | Words In Category |
|---|---|---|
| Total function words | | 464 |
| Total pronouns | I, them, itself | 116 |
| Personal pronouns | I, them, her | 70 |
| 1st pers singular | I, me, mine | 12 |
| 1st pers plural | We, us, our | 12 |
| 2nd person | You, your, thou | 20 |
| 3rd pers singular | She, her, him | 17 |
| 3rd pers plural | They, their, they'd | 10 |
| Impersonal pronouns | It, it's, those | 46 |
| Articles | A, an, the | 3 |
| Common verbs | Walk, went, see | 383 |
| Auxiliary verbs | Am, will, have | 144 |
| Past tense | Went, ran, had | 145 |
| Present tense | Is, does, hear | 169 |
| Future tense | Will, gonna | 48 |
| Adverbs | Very, really, quickly | 69 |
| Prepositions | To, with, above | 60 |
| Conjunctions | And, but, whereas | 28 |
| Negations | No, not, never | 57 |
| Quantifiers | Few, many, much | 89 |
| Numbers | Second, thousand | 34 |
| Swear words | Damn, piss, fuck | 53 |
| Social processes | Mate, talk, they, child | 455 |
| Family | Daughter, husband, aunt | 64 |
| Friends | Buddy, friend, neighbor | 37 |
| Humans | Adult, baby, boy | 61 |
| Affective processes | Happy, cried, abandon | 915 |
| Positive emotion | Love, nice, sweet | 406 |
| Negative emotion | Hurt, ugly, nasty | 499 |
| Anxiety | Worried, fearful, nervous | 91 |
| Anger | Hate, kill, annoyed | 184 |
| Sadness | Crying, grief, sad | 101 |
| Cognitive processes | cause, know, ought | 730 |
| Insight | think, know, consider | 195 |
| Causation | because, effect, hence | 108 |
| Discrepancy | should, would, could | 76 |
| Tentative | maybe, perhaps, guess | 155 |
| Certainty | always, never | 83 |
| Inhibition | block, constrain, stop | 111 |
| Inclusive | And, with, include | 18 |
| Exclusive | But, without, exclude | 17 |
| Perceptual processes | Observing, heard, feeling | 273 |
| See | View, saw, seen | 72 |
| Hear | Listen, hearing | 51 |
| Feel | Feels, touch | 75 |

continued . . .

Table A1 (continued)

| Category | Examples | Words In Category |
|---|---|---|
| Biological processes | Eat, blood, pain | 567 |
| Body | Cheek, hands, spit | 180 |
| Health | Clinic, flu, pill | 236 |
| Sexual | Horny, love, incest | 96 |
| Ingestion | Dish, eat, pizza | 111 |
| Relativity | Area, bend, exit, stop | 638 |
| Motion | Arrive, car, go | 168 |
| Space | Down, in, thin | 220 |
| Time | End, until, season | 239 |
| Work | Job, majors, xerox | 327 |
| Achievement | Earn, hero, win | 186 |
| Leisure | Cook, chat, movie | 229 |
| Home | Apartment, kitchen, family | 93 |
| Money | Audit, cash, owe | 173 |
| Religion | Altar, church, mosque | 159 |
| Death | Bury, coffin, kill | 62 |
| Assent | Agree, OK, yes | 30 |
| Nonfluencies | Er, hm, umm | 8 |
| Fillers | Blah, Imean, youknow | 9 |

Accessed on May 8, 2015 from `http://www.liwc.net/descriptiontable1.php`

**Table B2 Test Set Precision-Recall Metrics for Imputations**

| | Precision Low-High | Precision Low-Mid | Precision Mid-High | Recall Low-High | Recall Low-Mid | Recall Mid-High |
|---|---|---|---|---|---|---|
| Perceptual Congruence | 0.857 | 0.726 | 0.767 | 0.267 | 0.651 | 0.711 |
| Value Congruence | 1 | 0.952 | 0.950 | 0.667 | 0.952 | 0.934 |

**Table B3 p-Values for Difference in Means between Low and High**

| | P-Value |
|---|---|
| Perceptual Congruence | $2.661e{-3}$ |
| Value Congruence | $8.500e{-6}$ |

**Table B4 Areas under the ROC Curve**

| | ROC AUC |
|---|---|
| Perceptual Congruence | 0.740 |
| Value Congruence | 0.950 |