Shades of Gender in Employee Discourse

Heather A. Haveman and William Rathje[1]
University of California, Berkeley

25 October 2023
[word count 200 for abstract, ~14,800 for main text + refs, ~3,900 for appendix + refs]

Abstract

In many organizations, ideal workers are conceived of as male, which disadvantages female employees. To investigate this phenomenon, we integrate cultural theory, structural linguistics, and natural-language processing to capture shades of gender in cultural conceptions of workers and organizations. We analyze large-scale data on employee discourse about organizations and use word embeddings to extract a gender axis in semantic space. We study tech firms, which have highly masculine cultures, and analyze associations in tech-worker discourse between the gender axis and cultural constructs related to gender. We find that discourse about tech firms is sometimes "degendered": the stereotypically male traits independence and leadership competence appear gender-neutral, while instrumental competence appears female-shaded. Although discourse about tech firms is generally male-shaded, there is considerable variation across employees and firms. Male employees, less-satisfied employees, and those in privately held and smaller firms use more male-shaded language; language-use differences between men and women are wider among less-satisfied employees and those in publicly traded and larger firms. Our approach to quantifying shades of gender in organizational cultures moves us closer to determining how those cultures promote or reduce inequality and exclusion. It also points the way to quantifying other dimensions of organizational culture content at scale.

[1] We thank Glassdoor.com for access to the data, and Sameer Srivastava for facilitating an introduction to the firm. For funding, we acknowledge the Center for Equality, Gender, and Leadership at the Haas School of Business. For sharing their R code and answering our questions about how best to implement it, we thank Marshall Taylor and Dustin Stoltz. For comments, we thank the Macro Research Group and participants at the Berkeley Culture Conference. Minori Jaggia, Emily Guo, and Priyanka Krishna contributed to the paper through their programming skills, as did as Lily Bhattarjee, Grace Kull, and Sandeep Sainath on earlier parts of the analysis.

Gender equality and inclusion at work has long been a goal for activists and governments. But despite activists mobilizing for women's rights, legislators passing statutes mandating equal opportunity for women, regulators and courts enforcing those statutes, and employers developing policies and procedures to comply with the law, gender inequality and exclusion persists in many workplaces. This is especially true in sectors like information technology (Neely et al. 2023) and in management jobs (Haveman and Beresford 2012).

A common explanation for gender inequality and exclusion is that cultural conceptions of jobs, workers, and organizations are culturally "shaded" (or "coded" or "typed") as male (Kanter 1977; Acker 1990; Heilman 2012) – meaning that beliefs about gender influence organizational norms, structures, practices, and interactions.[2] Organizational staff usually assume the "ideal worker" is male when they arrange tasks into jobs and workers into teams, describe jobs in recruiting sessions and job postings, design policies and procedures, and devise evaluation and promotion criteria (Acker 1990; Williams 2000). The resulting male-shaded conceptions of jobs, workers, and organizations are reflected in and reinforced through everyday interaction.[3] Male-shaded cultural conceptions of jobs, workers, and organizations raise questions about how well women fit into organizational cultures and whether they are competent in many jobs. Because male-shaded cultural conceptions of jobs value men over women, women are less likely to apply for male-typed jobs (Gorman 2005). When they do apply for male-typed jobs, women are less likely to be hired, less likely to feel comfortable and supported in male-shaded organizations, less likely to be evaluated positively and promoted, and more likely to exit (Blair-Loy 2003; Gorman 2005; Wynn and Correll 2018).

We focus on the tech sector: industries with computing at their core, such as hardware, software, and video games. Tech products are used in all sectors of the world economy,

---

[2] Throughout the paper, we use the terms "gender-shaded," "gender-coded," and "gender-typed" interchangeably to refer to how beliefs about gender influence how people understand and evaluate the firms where they work and they jobs they hold, and how they interact with others and design organizational routines and policies.

[3] Most previous research has argued that conceptions of jobs, workers, and organizations are male-shaded. Our analysis allows for the possibility of female-shaded and gender-neutral conceptions as well.

including consumer services, education, finance, healthcare, government, manufacturing, professional services, and retail and wholesale operations. The tech sector accounts for 10.5% of U.S. GDP, despite employing just 7.9% of the workforce (Zippia 2023). Tech jobs pay well because many workers are engineers and scientists, and offer good opportunities for upward mobility because many firms grow rapidly. But tech has persistent gender problems, with women underrepresented and with complaints about cultures and practices that disadvantage and exclude women (e.g., Fowler 2017; Chang 2018; Wynn and Correll 2018). Despite these problems, we expect to see variation in the gendering of cultural conceptions of tech firms because they differ in many ways – e.g., in terms of size, age, ownership, product portfolio, and location – all of which may influence the gender shading of workplace cultural conceptions.

To investigate gender-shaded cultural conceptions, we focus on the language employees use to describe their workplaces. Language expresses cultural conceptions because it has shadings of beliefs and values baked into it (Barley 1983). For example, the language of performance evaluations reveals how gender stereotypes infuse managers' perceptions of workers (Correll et al. 2020). We analyze data on over 900,000 employee reviews of over 36,000 tech firms across the U.S. from Glassdoor.com, an online job-search platform. We develop an unobtrusive indicator of gendered cultural conceptions of workplaces – how much employees use language that is coded male vs. female. Our approach creates a valid and unbiased measure of to what extent discourse is gender-typed, which is difficult to achieve with interviews or standard surveys (Reader et al. 2020), and which can be applied at scale to study many people in many organizations.

Following research in computer science and sociology that aligns structural linguistics, cultural theory, and natural-language-processing (NLP) techniques (e.g., Caliskan et al. 2017; Kozlowski et al. 2019), we train a word-embedding model to project vectors representing word meanings onto a 200-dimension semantic space and use those vectors to derive a gender axis in that space. This axis is a dimension of meaning running from a male pole to a female pole. Measuring semantic relationships between this axis and the language used by employees when they describe their firms allows us to quantify the gendering of that language. We begin by examining semantic relationships between this gender axis and two sets of cultural concepts that are likely to be gender-coded: (1) *gender stereotypes*, such as assertiveness (male-coded) and concern for others (female-coded), and (2) *corporate values*, such as innovation and rationality (both male-coded). Concretely, for each concept, we calculate the association between the gender axis and the word embedding for each word denoting that concept. We also conduct *document-level and firm-level analysis*, calculating how much employee discourse about their workplaces is male- or female-shaded (i.e., where along the gender axis the language in individual reviews lies) and how much discourse about *specific firms* is male- or female-shaded. Last, we leverage the literature on gender at work to analyze which employees in which firms use more male-shaded language.

Semantic associations between the gender axis and most gender-stereotype concepts are as expected, except for three male-stereotype concepts: "independence" and "leadership competence" are gender-neutral, while "instrumental competence" is female-shaded. Words denoting corporate values are largely male-shaded, as expected – but some of those words are gender-neutral and a few are female-shaded. When we shift from analyzing individual concepts to analyzing entire documents, we find that tech-worker discourse generally uses male-shaded language, although how much varies considerably across employees and firms. Employees who are female, less satisfied, and working in privately held and smaller firms tend to use more male-shaded language. The difference between the gender shading of language used by male and female employees (i.e., how much more male-shaded men's language is than women's language) is greater among less-satisfied workers and those in publicly traded and larger firms. These results indicate that employee discourse in the tech sector is "degendered" to some extent: not all firms, even in the tech sector, are bastions of masculinity.

This paper makes two contributions. Methodologically, we use word embeddings to analyze cultural conceptions of entire documents (rather than individual words) and the organizations described in those documents. This method can be generalized to study other social phenomena, including objects, events, activities, individuals, and groups. Substantively, we demonstrate how to create an unobtrusive indicator of gendered discourse and investigate which employees in which firms, use more male-shaded language. Our approach can be used to advance theories of gender inequality by pinpointing the causes or consequences of gendered discourse for employees and firms. It can be extended to other dimensions of culture, such as race, class, and equality vs. hierarchy, forging new directions in the study of organizational culture, which has generally focused on culture strength, rather than content.

Theory

For almost 50 years, research has demonstrated that employing organizations directly influence gender inequality and exclusion (e.g., Kanter 1977; Gorman 2005; Kelly and Moen 2020). Much work has focused on organizational structures (e.g., Baron et al. 1986; Reskin 1993) and policies (e.g., Kelly and Dobbin 2009; Dobbin et al. 2015). Far less studied is culture (but see Gorman 2005; Wynn and Correll 2018) because culture is less tangible than structure or policies. Much of culture is invisible, so it is difficult to observe, especially when it touches on sensitive topics like beliefs about gender. But culture both reflects and reinforces organizational structures and policies, so it merits greater attention, especially in an era when new sources of data (such as large-scale text records and social-media connections) and new computational techniques (such as natural-language processing and large-scale network methods) make it possible to advance theory with new empirical evidence (Merton 1948). Indeed, previous research has shown that, even after considering formal structures and policies, gender inequality persists, often due to cultural factors (Edelman 2016).

We probe the extent to which organizations are gendered – by which we mean that cultural schemas about differences between men and women determine what work means, how work should get done, who can and should do the work, and who should have status and power (Acker 1990; Ridgeway 2011). We begin by discussing cultural conceptions of gender in society at large before narrowing our focus to cultural conceptions of gender in workplaces.

We then explain the importance of language in understanding gendered cultural conceptions. Next, we examine two cultural phenomena that are likely to be culturally gendered: gender stereotypes and corporate values. We end by shifting the level of analysis upward, from concepts (words and phrases) to entire documents and then upward again to firms, to investigate variation in gendered discourse across employees and firms.

Cultural conceptions of gender

Gender is a cultural schema. Like other cultural schemas, gender consists of widely shared (although sometimes contested) clusters of mental associations about objects, individuals, groups, organizations, activities, and events that develop with experience and that provide default associations about the characteristics of objects, individuals, and so forth (Carley and Palmquist 1992; Hunzaker and Valentino 2019; Boutyline and Soter 2021). All cultural schemas, including gender, are inherently relational: they "work" through contrasts between categories. Concepts in cultural schemas and the relationships among them are represented in the ways we think and therefore in the ways we speak and write.

Historically, gender has been understood as a binary, contrasting male with female. Early structural linguists argued that we cannot understand the concept "male" without reference to the concept "female" because they are defined oppositionally (e.g., Lévi-Strauss 1963 [1983]).[4] Binary cultural schemas transform a complex of practices and expressions into simple categories that are cognitively easier for people to navigate in everyday life. These simplified representations are integrated into cognitive schemas and codified in discourse, creating hierarchical, essentialized, and fixed understandings of objects, individuals, groups, organizations, activities, and events, and thus of the interactions among them. For example, in patriarchal societies, men have long been systematically valorized and accorded more

[4] The male/female binary is just one of the many we use to order and navigate the social world; others include young/old, success/failure, and politically right/left. But not all cultural schemas are binary. Some are multidimensional; for example, the common cultural distinction between market, state, and civil society, or different societies' classifications of race. Others involve hierarchies; for example, military ranks.

resources while women have long been marginalized (e.g., Miller 1998). As a result, male-coded traits and behaviors (such as assertiveness or leadership) have been celebrated while female-coded traits and behaviors (such as cooperation and nurturing) have been devalued.

Over time, relationships develop between cultural schemas, conjoining social positions and practices. For example, the Protestant ethic holds economic success to be a sign of virtue and God's favor, which may explain why cultural conceptions of wealth are positively associated with cultural conceptions of success (Kozlowski et al. 2019). Sometimes relationships between cultural schemas are unexpected. For instance, the schema for gender is associated with the schema for body weight. Fat is not just a feminist issue, as the saying goes, but a widely perceived attribute of femaleness because women's bodies tend to be subject to greater scrutiny than men and the socially desirable weight range for women is narrower, so maintaining one's weight within that range is more difficult for women and deviations are more likely to be noticed (Arseniev-Kohler and Foster 2022). Relationships between cultural schemas can vary within societies. For example, for liberals, the poverty end of the affluence/poverty axis is associated with structural factors like racism that are individuals cannot control; for conservatives, it is associated with actions and attributes like laziness that individuals can (and should) control (Hunzaker and Valentino 2019). Relationships between cultural schemas can change over time. For instance, over the twentieth century, as education became increasingly important for securing well-paying jobs, the affluence end of the affluence/poverty axis became more strongly associated with the high end of the high/low education axis (Kozlowski et al. 2019). And the increase in women working outside the home spurred by the feminist movement weakened some associations between the cultural schemas of gender and work (Eagly et al. 2020).

Gendered cultural schemas at work

We focus on cultural schemas at work because workplaces are fundamental determinants of psychological, social, and economic outcomes (Baron and Bielby 1980).

Cultural schemas at work encompass expectations and judgements about organizational goals, the tasks required to achieve those goals, and the qualities of the employees who are best suited to carrying out those tasks. For some organizations, cultural schemas are gendered in a particular way – shaded male – because of perceptions that core tasks are male-typed, so men are the “natural” or “ideal” people to do those tasks (Acker 1990; Williams 2000).

Male-shaded cultural conceptions of what work is and who should do it influence everyday interactions at work and are reinforced by those interactions. Male-shaded cultural conceptions also influence the design of workplace policies and practices, including recruiting, evaluation, and promotion. Male-shaded cultural conceptions raise questions about how well women fit the culture of organizations and whether they are competent in many jobs. In particular, organizations with male-shaded cultural schemas tend to be unfriendly toward and unsupportive of women. For example, a study of elite law firms showed that when selection criteria for new associates included more stereotypically masculine characteristics such as decisiveness and assertiveness (versus stereotypically feminine characteristics such as friendliness and cooperation), fewer female lawyers were hired (Gorman 2005). In essence, then, organizations can be culturally conceived as male-typed if they frame ideal workers as male and core tasks as naturally done by men; this framing raises questions about whether women are competent and fit the culture of those organizations (Kanter 1977; Acker 1990; Williams 2000).

Yet because organizations differ along many dimensions, there is likely to be considerable variation in how much cultural schemas about organizations are gendered – coded male, gender-neutral, or female (Britton 2000). Uncovering and explaining such variation, which has seldom been studied, is our objective. Only when we have a way of comparing how much cultural conceptions of organizations are gendered can we understand the consequences of cultural conceptions of organizations for gender inequality and exclusion.

Using language to capture cultural schemas at work

To probe the norms and values that characterize gendered cultural schemas and drive gender inequality and exclusion at work, we focus on the language employees use. Language reveals employees' cultural conceptions of workplaces because language is a medium for expressing fundamental assumptions about how firms can and should operate (Barley 1983; Van Maanen 1991; Corritore et al. 2020). Thus language reveals employees' understandings of their workplaces – their mental models (Carley and Palmquist 1992) or cognitive schemas (DiMaggio 1997; Hunzaker and Valentino 2019). Mental models expressed in language allow people to categorize the world around them in symbolic terms that "have meaning, are cues to behavior, and organize behavior" (Stryker 1980: 56). For instance, the language used in job posting signals to prospective applicants whether women are expected to fit those jobs (Gorman 2005), while the language used in performance evaluations reveals the extent to which gender stereotypes influence managers' perceptions of workers (Correll et al. 2020).

Analyzing language allows us to create an unobtrusive indicator of workplace cultural schemas, one that is not prompted by researchers' actions. Unobtrusive measures are especially useful for capturing elements of culture that are socially contested or sensitive, such as shades of gender (Reader et al. 2020). It is very difficult to study such topics using interviews or surveys because of social-desirability and demand effects. In contrast, when employees volunteer to describe their workplaces, especially in an anonymous forum, they are more likely to "tell it like it is," even when they bring up contested or sensitive topics.

We take into consideration how two aspects of language, content and context, reveal gendered cultural conceptions. The *content* of language indicates what matters: words and phrases that are common are associated with central cultural elements; those that are rare with peripheral (Whorf 1956; Sapir 1958). That is not to say that all central cultural elements are valued; indeed, some may represent common expectations or practices that are disdained or contested. If so, they will be prevalent *and* take the form of complaints about the prevailing culture or practices. For example, following Covid-19, many supervisors have required employees to return to working onsite rather than remotely, but many employees oppose this. So terms like "remote work" and "work from home" may be central to today's workplace discourse but contested.

The cultural meaning of language also derives from its *context* – i.e., nearby grammatical elements such as words and punctuation marks (Harris 1954; Firth 1957). Most germane to our analysis is that the proximity of concepts to words that denote gender (e.g., "he," "woman," "male") determines how gender-typed those concepts are. For example, the presence of words like "engineer" or "nurse" in close proximity to words denoting gender like "he" or "she" signals strongly gendered expectations about occupations.

Gender stereotypes

Gender stereotypes, like all stereotypes, are cultural schemas: widely shared generalizations about men and women that are applied to individuals and can be captured in discourse. In Western cultures, the male stereotype tends to emphasize *agency*; the female stereotype, *communality* (Ridgeway 2011; Heilman 2012).[5] Common understandings of agency have multiple components: assertiveness, independence, instrumental competence, and leadership competence (Hentschel, Heilman, and Peus 2019). *Assertiveness* involves dominance, forcefulness, and boldness. *Independence* involves self-reliance, taking responsibility, and autonomy. *Instrumental competence* involves diligence, hard work, and productivity. *Leadership competence* involves achievement orientation, taking charge, and persuasiveness. Similarly, common understandings of communality have multiple components: concern for others, sociability, and emotional sensitivity (Hentschel et al. 2019). *Concern for others* involves understanding others and being kind. *Emotional sensitivity* involves intuition and attention to one's own and others' feelings. *Sociability* involves a relationship orientation, collaboration, and interpersonal communication.

---

[5] Another conceptualization of gender stereotypes, which distinguishes between competence (male) and warmth (female) (Fiske et al. 2002), largely parallels the one we adopt: competence is a component of agency, while warmth is closely related to the three components of communality.

We consider one final gender-stereotype concept, *rationality*, because the core work done by the tech firms we study is based on science and engineering. These fields prize logic and reasoning, which are constitutive of rationality. Rationality has long been considered a male characteristic (Heilman 2012).

Gender and organizational values

The next step in our analysis of gendered discourse involves analyzing three corporate values: innovation, speed, and performance. *Innovation* helps firms outcompete rivals. Innovation is especially critical in the tech sector, which continually invents new technologies or improves existing ones. That is why tech firms spend more on research and development than other firms (Cohen and Klepper 1992). For example, Microsoft highlights this quotation from a product developer: "Our 'growth mindset' culture lets us try amazing things; we are innovating like crazy right now" https://careers.microsoft.com/us/en/culture, viewed 2023-01-17), while OpenAI's charter declares the firm "must be on the cutting edge of AI capabilities" (https://openai.com/charter, viewed 2023-01-17). Following in the lauded footsteps of Bell Labs (which developed the transistor and the Unix programming language) and XEROX PARC (which developed the graphical user interface and laser printing), many large tech firms fund "innovation labs" such as Amazon Lab126 (which developed the Kindle) that are tasked with "moonshot" research on cutting-edge technologies.

Speed is intimately connected to innovation: coming up with new products and processes is necessary, but firms must do this quickly enough to keep up with or – better – stay ahead of rivals. Being first to market with new or improved products is therefore a common goal. First movers can establish their products as industry standards and create strong brand recognition (Lieberman and Montgomery 1988). And when there are high switching costs, first movers have lasting advantages over rivals. Although first-mover advantages are far from assured, many businesspeople believe they are, based on popular management books that extol being first to market with new products. This belief is endemic to the tech sector;

consider, for example, the motto "move fast and break things" promoted by Facebook founder Mark Zuckerberg.

Finally, all firms care about *performance*. We focus on economic performance, a goal common to all for-profit firms.

We expect innovation, speed, and performance to be more strongly associated with the male end of the gender axis than the female end (i.e., to be male-shaded). Innovation and speed are core values for tech firms, whose jobs and employees tend to be male-typed. And performance is a focus of managers, especially top managers, whose ranks remain male-dominated and which are culturally coded as male (Eagly and Karau 2002).

From concepts to documents and organizations

Most research on cultural schemas in discourse focuses on specific concepts rather than the social phenomena – which could be objects, individuals, groups, organizations, activities, or events – that are culturally coded by or generate discourse invoking those concepts. Above we discussed concept-level analysis, following previous research. But our ultimate objective is to understand the gendering of discourse in employing organizations. To do this, we need to shift the level of analysis upward from *concepts* to *documents*, each of which contains many different concepts, then upward again from documents to *organizations*, which are culturally coded by the discourse distributed over multiple documents. Only by analyzing entire documents can we see how all the concepts expressed in a document combine to create gender-shaded discourse. And only by analyzing all the documents about an organization can we see how discourse about it is gender-shaded. Shifting the level of analysis upward from the concept to the document and then again to the organization[6] allows us to examine both central tendencies *and* dispersion in the gender-shading of organizations in discourse, and thus allows

[6] The same logic applies to other social phenomena, such as objects, individuals, groups, events, or activities, that may be under discussion.

us to investigate associations between cultural conceptions of organizations and features of those organizations.

Feminist scholars originally conceived of all employing organizations as culturally coded male (Kanter 1977; Acker 1990), but later scholars recognized there may be variation across organizations in their gender coding (Britton 2000). In line with this argument, sociologists have found that having more women in senior corporate leadership can "undo" the gendering of firms, leading to less gender segregation at lower levels (Stainback et al. 2016). In the same vein, others have found that gender gaps in hiring and pay are greater in male-dominated STEM (science, technology, engineering, and mathematics) fields like physics than in STEM fields with more equal gender representation like biology (Smith-Doerr et al. 2019). To examine variation in cultural conceptions of organizations, we use entire documents – employee reviews of their firms. This allows us to pinpoint which employees in which organizations perceive (and discuss) those organizations in more male- or female-shaded language, and which organizations' employees have, in the aggregate, more strongly gendered (male-shaded or female-shaded) workplace cultural schemas.

To understand variation in the gender shading of cultural conceptions of organizations, we build on theories of gender at work. We begin with two important individual-level factors: gender and job satisfaction. Concerning *gender*, our baseline hypothesis is that female employees are less likely than male employees to use male-shaded language because women are more likely to be sensitive to gender differences and inequality at work (Chatman et al. 1998). Women's greater sensitivity to gender differences and inequality makes them more likely to use gender-inclusive language, such as referring to managers or other workers as "they" or "he and she," rather than just "he." Additionally, socialization into gender roles and stereotypes makes women more likely than men to discuss female-stereotyped topics and traits, such as collaboration and friendliness, and men more likely than women to discuss male-stereotyped topics and traits, such as independence and rationality. As a result, female employees are likely to use less language that reflects male-coded conceptions of workers and organizations. Hence:

> **Hypothesis 1**: Female employees tend to use less male-shaded language than male employees.

Now consider *satisfaction* with jobs and workplaces, which helps capture inclusion and acceptance. The relationship between satisfaction and male-shaded language depends on local (i.e., workplace) culture and broad societal norms, both of which encode gender schemas. Focusing on workplace culture suggests that the relationship between gendered discourse and satisfaction is positive. Tech employees who use more male-shaded language may be more satisfied with their jobs and workplaces because that language indicates fit with the prevailing male-dominated tech culture, and cultural fit is positively associated with satisfaction (Chatman 1991). For example, tech employees, both male and female, who are more satisfied with their workplaces may be more likely to discuss the male-typed topics that are associated with a male-dominated culture, such as assertiveness and logic. Therefore we predict:

> **Hypothesis 2a**: Employees who are more satisfied with their workplaces tend to use more male-shaded language than employees who are less satisfied.

The strongly masculine cultural orientation of the tech sector, however, deviates from societal norms, as most Americans prefer gender equality (Barroso 2020). Focusing on societal instead of tech-sector norms suggests that the relationship between gendered discourse and satisfaction could be negative. Reflecting the societal concern for gender inclusion, tech employees who are more satisfied, both male and female, may use more female-shaded or gender-neutral language. For instance, driven by societal concerns for gender inclusion, employees who are more satisfied may bring up topics and behaviors that are female-shaded, like collaboration and emotional sensitivity, or use gender-neutral phrasings like “he or she” when referring to coworkers and bosses. Hence:

> **Hypothesis 2b**: Employees who are more satisfied with their workplaces tend to use less male-shaded language than employees who are less satisfied.

Satisfaction may moderate the relationship between employee gender and use of gendered language. Specifically, any observed difference in the language used by male and female employees may be smaller among more-satisfied employees than among less-satisfied employees because more satisfied employees tend to have better cultural fit with their firms (Chatman 1991). Better cultural fit should lead to more similarity in employee discourse (Goldberg et al. 2016), including how gendered that discourse is. Put another way, satisfied employees who "fit" their firm's culture are likely to express themselves in similar ways, but those who are dissatisfied may express themselves in many different ways. Paraphrasing Tolstoy, "all happy employees are alike, but each unhappy employee is unhappy in their own way." Therefore, we predict:

> **Hypothesis 2c**: Any observed difference between male and female employees in the use of male-shaded language is less among more-satisfied workers than among less-satisfied workers.

Next, we consider two important organizational characteristics that are likely to affect the gendering of employee discourse: ownership and size. *Ownership* (public vs. private) is seldom studied because organizations scholars focus mostly on publicly traded firms, on which we generally have better data. In this study, though, we are fortunate to have data on both publicly traded and privately owned firms. Several fundamental differences between publicly traded and privately owned firms may affect the gendering of employee discourse. Most basically, publicly traded firms are more visible because they are legally required to report on their operations. Visibility may induce employees in highly visible publicly traded firms to less use male-shaded language, to conform to societal norms about gender equality and inclusion. Moreover, in response to legal requirements and shareholder expectations, publicly traded firms are more likely to develop formal structures, including gender-equity programs that signal congruence with societal norms about gender equality and inclusion. These programs may influence employees to adopt less male-shaded language. In sum, then, we expect:

> **Hypothesis 3**: Employees in publicly traded firms tend to use less male-shaded language than employees in privately owned firms.

Firm ownership is likely to moderate the relationship between employee gender and use of gendered language. Assuming hypothesis 1 is supported, publicly traded firms' gender-inclusion programs may influence all employees, but *especially* male employees, to adopt less male-shaded language. If so, the discourse of male employees in publicly traded firms will be closer to that of their female counterparts. Therefore:

> **Hypothesis 3a**: Any observed difference between male and female employees in the use of male-shaded language is less among employees in publicly traded firms than those in privately owned firms.

Firm size has fundamental but complex effects on firms. First, larger firms tend to be more formalized than smaller firms, with more written rules and procedures (Blau and Schoenherr 1971). Second, larger workplaces are subject to more legal requirements concerning gender equality; for example, U.S. workplaces with more than 100 employees must report gender-composition data to the federal government. Third, larger firms are more visible than smaller ones (Scott and Davis 2007), which makes them more susceptible to social norms about gender equality and inclusion; if so, they will be more likely to have gender-inclusion policies. Larger firms' visibility and formal rules, procedures, and reporting requirements concerning gender inclusion may influence their employees to use less gender-biased language. Therefore, we predict:

> **Hypothesis 4**: Employees in larger firms tend to use less male-shaded language than employees in smaller firms.

Firm size is likely to moderate the relationship between employee gender and use of gendered language. Assuming hypothesis 1 is supported, larger firms' formal gender-inclusion programs may induce all employees, *especially* men, to use less male-shaded language. This would reduce the difference between male and female employees' language use:

> **Hypothesis 4a**: Any observed difference between male and female employees in the use of male-shaded language is less among employees in larger firms than those in smaller firms.

But considering three other correlates of firm size leads to the opposite prediction. First, as firms grow, jobs and work groups become more different from each other (Blau and Schoenherr 1971). Second, employees of larger firms are more likely to work in different locations than those of smaller firms. Third, because larger firms have more employees, the likelihood that any two employees will interact declines with size. All three factors promote the development of subcultures (Koene et al. 1997), which tend to coalesce around shared social positions, such as personal characteristics, department, and job level (Hofstede 1998; Trice 1984). Gender is the most salient personal characteristic in many workplaces (Ridgeway 2011). Department and job level are often correlated with gender; e.g., women are more likely to work in human resources and men in finance. Therefore, gender-based subcultures are likely to emerge in large firms. These subcultures hold different norms and values and speak different languages – more male-shaded in subcultures dominated by men, less in subcultures dominated by women. Finally, although larger firms are more visible and therefore more likely to develop policies and programs to promote gender inclusion, those tend to be led by women. Therefore, in larger firms, female-dominated subcultures may be more focused on gender inclusion than male-dominated subcultures. This suggests that as firm size increases and gender-based subcultures develop, the difference between male and female employees in the gendering of their language gender may be greater, not less:

> **Hypothesis 4b**: Any observed difference between male and female employees in the use of male-shaded language is greater among employees in larger firms than those in smaller firms.

Methods

Research site: The tech sector

We study tech workplaces: those with computing technology at their core, such as computer hardware and software, networking systems, and video-game firms.[7] Despite

---

[7] We studied nine industries: computer hardware and software manufacturing, electronics manufacturing, enterprise software and network systems, information-technology services, internet-

general evidence of gender inequality and exclusion in the tech sector (e.g., Fowler 2017; Chang 2018; Wynn and Correll 2018), the sector is also highly varied. Tech firms sell many different products and serve many different markets in locations across the country. They vary greatly in age (e.g., Hewlett Packard was founded in 1939, Lyft in 2012), scale of operations (e.g., in 2020, Apple had $275 billion in sales and over 150,000 employees, IEC Electronics $125 million in sales and fewer than 1,000 employees), and ownership (public vs. private). All of these factors are likely to influence the gender shading of employee discourse.

Data sources

Our data consist of 946,653 employee reviews of 36,957 firms in the tech sector from January 2014 to September 2020[8] from Glassdoor.com, an online platform where employees rate their current or former workplace. To rate firms, people fill out a form online. The main components are required open-ended questions about the "pros" and "cons" of workplaces, which elicit employees' descriptions of their jobs, coworkers, and firms. Reviewers can elect to respond to a prompt about "advice to management." The form also contains star ratings (1-5). Overall ratings are required; employees can optionally rate culture and values, diversity and inclusion, work-life balance, senior management, and career opportunities.[9] Figure 1 shows a typical recent review for Salesforce, a San Francisco-based software firm. The bottom half of the figure shows the optional ratings on specific topics.

[Figure 1 about here]

Glassdoor offers several advantages for conducting research on workplace cultures and practices. The company attracts a large group of job seekers – about 64 million unique users every month. Half of job-seekers use Glassdoor (DeMers 2014), so reviews serve as powerful

---

based services, internet-service providers, telecommunications manufacturing, telecommunications services, and video games. We selected those industries because large tech firms in the San Francisco Bay Area are focused on them.

[8] To see if the disruption of workplaces caused by Covid affected our results, we dropped reviews posted after February 2022. As we explain below, the results of that temporal subsample are almost identical to the results of the full temporal sample.

[9] Unfortunately, we do not have access to data on ratings on diversity and inclusion.

signals about potential employers. Glassdoor covers a wide array of firms: large and small, publicly traded and privately held, across the country, in all economic sectors. Glassdoor reviews are anonymous, which limits their susceptibility to bias stemming from fear of employer retaliation (Marinescu et al. 2021). Given its advantages, it is not surprising that Glassdoor reviews have been used in studies of organizational culture (e.g., Schmiedel et al. 2019; Corritore et al. 2020) and employee satisfaction (e.g., Storer and Reich 2021).

Glassdoor reviews have two potential downsides, however. First, rather than being written by a random sample of employees, the data are limited to those who contribute reviews. If people are more motivated to contribute reviews when they have strong emotions, the data might be biased both positively *and* negatively. This would result in extreme bimodal distributions, with many one- and five-star reviews, like in data from platforms like Yelp. But because Glassdoor has a "give-to-get" model that requires users to provide reviews before gaining unlimited access to data (such as interview questions) for job searches, Glassdoor reviews are less likely to be biased in this way (Marinescu et al. 2020) and more representative of the experiences and perceptions of all employees. Indeed, the distribution of overall ratings (1-5 stars) among the reviews in our sample is nearly uniform, with a slight positive skew: the mean is 3.48 out of 5.

Second, reviews may disproportionately represent the perceptions of men. But in our data, there is no evidence of gender bias, as 38% of reviewers identified as male, 19% percent as female, and 42% gave no answer. Thus, of those who reported their gender, 33% were female, close to the 28% of employees in the IT sector who are female.

Measures

We applied natural-language-processing (NLP) techniques to the full corpus of reviews, enabling us to analyze text data at a very large scale: across the country, over seven years, and for over 36,000 firms. People volunteer reviews so they can gain access to Glassdoor data for their own job searches, so it is up to them when to write reviews and what they write. This approach goes beyond most previous research on workplace cultural schemas based on ethnographies (e.g., Barley 1983; Van Maanen 1991), surveys (e.g., Hofstede 1980; O'Reilly et al. 1991), or administrative data like annual reports or web sites (e.g., Nguyen et al. 2019). Ethnographies provide rich details but cover only one or a few organizations. Surveys pick up variation across organizations but they rely on top-of-the head responses prompted by survey designers, which may not mirror on-the-ground reality. Administrative data can also cover many organizations but may be merely symbolic or aspirational. In contrast, our approach captures on-the-ground reality volunteered by employees on their own time, does so in rich detail, allows for variation across organizations, and covers many, many employees in many, many firms.[10]

Capturing the gendering of organizations. We began by measuring the gender shading of employee discourse about organizations using *word embeddings* (Jurafsky and Martin 2023), an unsupervised machine-learning technique whose underlying assumption is relational; i.e., that word context determines meaning (Harris 1954; Firth 1957). This relational conceptualization of meaning is compatible with both structural-linguistics arguments that signifiers like words acquire meaning only through their location in systems of signification (e.g., Saussure 1916 [2011]; Lévi-Strauss 1963 [1983]). It is also compatible with sociological theories that cultural schemas are inherently relational, as they work through contrasts between cultural elements such as norms, values, rituals, and symbols (e.g., Hunzaker and Valentino 2019; Arseniev-Kohler and Foster 2022).

Concretely, word-embedding models map words onto points (represented as vectors) in N-dimensional spaces, where N is typically 100-300 (Mikolov et al. 2013a; Jurafsky and Martin 2023).[11] Words that share many contexts, like "engineer" and "scientist," are located near each

[10] Before we applied NLP techniques, we pre-processed and cleaned the data; e.g., by tokenizing text and removing most punctuation marks. The Appendix details the steps we took.
[11] Word-embedding algorithms can model meanings in texts with millions of unique words using a much smaller number of dimensions because many words have similar meanings and because the algorithms incorporate relationships between meanings. Despite our focus on a handful of meaning dimensions, we need to model many other dimensions to locate the ones we study in linguistic structure.

other in semantic space (i.e., their word vectors point in similar directions), and words that do not, like "engineer" and "day," are located farther away from each other (i.e., their word vectors are orthogonal). Thus these models map semantic locations onto geometric locations.

Because word embeddings capture relational meanings, they can be used in analogical reasoning (Mikolov et al. 2013b). For example, the query "bad:worse::hard:?" should yield the answer "harder." With word embeddings, analogical reasoning takes the form of arithmetic operations on word vectors: vector(bad) – vector(worse) + vector(hard) ≈ vector(harder). Word embeddings' ability to capture relational meanings makes it possible for them to measure the degree to which the meanings of words are associated, including associations between words denoting gender and other concepts – even concepts that are not obviously related to gender, such as class (Kozlowski et al. 2019) or body weight (Arseniev-Kohler and Foster 2022). Such assessments allow us to uncover how the structure of a culture (as expressed linguistically) classifies objects, individuals, groups, organizations, activities, and events along divisions corresponding to binaries like gender.

The semantic relationships quantified by word embeddings between gender and other concepts have been validated by external measures, including implicit association tests, surveys, and neural activity (e.g., Caliskan et al. 2017; Kozlowski et al. 2019; Arseniev-Kohler 2022). For instance, relationships between word vectors for terms denoting gender (e.g., he, she, woman, man) and other concepts (e.g., art, math, job, children) are strongly correlated with the latter concepts' gendered implicit associations (Caliskan et al. 2017). Mirroring widely shared cultural schemas, words related to art and family tend to be associated with women and words related to science and work with men.

To create a gender axis in semantic space, we follow established practice (e.g., Caliskan et al. 2017; Kozlowski et al. 2019; Arseniev-Kohler and Foster 2022) and use antonym pairs that denote the gender binary, such as man/woman, he/she, and male/female. Subtracting each word vector for a male term from the word vector for its antonymic female term yields a new word vector (a resultant). If the meaning of "woman" is equivalent to the meaning of "man," except that they denote opposite genders, then subtracting one vector from the other cancels out all but the gender differences between them (Arseniev-Kohler and Foster 2022). The resultant captures the semantic difference between "man" and "woman," so it can be conceived of as an axis in semantic space running from a male to a female pole. Many different antonym pairs denote the gender binary. The resultants created by subtracting male words from their female antonyms have slightly different semantic orientations (point in different directions in semantic space) because different words denoting gender tend to be used in different ways and in different contexts. To develop a robust measure of the gender axis in semantic space, we took the average across all gender-antonym pairs.

Figure 2a illustrates how we constructed the gender axis. The male pole is on the left, female on the right. Words like "she," "he," "female," and "male" that denote gender are listed along the surface of the semicircle, reflecting the fact that word vectors are normalized so their projections in semantic space are the same length. Solid lines running from the midpoint of the gender axis to the surface represent word vectors. Dotted lines represent resultants for pairs of terms.

[Figure 2 about here]

We analyzed semantic associations between the gender vector, on the one hand, and words denoting gender stereotypes and corporate values, on the other, using cosine similarity, the standard measure of association in high-dimensional vector spaces.[12] This yielded each word's gender polarity – the degree to which it is related to the female vs. male pole of the gender binary (Garg et al. 2018; Kozlowski et al. 2019; Arseniev-Kohler and Foster 2022). Conceptually, this is shown in Figure 2b for several common words in our corpus, such as "opportunity," "team," and "management." Negative cosine similarity scores point toward the male pole; positive scores point toward the female pole. Cosine scores with larger absolute values indicate stronger associations with the gender axis.

[12] This entails calculating the cosine of the angle between two word vectors, which measures their semantic overlap. This measure ranges from 1 (the same) to 0 (orthogonal) to -1 (opposite).

To capture the sensitivity of these associations to sampling (i.e., the specific documents under study), we used bootstrapping methods (Antoniak and Minmo 2018) and established confidence intervals for associations between the gender axis and terms denoting concepts of interest. We generated 20 datasets through random sampling, each containing 90% of the full dataset. We trained a different word2vec model on each subsample. Using each subsample's word-embedding model, we calculated the gender axis and cosine similarity scores between the gender axis and each term of interest. Across the 20 subsamples, we then calculated means and confidence intervals for each cosine similarity measure.

Measuring gender stereotypes. We created lexicons for each of the seven components of gender stereotypes (Hentschel et al. 2019): concern for others, emotional sensitivity, and sociability (female); assertiveness, independence, instrumental competence, and leadership competence (male). We also created a lexicon for rationality, which is especially relevant to the tech sector and is stereotypically male (Heilman 2012). For all concepts, we compiled lists of words. For the first seven concepts, we built initial lexicons starting with the terms listed in Hentschel et al. (2019); for rationality, we started with terms mentioned by Heilman (2012). We developed comprehensive lexicons by adding synonyms of the original terms, synonyms of those synonyms, antonyms of antonyms, and morphemes (different word forms; e.g., adamantly for adamant).

Words denoting cultural concepts have graded relationships with each other – some are more similar in meaning than others (Rosch and Mervis 1975). To ensure that the lexicons we analyzed were semantically coherent (closely related in meaning), we "trimmed" each lexicon by iteratively comparing cosine similarities between each term and all other terms in that lexicon. We did this in stages, starting in each stage by eliminating the word with the lowest average cosine similarity score compared to all other words. We stopped when all remaining words had average cosine similarity scores of 0.50 or more, which corresponds to an angle of 60° in semantic space. Some words were rare; to ensure robust results from bootstrapping, we limited each lexicon to words that appeared in at least 10 of the 20 subsamples.

Measuring corporate values. To generate lexicons of terms associated with concepts denoting innovation, speed, and performance, we followed the same procedure: generating comprehensive lexicons by using synonyms of those three words, synonyms of synonyms, antonyms of antonyms, and morphemes. Again, we ensured that lexicons were semantically coherent, iteratively trimming each lexicon by comparing cosine similarities. Again, analysis was limited to words that appeared in at least 10 subsamples.

Gendered language: from words to documents. Almost all social-science research using word embeddings has analyzed data at the level of the word or phrase (for an exception, see McCumber and Davis 2022). But we want to analyze data at the level of the *document* so we can compare reviews along theoretically and substantively important dimensions such as reviews written by men vs. women or by employees in large vs. small firms. To measure the gender valence of *entire reviews*, rather than individual words, we used *concept-mover's distance,* which can capture semantic associations between cultural binaries, like gender, and entire documents (Taylor and Stoltz 2021a, 2021b) to create *review gender scores*. To explain briefly (the Appendix provides details), this measure is a function of the semantic distances between the words in a document (review) and the words denoting target concepts, as defined by the embedding vectors associated with those words. Reviews closer to the male pole of the gender axis have negative scores; those closer to the female pole have positive scores. This allows us to assess variation in the gender shading of language across reviews.

We used the review gender score as the dependent variable in ordinary-least-squares regressions to test hypotheses about which employees in which firms used more male-shaded (vs. female-shaded) language. To test hypothesis 1, we used *gender*, a dummy variable equal to 1 for employees who self-reported as female and 0 for those who self-reported as male. We also created a dummy variable, *gender unknown*, equal to 1 for the 42% of employees who did not report their gender and 0 otherwise.[13] Our main results, shown below, include only those

[13] Reviews where gender was *not* reported tended to be for smaller firms, whose employees may be more reluctant to reveal their attributes because it is easier to identify them, even though reviews are

reviews where gender was reported. In a robustness check, we analyzed data on all reviews and included the *gender unknown* variable as a control.

To test hypothesis 2, we used the overall star rating of the employing organization, which is available for all reviews. This is measured on a scale of 1 to 5, where 5 represents the highest level of satisfaction. To test hypothesis 3, we created a dummy variable *ownership public*, which equals 1 for firms that are publicly owned and 0 otherwise. About 7% of reviews come from the nonprofit sector (including government agencies and educational institutions) or subsidiaries of for-profit firms (which could be privately owned or publicly traded). For these, we created a dummy variable called *ownership other*, which equals 0 for firms that were publicly traded or privately held and 1 otherwise. Finally, to test hypothesis 4, we used *firm size*, in terms of number of employees. Data on firm size were categorical (1-50, 51-200, … 10,001+); we created three categories, each with approximately the same number of reviews (*small* ≤200 employees, *medium* 201-10,000, *large* >10,000).

We included dummy variables for *industry*, *region*, and *year* to capture similarities among firms and control for non-independence of observations across reviews. Firms in the same industry or region may be similar to each other due to coercive forces (state legal regimes), normative forces (industry standards or regional cultures), or mimetic forces (local role models), so their employees' descriptions of their workplaces may be similar. We measured *industry* using the nine categories listed above. We measured *region* using the U.S. Bureau of Economic Analysis definition (2014), which takes into consideration states' economic, demographic, social, and cultural characteristics. There are eight regions: New England, Mideast, Southeast, Great Lakes, Plains, Southwest, Rocky Mountain, and Far West.[14] When

---

posted anonymously. Indeed, many employees in small firms refused to share not just gender but also job title, reporting instead phrases that signaled concerns about privacy, such as "Can't Say - Only A Few of Us." As we explain below, we conducted a robustness check where we dropped reviews from firms in the bottom size category (≤200 employees) to investigate selection bias in reviews of small firms.

[14] New England includes Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. The Mideast contains Delaware, District of Columbia, Maryland, New Jersey, New York, and Pennsylvania. The Southeast includes Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana,

the state was not recorded, we created an "unknown" region category. Controlling for year captures time-specific similarities between firms. We measured *year* with calendar year.

For a more conservative test of our hypotheses, we estimated models with firm fixed effects to control for differences between firms. Because some of the variables of interest – ownership and size category – do not change over time for the vast majority of firms, we did not estimate standard fixed-effects models. Instead, we estimated "hybrid fixed-effects random-effects" (hybrid-FE-RE) models, also known as "between-within" models (Allison 2009; Vaisey and Miles 2017). We used this estimating equation:

$$y_{ijt} = \alpha_j + \beta_1(X_{ijt} - \mu[X_{ijt}]) + \beta_2(\mu[X_{ijt}]) + \beta_{3z}Z_{ij} + \gamma_1 R_{ijt} + \gamma_2 Y_t + \upsilon_j + \varepsilon_{ijt}\ ,$$

where i indexes individuals, j firms, and t time (year). y_{ijt} is the review gender score for individual i in firm j at time t, X_{ijt} is a set of covariates that vary over time within firms (employee gender and satisfaction level), $\mu[X_{ijt}]$ is a vector of firm-specific means for those covariates, Z_{ij} is a vector of covariates that are (mostly) time-invariant within firms (ownership and size category), R_{ijt} is a vector of region dummies, and Y_t is a vector of time (year) dummies. This model allows each cluster (each firm) to have its own intercept, α_j. It splits the total error into a firm-level error, υ_j, which is assumed to be independent of time-varying covariates, and a review-level error, ε_{ijt}. This modeling strategy treats observations as clustered within firms and estimates robust standard errors for all parameters. The parameter β_1 is equivalent to parameters estimated by standard FE models. The hybrid-FE-RE model yields unbiased estimates of coefficients for time-varying covariates even when those covariates are correlated with the firm-level error, υ_j (Vaisey and Miles 2017).

To ensure robust measures for the means of the time-varying covariates, we limited this analysis to the 5,406 firms that received 25 or more reviews in the focal year. Those firms

---

Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia. The Great Lakes comprises Illinois, Indiana, Michigan, Ohio, and Wisconsin. The Plains includes Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota. The Southwest contains Arizona, New Mexico, Oklahoma, and Texas. The Rocky Mountain region includes Colorado, Idaho, Montana, Utah, and Wyoming. Last, the Far West comprises Alaska, California, Hawaii, Nevada, Oregon, and Washington.

constituted only 14% of the firms in the entire dataset, but they received 83% of reviews. when we eliminated reviews with no data on employee gender, the number of firms dropped to 2,973 and the number of reviews to 44% of the total.

Results (I): Understanding the Gender Axis

Because the idea of using language to capture gender as a cultural axis is new, we conducted validity checks. We began by assessing how each gender-denoting term was related to the gender axis, based on cosine similarity scores. Figure 3 shows the results of this analysis. Words denoting male are blue; words denoting female, pink.

[Figure 3 about here]

As expected, all words denoting *female* were female-shaded. They were all similarly situated semantically: their cosine similarity scores with the gender axis were fairly close, with a coefficient of variation (standard deviation/mean) of 0.34. In contrast, words denoting *male* varied greatly in gender shading; their coefficient of variation was almost triple that of the female words (0.92). Seven of nine male-denoting words were clearly male-shaded, one ("he") was gender-neutral because its 99% confidence interval straddled the zero line, and one ("male") was male-shaded. The gender neutrality of "he" suggests that men were the default employees, suggesting that tech workers tended to assume that others in their workplace were male. So when tech workers said "he," they were talking in gender-neutral terms. In contrast, when tech workers said "she" or "her," they used clearly female-shaded terms, indicating that they expected women to be in female-shaded jobs doing female-typed tasks in female-typed ways. To explore why "boys" and "boy" were so strongly male-shaded, we read through a random sample of 100 reviews that contained those words. Most of the time when tech workers used them, they were talking about "good old boys" or the "old boys club." We show some examples below in Table 2.

We also investigated which words, not just those that denote gender, were most male- vs. female-shaded. The top 30 results are shown in Table 1. The columns on the left focus on male-shaded words; those on the right, on female-shaded words. The 30 most male-shaded words included two of the words used to create the gender axis (“boy” and “boys”); “man” also ranked in the top 100. Notably, “ol,” “ole,” “cronyism” were included, while “club” was also ranked in the top 100. Some words on the top-30 list (“arrogance,” “entrepreneurial,” “trenches,” and “analysis”) fit with the male stereotype components of emotional coldness, initiative, logic, and hard work. But others were not as obviously gendered (“shareholders,” “kool,” “legacy,” and “mentality”), suggesting that they have acquired implicit male cultural associations in tech-worker discourse.

[Table 1 about here]

The 30 most female-shaded words included three of the words used to create the gender axis (“her,” “female,” and “woman”); “women” also ranked in the top 100. Some top-30 words were associated with the stereotype that women belong at home and men at work (“child” was in the top 30, while “kid” and “kids” were lower on the top-100 list, or related women to men (“wife” was ranked #36). Others represented roles and relationships (“coworker”), reflecting the gender stereotype that women at work are more collaborative than men. Overall, these results indicate that the gender axis does indeed capture semantic associations that track the gender cultural binary.

Results (II): Associations between the Gender Axis and Gender Stereotypes

Figure 4 shows projections of the gender-stereotype lexicons onto the gender axis. Figure 4a shows associations between the gender axis and female-stereotype words (three figures); Figure 4b, associations between the gender axis and male-stereotype words (five figures). In each figure, words are arrayed from the most male-shaded to the most female-shaded. For each word, the dot represents the mean across 20 subsample estimates and the ends of the bars represent the 95% confidence interval. Words in pink are female-shaded, in blue male-shaded, and in gray, gender-neutral (their confidence intervals straddle the zero line).

[Figure 4 about here]

We begin with female-stereotype concepts in Figure 4a. The top left panel shows that three-quarters (20/27) of words denoting *concern for others* were female-shaded. Three words were gender-neutral, while four others were male-shaded. The top right panel shows that about half (7/13) of words denoting *emotional sensitivity* were female-shaded. Three words were gender-neutral, while three others were male-shaded. The bottom panel shows that about three-fifths (14/24) of words denoting *sociability* were female-shaded, five were gender-neutral, and five were male-shaded. Together, these results suggest that when tech employees discussed their jobs and firms, they usually aligned all three female stereotype concepts with female-shaded language. Nevertheless, there is evidence that these concepts were not entirely female-shaded, as over one-third of words denoting these concepts were gender-neutral or male-shaded.

Now we turn to male stereotype concepts in Figure 4b. The top left panel shows that just over two-thirds (19/28) of words denoting *assertiveness* were male-shaded. Two words were gender-neutral, while seven were female-shaded. Surprisingly, the top right panel shows that just over half (8/15) of words denoting *independence* were gender-neutral, rather than male-shaded as expected. Only one word was male-shaded, while six were female-shaded. This suggests that when tech employees discussed assertiveness in the context of their workplaces, they conceived of it as generally (but not exclusively) a male trait. But when they discussed independence, they conceived of it as a gender-neutral or slightly female-shaded trait.

The bottom left panel shows that 40% (15/36) of words denoting *leadership competence* were male-shaded. One-quarter (9/36) were gender-neutral and one-third (12/36) female-shaded. The bottom right panel shows that two-thirds (25/38) of words denoting *instrumental competence* were female-shaded, not male-shaded, as most previous research would predict. Only six words were male-shaded, while seven were gender-neutral. Taken together, these results indicate that tech employees conceived of success at leading teams (leadership competence) as something that both men and women did, and getting things done (instrumental competence) as something that women were more likely to do than men. These findings may be surprising, given decades of analysis using experimental subjects to label these terms as male- or female-shaded. But they are congruent with research showing that that in recent years, women were more likely to be judged by experimental subjects as competent (Eagly et al. 2020; Bongiorno et al. 2021). These findings suggest that women's increasing participation in the workforce and representation in management are, to some extent, degendering cultural conceptions of leadership.

We now turn to the concept of *rationality*. Almost 60% (11/19) of words denoting rationality were male-shaded. Among the other words denoting rationality, four were gender-neutral and four female-shaded. This set of results indicates that rationality was mostly, but not exclusively, conceived by employees as a male-typed trait.

These results indicate that gender stereotypes in discourse about jobs and workplaces – even in the tech sector, which is heavily male-dominated – are weaker than expected. Just under half of associations between the gender axis and words denoting gender stereotypes were in the expected direction, while one-third were in the opposite direction. Even the most heavily gendered stereotype concept – *concern for others* – showed the expected associations with the gender axis for only three-quarters of lexicon terms. Notably, *assertiveness* and *leadership competence*, which previous research has found to be male-shaded, were in this analysis gender-neutral, and the concept of *instrumental competence*, also male-shaded in most previous studies, was female-shaded.

Overall, relational language (e.g., responsibility, coordination, advising, friendliness, professionalism) was consistently female-coded, even when describing putatively male attributes. In contrast, individualistic language (e.g., autonomy, achievement) and language denoting hierarchy (e.g., power, dominance, aggressiveness) was usually male-coded. Some results are puzzling – e.g., words denoting power and domination were clearly male-typed, but "assertiveness" was female-typed. These puzzles merit further investigation using other sources of data and other methods.

Results (III): Associations between the Gender Axis and Firm Values

To investigate tech firms' values, we focused on three concepts – innovation, speed, and performance. Figure 5 shows projections of the words denoting these concepts onto the gender axis; it has three parts, one for each concept. In each part, words are arranged from the most male-shaded to the most female-shaded. For each word, the dot represents the mean across 20 random subsample estimates; the ends of the bars represent 95% confidence intervals. Words in pink are female-shaded, blue male-shaded, and gray gender-neutral.

[Figure 5 about here]

The top left panel shows that almost all words denoting *innovation* (17/19) were male-shaded; two were gender-neutral. This suggests that in tech firms, innovation was a masculine value: most likely to be perceived as what men did. The top right panel shows that just over half of words that denote *speed* (8/15) were male-shaded, while just under half (7/15) were gender-neutral. This indicates that as a corporate value, speed was only sometimes conceived of as masculine. Men were perceived as a little more likely than women (or anyone) to achieve rapid results. The bottom panel shows that three-quarters of words denoting *performance* (14/18) were male-shaded. One word ("accomplishment") was gender-neutral and three ("achievements," "accomplishments," and "succeeded") were female-shaded. This suggests that tech-firm employees usually, but not always, conceived of performing well as masculine. Performance, a general corporate goal, was largely but not completely male-shaded. It certainly was more "degendered" than the values of innovation and speed, which are central to tech firms.

Results (IV): Review-level Analysis

In this section, we first explain how we validated the gender score at the review level. We then discuss the review-level analysis: the distribution of review gender scores and regression models predicting those scores. After that, we discuss the firm-level analysis.

Validating the review gender score

To validate this measure, we read hundreds of reviews. To provide examples without exhausting readers, Table 2 shows 10 reviews with the lowest scores (most male-shaded, Table 2a) and 10 with the highest scores (most female-shaded, Table 2b). Male-shaded reviews tended to use stereotypically male-shaded language, such as “disruptive,” “cut throat,” “upper management,” “calls the shots,” and “innovation.” They also tended to focus on top-level management (“The Board,” “the founder / CEO,” “divisional leadership,” “upper management”). Interestingly, masculinity was associated with hierarchical structures, rigid chains of command, red tape, and slowness; e.g. “Matrix org can slow down execution,” “Bureaucracy can sometimes get in the way,” “Too many silos in the company,” “Old guard divisional leadership hinders innovation and growth.” Female-shaded reviews, by contrast, tended to include stereotypically female-shaded language; e.g. “Friendly office environment,” “Nice environment,” “Great coworkers,” “Treated like family, friendly, helpful people.” Female-shaded reviews differed from male-shaded reviews by emphasizing egalitarian structures and work environments. And they often discussed low-power temporary, intern, or executive-assistant jobs.

[Table 2 about here]

We also conducted most-similar-word queries; i.e., searches for words closest in semantic space to words denoting gender. We manually assessed whether it makes sense for associated words to be in close proximity to gender-denoting words. The results of that validation test, which are shown and discussed in the Appendix, reinforced our conclusion that the review gender score captures the gender shading of employee discourse.

Analyzing review-level gender scores

Figure 6 shows the distribution of review gender scores. As expected, the mean of this variable was negative (-0.042), indicating that the typical review was written in male-shaded language. But there was considerable variation: the minimum (most male-shaded) was -0.147, the maximum (most female-shaded) 0.127. These statistics indicate that tech firms were generally bastions of strong masculine cultures, as expected given academic and journalistic research (e.g., Chang 2018; Wynn and Correll 2018) and complaints by tech veterans (e.g., Fowler 2017). But they also reveal that employee discourse about tech firms had varying shades of gender: most was male-shaded, but a little was gender-neutral or female-shaded. Finding variation in the gendering of cultural conceptions of tech firms is congruent with feminist ideas about the gendering of organizations as a variable, not a constant (e.g., Britton 2000) and with empirical work showing that cultural gender shading varies across organizations (e.g., Stainback et al. 2016).

[Figure 6 about here]

Table 3 shows the main regression results predicting review gender scores, based on reviews where gender was reported. This table does not show fixed effects for industry, region, or year; those coefficients are shown in Table A3 of the Appendix. The dependent variable, gender score, is positive for female-shaded language and negative for male-shaded language. None of the variance-inflation factors were very high, so multicollinearity is not an issue, except of course for interactions.

[Table 3 about here]

Models 1 through 4 show the main effects of gender (female=1), satisfaction (overall star rating, 5=highest), ownership (public or other vs. private), and firm size (number of employees; medium or large vs. small). Model 1 supports hypothesis 1: the language used by female employees was more female-shaded (i.e., less male-shaded) than the language used by their male counterparts. Based on the coefficients in model 1, female employees' gender review scores were 7.7% higher (more female-shaded, less male-shaded) than those of their male counterparts. Model 2 supports hypothesis 2b rather than hypothesis 2a: the language used by employees with the highest satisfaction ratings (5 stars) was 1.3% more female-shaded than the language used by employees with sample-average (3.5 stars) satisfaction ratings. Model 3 supports hypothesis 3: the language used by employees in publicly traded firms was 6.7% more female-shaded than the language used by employees in privately owned firms. Model 4 supports hypothesis 4: the language used by employees in large firms was more female-shaded than the language used by employees in small firms (11% more) or medium-sized firms (6.6% more).

Models 5 through 7 add interactions with gender. In model 5, the coefficients on gender and satisfaction remained significant, and the coefficient on their interaction was negative and significant. The difference between male and female employees in the use of male-shaded language was 0.8% narrower among employees who rated their satisfaction at the maximum (5 stars) than among those with sample-average satisfaction ratings (3.5 stars) – small in magnitude but supporting hypothesis 2c. In model 6, the coefficients on gender and public ownership remained positive and significant, and the coefficient on their interaction was positive and significant, which fails to support hypothesis 3a. The difference in language use between among male and female employees in publicly traded firms was narrower that than between those in privately held firms, contrary to our expectation. In model 7, the coefficients on gender and both firm size dummies remained positive and significant, while the coefficient on their interactions were positive and significant. The difference between language use was 1.6% wider between male and female employees in medium-sized firms than between those in small firms and 2.0% wider among male and female employees in large firms than between those in small firms. These results support hypothesis 4b rather than hypothesis 4a. Finally, model 8 includes all variables and shows that all effects (except for public vs. private ownership) were robust.

Analyzing review gender scores at the firm level

Figure 7 shows the distribution of firm-average review gender scores. As in the review-level analysis, the mean was negative (-0.045). There was much less variation across firms than across reviews, as expected when aggregating data. Different from the review-level analysis, no firm-average gender score was gender-neutral or female-shaded. Among the firms with the most male-shaded discourse were a firm in Missouri that sells public-safety and incident-reporting software to police forces (a highly male-cultured sales target) and a Boston-based cybersecurity firm (a male-dominated specialties). Among the firms with the least male-shaded discourse were a firm in Iowa that develops user-experience software and an Indiana-based wireless internet provider (both less male-dominated specialties). Among the firms with the mean gender review score was a cloud software firm in Pennsylvania.

[Figure 7 about here]

We conducted a second regression analysis to understand variation within firms, using hybrid-FE-RE models. These results are shown in Table 4. As explained above, we limited the analysis to firms with 25 or more reviews in the focal year in order to yield robust estimates of firm means for time-varying variables (gender and satisfaction). These results are net of fixed effects for region and year, which are not shown to save space. For time-varying variables, the coefficients on the mean-deviated covariates, which are labelled "FE," are equivalent to the coefficient on those variables from a standard FE model. The coefficients on variables that rarely varied over time within firms (ownership and size category) must be treated with the same caution as the results in Table 3.

[Table 4 about here]

Model 1 shows that the language used by female employees was 7.7% more female-shaded than the language used by their male counterparts, bolstering support for hypothesis 1. Model 2 shows that more-satisfied employees also used more female-shaded language, further supporting hypothesis 2b over 2a. The language used by employees that rated their satisfaction at the maximum (5 stars) was 3.4% more female-shaded than the language used by employees that rate their satisfaction at the mean (3.5). Model 3 again supports hypothesis 3: the language used by employees in publicly traded firms was 4.8% more female-shaded than the language used by employees in privately owned firms. Model 4 again supports hypothesis 4: the language used by employees in large firms was more female-shaded than the language used by employees in small firms (14% more) or medium-sized firms (6.4% more). The magnitudes of most coefficients are similar to those in Table 3; the only exception is satisfaction; its magnitude in Table 4 was almost three times as large as in Table 3.

Models 5 through 7 add interactions with gender. In model 5, the FE coefficients on gender and satisfaction remained significant, and the FE coefficient on their interaction was negative and significant. The difference between male and female employees in the use of male-shaded language was 1.9% narrower among employees who rated their satisfaction at the maximum (5 stars) than among those with sample-average satisfaction ratings (3.5 stars), supporting hypothesis 2c. In model 6, the coefficients on gender (FE) and public ownership remained positive and significant, and the coefficient on their interaction (FE) was positive and significant. This fails to support hypothesis 3a: the difference between male and female employees in language use among employees in publicly traded firms was wider, not narrower. This finding parallels that shown in Table 3. In model 7, the coefficients on gender (FE) and both firm size dummies remained positive and significant, while the coefficient on their interactions (FE) were positive and significant. The difference between male and female employees' language use was 1.4% wider among employees of medium-sized firms than those of small firms and 2.2% wider among employees of large firms than those of small firms, supporting hypothesis 4b rather than hypothesis 4a. Again, these results mirror those of Table 3, increasing our confidence in those findings.

As explained in the Appendix, we conducted four robustness checks. We re-estimated models in Table 3 including reviews where employees did not report gender, adding a dummy variable for *gender unknown*. We dropped reviews from employees in small firms, where privacy concerns make employees less likely to report their gender. We dropped reviews of firms that received the highest satisfaction rating (5 stars) because those may have been biased by the coaching of managers. And we dropped all reviews posted after the end of February 2020, when Covid began to hit U.S. workplaces and many white-collar jobs went remote. The results of these alternative analyses, which are discussed in the Appendix, are all very similar to those shown in Table 3, bolstering our confidence in Table 3.

Discussion and Conclusion

We investigated one subtle but powerful attribute of employing organizations that contributes to gender inequality and exclusion: cultural conceptions of workers, jobs, and organizations as male. Such male-shaded cultural conceptions cause employees to question whether women fit the culture of organizations and whether they are competent in many jobs. As previous research has shown, male-shaded conceptions make it harder for women to thrive at work and drive them to exit (e.g., Blair-Loy 2003; Gorman 2005; Wynn and Correll 2018). We focused on the tech sector because academics, tech veterans, and journalists have all revealed that tech has a gender problem in terms of demographics, practices, and culture (e.g., Fowler 2017; Chang 2018; Wynn and Correll 2018). Yet because tech firms differ greatly in many ways, we expected to see variation in the gendering of their cultural conceptions.

To probe how much cultural conceptions of tech firms are male-shaded, and for which employees and which firms, we applied word embeddings to employees' descriptions of their firms from Glassdoor.com. Word-embedding models map words onto points, represented as vectors, in semantic space, yielding insights into employees' perceptions of their firms' cultures. We leveraged word vectors' ability to capture semantic similarity, subtracting the vector for each word denoting "male" from the vector for its antonym denoting "female" to create a series of vectors mapping onto the binary concept of gender. We calculated the mean of these gender vectors to derive a robust measure of the gendering of discourse: the *gender axis*, which runs from the male pole of the gender cultural binary to the female pole. To validate this measure, we probed associations with words denoting gender; we also investigated associations with common words in our corpus.

We then investigated relationships between the gender axis, on the one hand, and gender stereotypes and tech-firm values, on the other. Most associations with words denoting gender stereotypes were as expected, but some were weak. Associations with three male-stereotype concepts were surprising: independence and leadership competence were largely gender-neutral; instrumental competence, largely female-shaded. Concepts denoting key firm values (innovation, speed, and performance) were largely male-shaded, as expected. But there was evidence for all the concepts being a little "degendered" because for each, some associations with the gender axis were either gender-neutral or female-shaded. Overall, words denoting relationships (e.g., responsibility, coordination, friendliness, professionalism) were consistently female-coded, even when they denoted putatively male attributes. In contrast, words denoting individualism (e.g., autonomy, achievement) and those denoting hierarchy (e.g., power, dominance, aggressiveness) were usually male-coded. Some results are puzzling – e.g., words denoting power and domination were male-typed, but "assertiveness" was female-typed – which merits further investigation using other sources of data and other methods.

When we shifted the level of analysis upward from words to documents and then to organizations, we found that, as expected, employee descriptions of tech firms were generally male-shaded, although there was considerable variation. We estimated regression models to test hypotheses about which employees, in which firms, were more likely to use male-shaded vs. female-shaded language, net of industry, region, and year. All hypotheses except one were supported (excluding opposing hypotheses). Female employees, employees who were more satisfied with their workplaces, and employees in publicly owned and large firms were all less likely to use male-shaded language. Differences in language use between male and female employees were greater among more-satisfied workers and workers in large (vs. small) firms. But, contrary to predictions, the difference in language use between male and female employees was greater, not less, among workers in publicly traded (vs. privately held) firms. All results held when we shifted the level of analysis upward again, from documents to firms, by including firm fixed effects.

Limitations. This analyses revealed intriguing associations between the characteristics of employees and employing organizations, on the one hand, and the gender cultural schema, on the other, but they could only suggest, not demonstrate, causation. Identifying causal mechanisms requires other methods, such as field experiments or instrumental variables, or exogenous shocks that create natural experiments. Consider, for example, a tech firm that experiences a gender-discrimination scandal and loses an expensive and highly publicized lawsuit. That shock may be strong enough to shift cultural mores and change how its employees talk about their workplaces, which could be investigated using the approach we used above, adding into the analysis information about the scandal and lawsuit.

Our analysis focused on the tech sector because it has well documented gender problems, in terms of demographics, culture, and practices. It would be useful to compare other sectors that have traditionally been more male-dominated, like finance and transportation, with those that have traditionally been more female-dominated, like retail and education. It would also be useful to tap into data sources over longer periods of time to capture secular trends. Glassdoor was founded in 2007, and there were few reviews – not enough for robust analysis – until several years later. Probing annual reports of publicly traded firms and industry-association documents could prove fruitful for more temporally extended analysis that could incorporate long-term trends like the rise of women's employment.

Implications for future research. Despite its limitations, our research reveals considerable variation across tech firms and workers in the degree to which cultural conceptions of tech firms are gendered. Cultural barriers to gender inclusivity persist in tech, but in some firms and for some workers, our analysis indicates that those barriers are relatively low. Thus, cultural barriers are neither uniform nor endorsed by all workers. The question remains as to how employee discourse can be harnessed to lower – even eliminate – cultural barriers to gender equality and inclusion without imposing some sort of "right-speak" reminiscent of George Orwell's dystopian world.

Note that while tech employees tended to perceive their workplace experiences in male-typed ways, they also perceived relational, care-oriented, and egalitarian concepts as female-typed. However, we must interpret these findings with care. Do they represent enduring features of organizational schema or processes of change? Do they reflect internalized gendered divisions of labor (e.g., do employees highlight female employees' care practices and overlook their drive and ambition?) or do they reflect that female employees actually do more care work?

This paper is one of the first to harness word-embedding algorithms to analyze entire documents (rather than individual words), as well as the organizations described in those documents (see also McCumber and Davis 2022). The methodology used here can be used to study many other phenomena, including objects, events, activities, individuals, and groups. We developed a direct but unobtrusive quantitative indicator of how employee discourse is gendered and revealed which employees in which firms use more male-typed language, which can be used at large scale. Our approach can be used to investigate the power of gendered discourse to shape gender inequality and exclusion through outcomes like hiring, evaluation, promotion, and organizational performance. More broadly, our approach can be used to develop indicators of other aspects of organizational culture that are carried in language, such as race, class, individualism vs. collectivism, equality vs. hierarchy, and long- vs. short-term orientation. Doing so would enrich research on organizational culture, which has largely focused on culture strength.

References

Acker, Joan. 1990. Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender and Society*, 4: 139-158.

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Los Angeles: Sage.

Antoniak, Maria, and David Minmo. 2018. Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics*, 6: 107-119.

Arseniev-Kohler, Alina. 2022. Theoretical foundations and limits of word embeddings: What types of meaning can they capture? *Sociological Methods and Research*, in press.

Arseniev-Kohler, Alina, and Jacob Foster. 2022. Machine learning as a model for cultural learning: Teaching an algorithm what it means to be fat. *Sociological Methods and Research*, 51 (4): 1481-1539.

Barley, Stephen R. 1983. Semiotics and the study of occupational and organizational cultures. *Administrative Science Quarterly*, 28: 393-413.

Baron, James N., and William T. Bielby. 1980. Bringing the firms back in: Stratification, segmentation, and the organization of work. *American Sociological Review*, 45: 737-765.

Baron, James N., Alison Davis-Blake, and William T. Bielby. 1986. The structure of opportunity: How promotion ladders vary within and among organizations. *Administrative Science Quarterly*, 31: 248-273.

Barroso, Amanda. 2020. *Key takeaways on American's views on gender equality a century after U.S. women gained the right to vote*. Pew Research Center. (https://www.pewresearch.org/fact-tank/2020/08/13/key-takeaways-on-americans-views-on-gender-equality-a-century-after-u-s-women-gained-the-right-to-vote/; viewed 2022-10-14)

Blair-Loy, Mary. 2003. *Competing Devotions: Career and Family among Women Executives*. Cambridge, MA: Harvard University Press.

Blau, Peter M., and Richard Schoenherr. 1971. *The Structure of Organizations*. New York: Basic Books.

Bongiorno, Renata, Paul G. Bain, Michelle K. Ryan, Peeter Kroonenberg, and Colin Wayne Leach. 2021. Think leader, think (immoral, power-hungry) man: An expanded framework for understanding stereotype content and leader-gender bias. Working paper. (https://www.sciencegate.app/document/10.31234/osf.io/p5uya)

Boutyline, Andrei, and Laura K. Soter. 2021. Cultural schemas: What they are, how to find them, and what to do once you've caught them. *American Sociological Review*, 86 (4): 728-758.

Britton, Dana M. 2000. The epistemology of the gendered organization. *Gender and Society*, 14: 418-434.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356: 183-186.

Carley, Kathleen, and Michael Palmquist. 1992. Extracting, representing, and analyzing mental models. *Social Forces*, 70 (3): 601-636.

Chang, Emily. 2018. *Brotopia: Breaking up the Boys' Club of Silicon Valley*. New York: Portfolio/Penguin.

Chatman, Jennifer A. 1991. Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36: 459-484.

Chatman, Jennifer A., Jeffrey T. Polzer, Sigal G. Barsade, and Margaret A. Neale. 1998. Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. *Administrative Science Quarterly*, 43: 749-780.

Cohen, Wesley D., and Steven Klepper. 1992. The anatomy of industry R&D distributions. *American Economic Review*, 82 (4): 773-799.

Correll, Shelley J., Katherine R. Weisshaar, Alison T. Wynn, JoAnne Delfino Wehner. 2020. Inside the black box of organizational life: The gendered language of performance assessment. *American Sociological Review*, 85 (6): 1022-1050.

Corritore, Matthew, Amir Goldberg, and Sameer B. Srivastava. 2020. Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly*, 65: 359-394.

DeMers, Jayson. 2014. How negative online company reviews can impact your business and recruiting. *Forbes*, September 9th. (https://www.forbes.com/sites/jaysondemers/2014/09/09/how-negative-online-company-reviews-can-impact-your-business-and-recruiting/?sh=2b604f1b1d9b; viewed 2020-08-02)

DiMaggio, Paul. 1997. Culture and cognition. *Annual Review of Sociology*, 23: 263-287.

Dobbin, Frank R., Daniel Schrage, and Alexandra Kalev. 2015. Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity. *American Sociological Review*, 80 (5): 1014-1044.

Eagly, Alice H., and Steven J. Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109 (3): 573-598.

Eagly, Alice H., Christa Nater, David I. Miller, Michèle Kaufmann, and Sabine Sczesny. 2020. Gender stereotypes have changed: A cross-temporal analysis of U.S. public-opinion polls from 1946 to 2018. *American Psychologist*, 75 (3): 301-315.

Edelman, Lauren B. 2016. *Working Law: Courts, Corporations, and Symbolic Civil Rights*. Chicago: University of Chicago Press.

Firth, John R. 1957. A synopsis of linguistic theory 1930-1955. In John R. Firth, ed., *Studies in Linguistic Analysis* (Special volume of the Philological Society): 1-32. Oxford: Basil Blackwell.

Fiske, Susan T., Amy J.C Cuddy, Peter S. Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82: 878-902.

Fowler, Susan. 2017. Reflecting on one very strange year at Uber. (viewed 12 June 2019)

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115 (16): E3635-E3644.

Goldberg, Amir, Sameer B. Srivastava, V. Govind Manian, William Monroe, and Christopher Potts. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *American Sociological Review*, 81 (6): 1190-1222.

Gorman, Elizabeth H. 2005. Gender stereotypes, same-gender preference, and organizational variation in the hiring of women: Evidence from law firms. *American Sociological Review*, 70: 702-728.

Harris, Zellig S. 1954. Distributional structure. *Word*, 10 (2-3): 146-162.

Haveman, Heather A., and Lauren S. Beresford. 2012. If you're so smart, why aren't you the boss? Explaining the persistent vertical gender gap in management. *Annals of the American Academy of Political and Social Science*, 639: 114-130.

Heilman, Madeline. 2012. Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32: 113-135.

Hentschel, Tanja, Madeline E. Heilman, and Claudia V. Peus. 2019. The multiple dimensions of gender stereotypes: A current look at men's and women's characterization of others and themselves. *Frontiers in Psychology*, 10: article 11.

Hofstede, Geert. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills: Sage.

Hofstede, Geert. 1998. Identifying organizational subcultures: An empirical approach. *Journal of Management Studies*, 35 (1): 1-12.

Hunzaker, M.B. Fallin, and Lauren Valentino. 2019. Mapping cultural schemas: From theory to method. *American Sociological Review*, 84 (5): 950-981.

Jurafsky, Dan, and Martin. 2023. *Speech and Language Processing, 3rd Edition*. (https://web.stanford.edu/~jurafsky/slp3/)

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation*. New York: Basic Books.

Kelly, Erin L., and Frank Dobbin. 1999. Civil rights law at work: Sex discrimination and the rise of maternity leave policies. *American Journal of Sociology*, 105: 455-492.

Kelly, Erin L., and Phyllis Moen. 2020. *Overload: How Good Jobs Went Bad and What We Can Do About It*. Princeton, NJ: Princeton University Press.

Koene, B.A., A.J.J. Boone, and J.J. Soeters. 1997. Organizational factors influencing homogeneity and heterogeneity of organizational cultures. In S.A. Sackman, Ed., *Cultural Complexity in Organizations*: 273-293. Thousand Oaks, CA: Sage.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84 (5): 905-949.

Lévi-Strauss, Claude. 1963 [1983]. *Structural Anthropology*. Chicago: University of Chicago Press.

Lieberman, Marvin B., and David B. Montgomery. 1988. First-mover advantages. *Strategic Management Journal*, 9 (Summer): 41-58.

Marinescu, Ioana, N. Klein, Andrew Chamberlain, and Morgan Smart. 2021. Incentives can reduce bias in online reviews. *Journal of Experimental Psychology: Applied,* 27 (2): 393-407.

McCumber, Andrew, and Adam Davis. 2022. Elite environmental aesthetics: Placing nature in climate change. *American Journal of Cultural Sociology*, forthcoming.

Merton, Robert K. 1948. On the bearing of empirical research for the development of theory. *American Sociological Review*, 13 (5): 505-515.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, CA (December 2013): 3111-3119.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Efficient estimation of word representations in vector space. (https://arxiv.org/abs/1301.3781)

Miller, Pavla. 1998. *Transformations of Patriarchy in the West, 1500-1900*. Bloomington: Indiana University Press.

Neely, Megan Tobias, Patrick Sheehan, and Christine L. Williams. 2023. Social inequality in high tech: How gender, race, and ethnicity structure the world's most powerful industry. *Annual Review of Sociology*, 49: 319-338.

Nguyen, Duc Duy, Linh Nguyen, and Vathunyoo Sila. 2019. Does corporate culture affect bank risk-taking? Evidence from loan-level data. *British Journal of Management*, 30: 106-133.

O'Reilly, Charles A., Jennifer A. Chatman, and David F. Caldwell. 1991. People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal*, 34: 487-516.

Reader, Tom W., Alex Gillespie, Julie Haid, and Meagan Patterson. 2020. Unobtrusive indicators of culture for organizations: A systematic review. *European Journal of Work and Organizational Psychology*, 29 (5): 633-649.

Reskin, Barbara F. 1993. Sex segregation in the workplace. *Annual Review of Sociology*, 19: 241-270.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford: Oxford University Press.

Rosch, Eleanor H., and Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7: 573-605.

Sapir, Edward. 1958. *Culture, Language and Personality*. Berkeley: University of California Press.

de Saussure, Ferdinand. 1916 [2011]. *Course in General Linguistics*. (Translated by Wade Baskin. Edited by Perry Meisel and Haun Saussy.) New York: Columbia University Press.

Schmiedel, Theresa, Oliver Müller, and Jan vom Brocke. 2019. Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example of organizational culture. *Organizational Research Methods*, 22 (4): 941-968.

Scott, W. Richard, and Gerald F. Davis. 2007. *Organizations and Organizing: Rational, Natural, and Open Systems Perspectives*. Upper Saddle River, NJ: Prentice-Hall.

Smith-Doerr, Laurel, Debra Fitzpatrick, Sharla Alegria, Donald Tomaskovic-Devey, and Kay Husbands Fealing. 2019. Gender pay gaps in U.S. federal science agencies: An organizational approach. *American Journal of Sociology*, 125 (2): 534-576.

Stainback, Kevin, Sibyl Kleiner, and Sheryl Skaggs. 2016. Women in power: Undoing or redoing the gendered organization? *Gender and Society*, 30 (1): 109-135.

Stryker, Sheldon. 1980. *Symbolic Interactionism: A Social Structural Version*. Menlo Park, CA: The Benjamin/Cummings Publishing Company.

Storer, Adam, and Adam Reich. 2021. "Losing my raise": Minimum wage increases, status loss and job satisfaction among low-wage employees. *Socio-Economic Review*, 19 (2): 681-709.

Taylor, Marshall A., and Dustin S. Stoltz. 2021a. Integrating semantic dimensions with concept mover's distance to measure binary concept engagement. *Journal of Computational Social Science*, 4: 231-242.

Taylor, Marshall A., and Dustin S. Stoltz. 2021b. Concept class analysis: A method for identifying cultural schemas in texts. *Sociological Science*, 7: 544-569.

Trice, Harrison M. 1984. *Occupational Subcultures in the Workplace*. Ithaca, NY: ILR Press.

U.S. Bureau of Economic Analysis. 2014. Regional Maps, GDP by State 2014. https://www.bea.gov/news/2015/gross-domestic-product-state-advance-2014-and-revised-1997-2013/regional-maps

Vaisey, Stephen, and Andrew Miles. 2017. What you can – and can't – do with three-wave panel data. *Sociological Methods and Research*, 46 (1): 44-67.

Van Maanen, John. 1991. The smile factory: Work at Disneyland. In Craig C. Lundberg, Joanne Martin, Larry F. Moore, Meryl Reis Louis, and Peter J. Frost, eds., *Reframing Organizational Culture*: 58-76. Newbury Park, CA: Sage.

Whorf, Benjamin L. 1956. Science and linguistics. In John B. Carroll, ed., *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Williams, Joan C. 2000. *Unbending Gender: Why Work and Family Conflict and What to Do about It*. New York: Oxford University Press.

Wynn, Alison T., and Shelley J. Correll. 2018. Puncturing the pipeline: Do technology companies alienate women in recruiting sessions? *Social Studies of Science*, 48 (1): 149-164.

Zippia. 2023. *Trending Tech Industry Statistics*. (https://www.zippia.com/advice/tech-industry-statistics/#:~:text=By%20the%20end%20of%202021,grow%20by%206.7%25%20in%202022, viewed 2023-08-14

Figure 1: Example Employee Review from Glassdoor.com

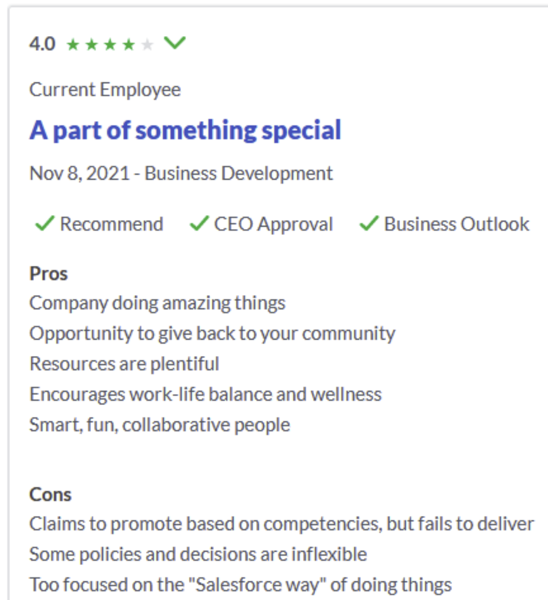


Figure 1a: Example Employee Review from Glassdoor.com: Optional Ratings on Topics

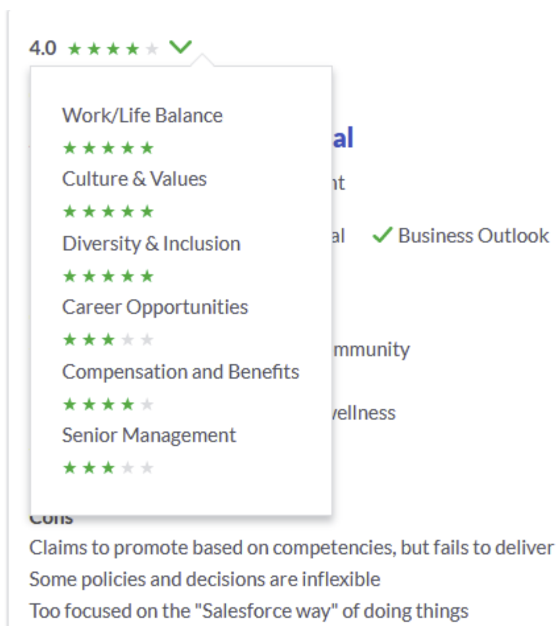


Figure 2a: Conceptual Diagram of Gender Word Embeddings and Resultants

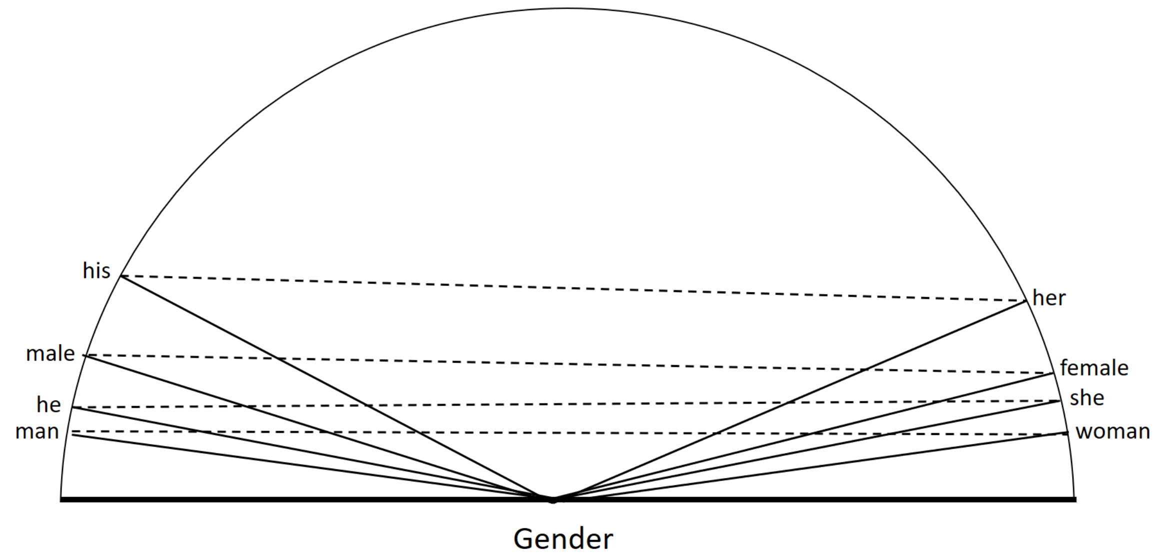
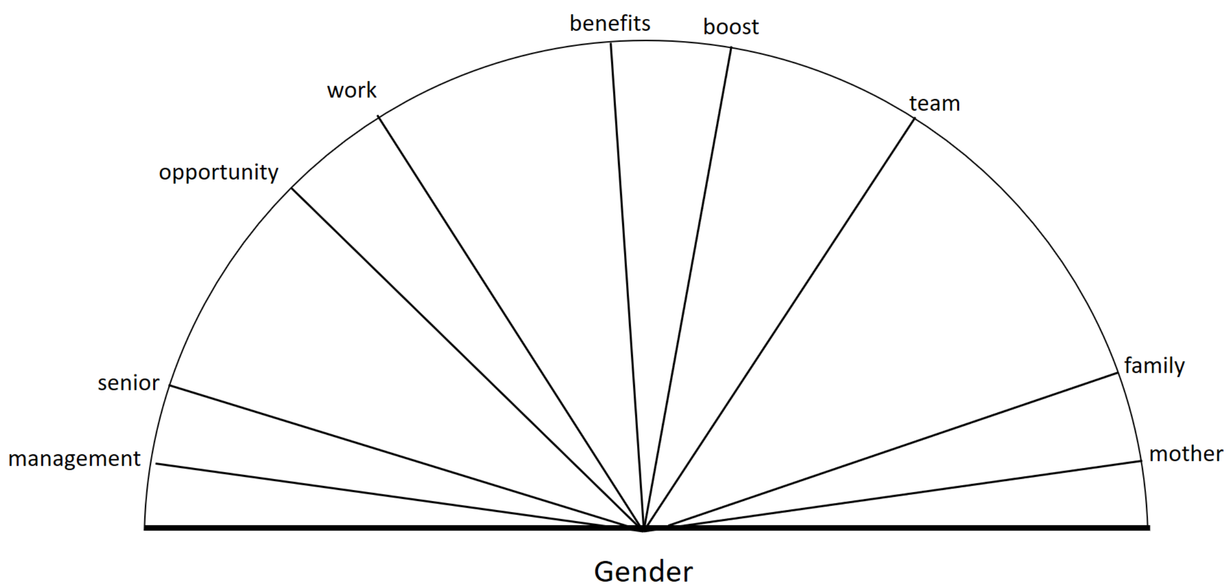
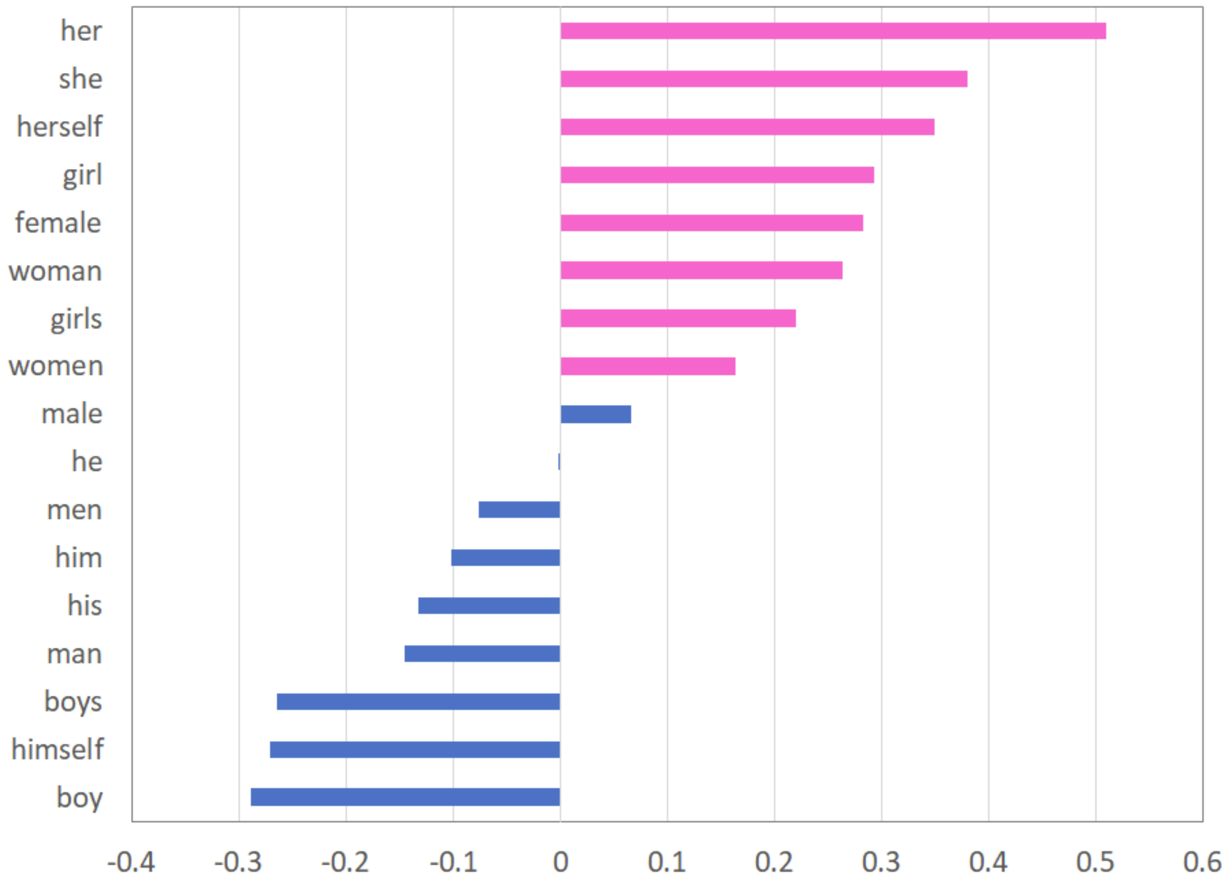


Figure 2b: Conceptual Diagram of Gender Axis and Words Describing Tech Firms



Adapted from Kozlowski et al. (2019: Figure 2).

Figure 3: Associations between Gender Terms and the Gender Axis

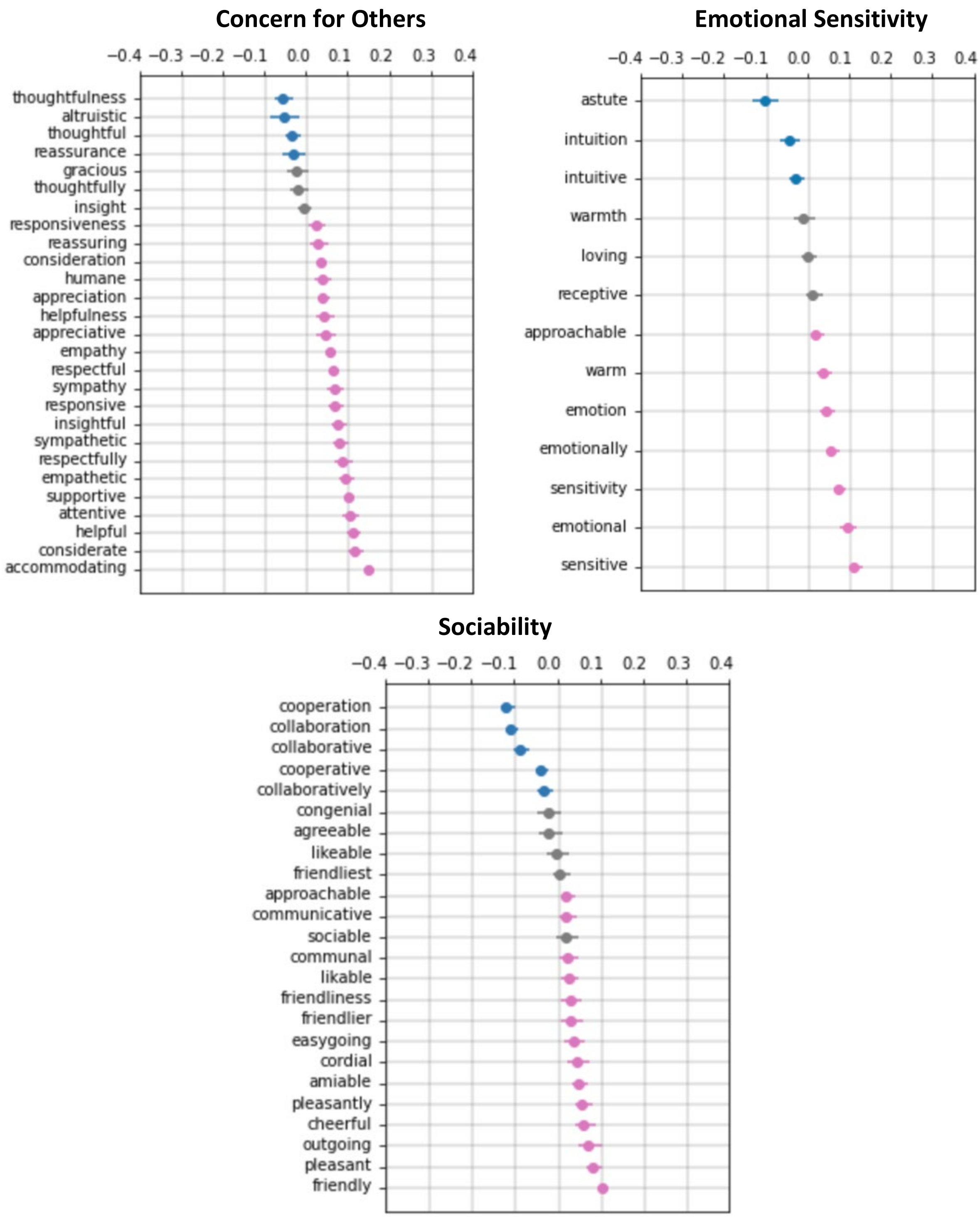


Notes: Pink indicates female-denoting words; blue indicates male-denoting words.
99% confidence intervals (not shown) around all words EXCEPT “he” do NOT cross the zero axis.
The mean for “he” is -0.002; the 99% confidence interval runs from -0.050 to 0.053.

Table 1: Common Words Most Strongly Associated with the Gender Cultural Binary

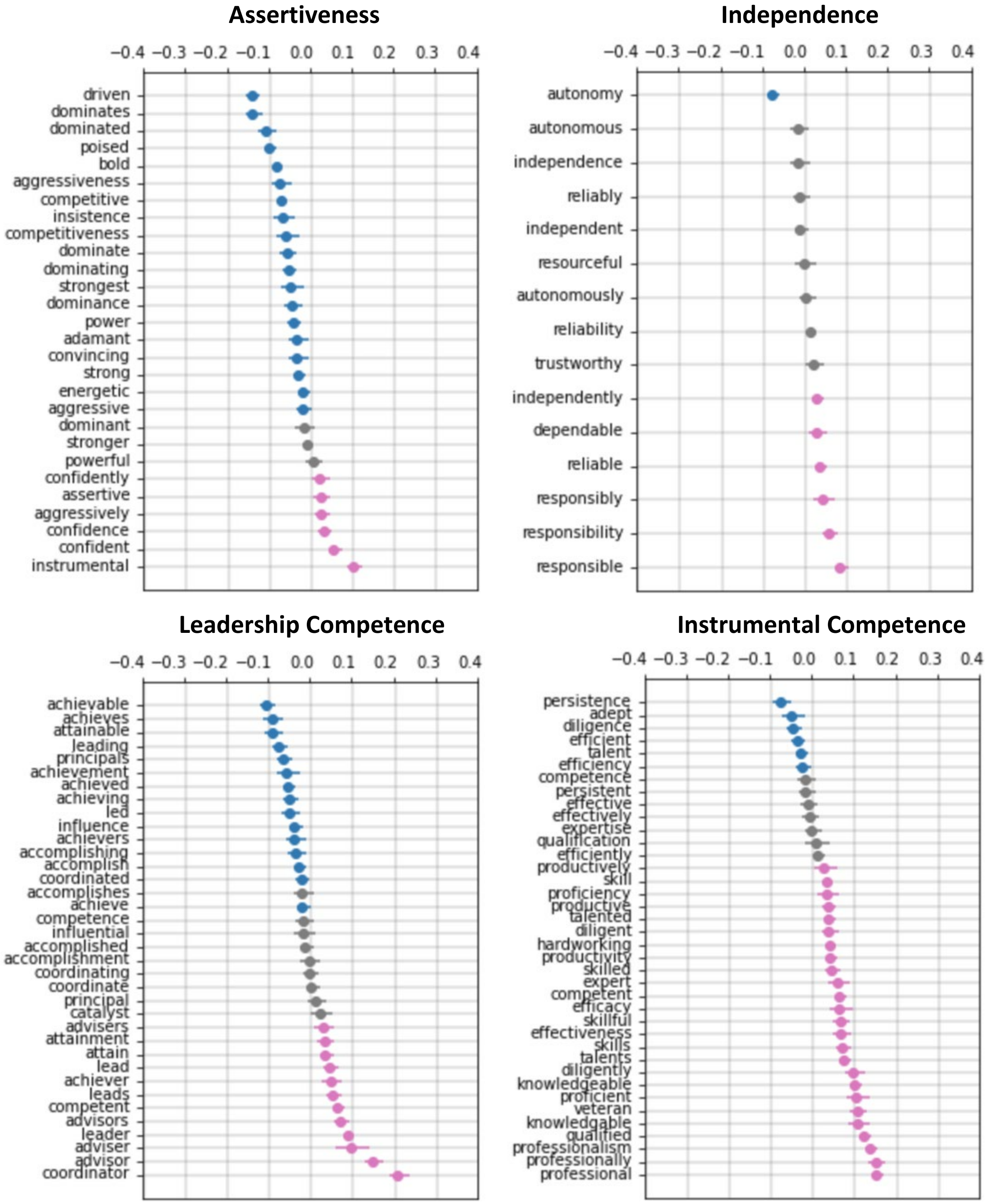
Male End of the Axis			Female End of the Axis		
Word	**Cosine Similarity**	**Frequency**	**Word**	**Cosine Similarity**	**Frequency**
boy	-0.318	2,759	her	0.455	10,645
ol	-0.310	1,081	assistant	0.390	1,252
shareholders	-0.306	2,233	myself	0.350	9,474
boys	-0.292	4,573	supervisor	0.334	8,921
emc	-0.270	1,205	coworker	0.322	1,455
ole	-0.269	1,211	female	0.289	3,395
kool	-0.268	1,723	assistance	0.268	4,770
vc	-0.265	1,024	supervisors	0.266	13,059
entrepreneurial	-0.240	3,657	unprofessional	0.266	8,740
investors	-0.237	4,276	degrees	0.263	1,813
cronyism	-0.237	1,082	specialist	0.255	1,629
legacy	-0.235	7,225	fortunate	0.250	2,027
aid	-0.233	2,744	child	0.246	2,405
mentality	-0.232	11,645	inappropriate	0.244	1,809
reactive	-0.230	2,315	coworkers	0.242	38,680
reality	-0.228	8,333	student	0.242	3,154
selling	-0.225	12,645	ended	0.242	3,936
trenches	-0.224	2,209	store	0.236	12,430
dell	-0.223	4,454	permanent	0.234	2,879
sighted	-0.220	1,328	woman	0.232	2,404
arrogance	-0.219	1,253	denied	0.229	1,175
silo	-0.217	1,323	manager	0.228	68,017
matrix	-0.216	1,146	coaches	0.227	1,369
embrace	-0.216	3,976	sat	0.222	1,321
mindset	-0.215	4,332	particular	0.220	5,774
mediocrity	-0.214	1,040	especially	0.219	35,336
analysis	-0.212	2,024	saw	0.217	6,450
chasing	-0.210	1,632	rude	0.214	6,070
yes	-0.209	11,740	position	0.214	48,160
acquisitions	-0.209	6,177	graduate	0.212	1,863

Figure 4a: Associations between the Gender Axis and Words Denoting Female Stereotypes



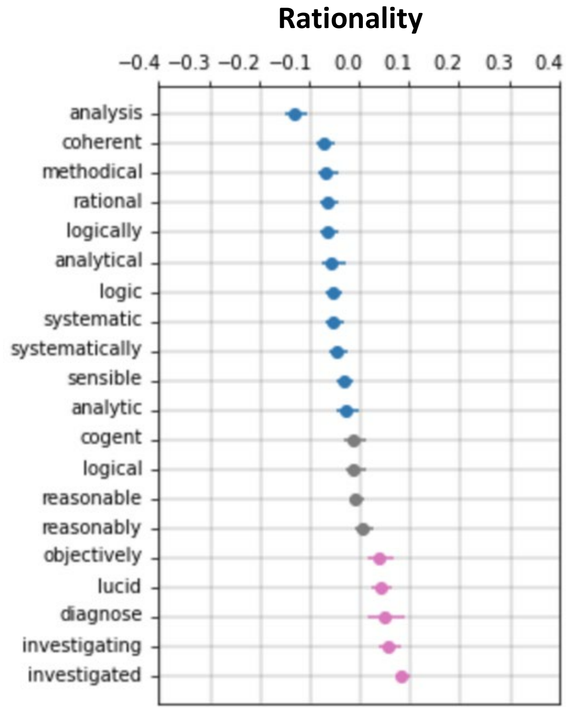
Notes: Dots represent means across 20 subsamples, while ends of the bars represent 95% confidence intervals. Blue indicates words that are male-shaded, gray indicates words that are gender-neutral, and pink indicates words that are female-shaded.

Figure 4b: Associations between the Gender Axis and Words Denoting Male Stereotypes



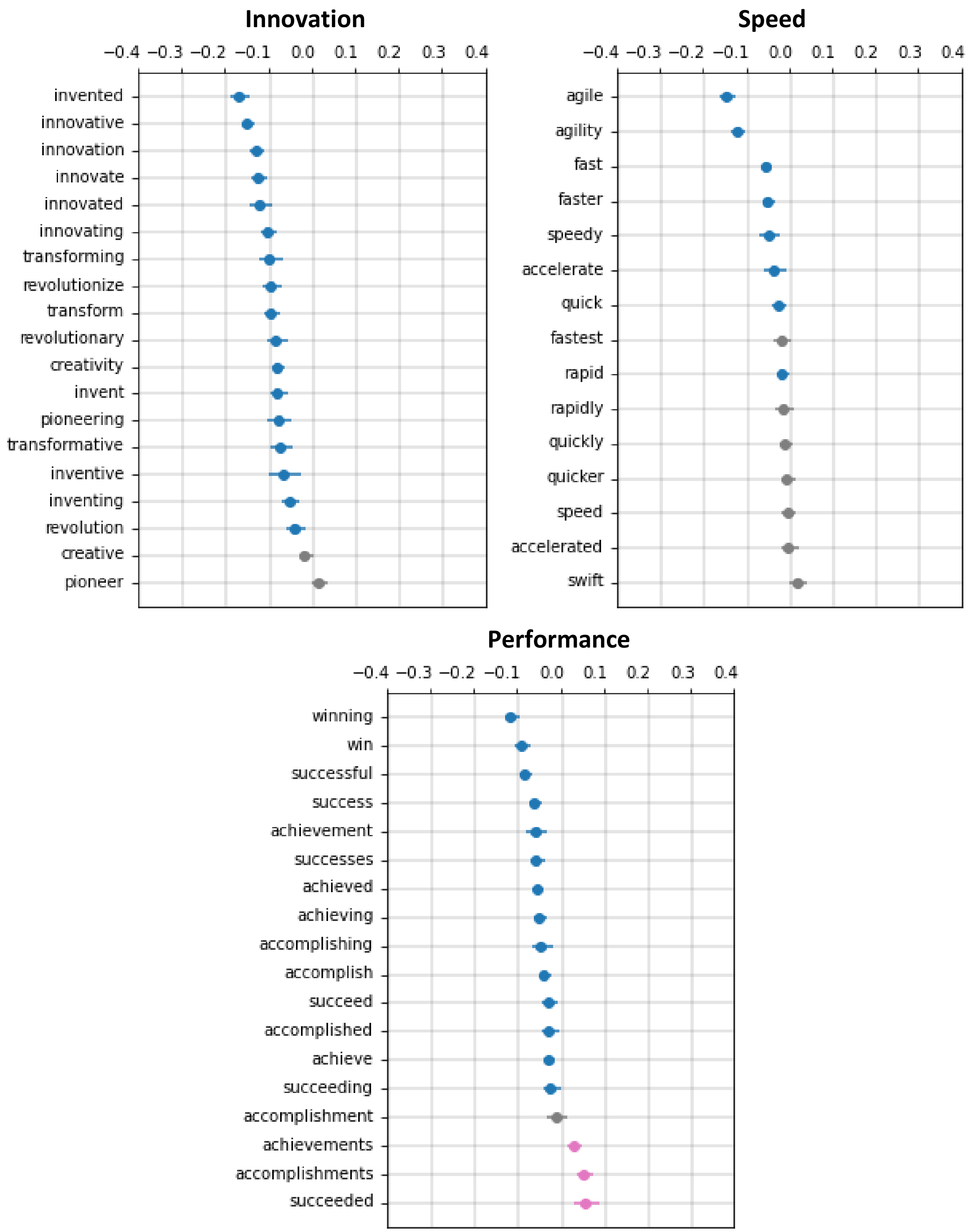
Notes: Dots represent means across 20 subsamples, while ends of the bars represent 95% confidence intervals. Blue indicates words that are male-shaded, gray indicates words that are gender-neutral, and pink indicates words that are female-shaded.

Figure 4b (cont'd):
Associations between the Gender Axis and Words Denoting Male Stereotypes



Notes: Dots represent means across 20 subsamples, while ends of the bars represent 95% confidence intervals. Blue indicates words that are male-shaded, gray indicates words that are gender-neutral, and pink indicates words that are female-shaded.

Figure 5: Associations Between Gender Axis and Words Denoting Corporate Values



Notes: Dots represent means across 20 subsamples, while ends of the bars represent 95% confidence intervals. Blue indicates words that are male-shaded, gray indicates words that are gender-neutral, and pink indicates words that are female-shaded.

Figure 6: The Distribution of the Review-Average Gender Score

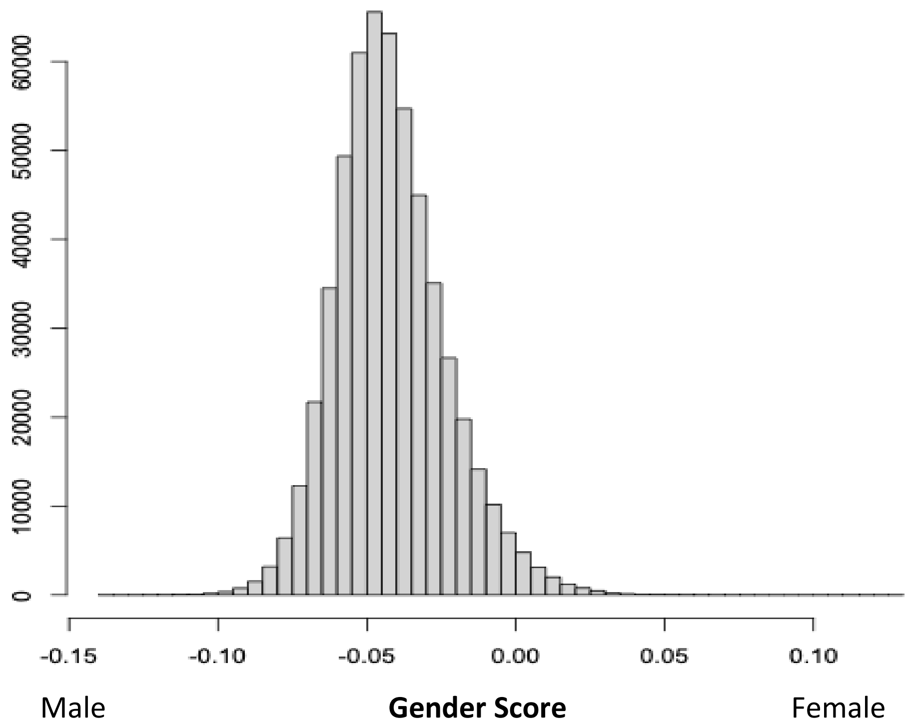


Figure 7: The Distribution of the Firm-Average Gender Score

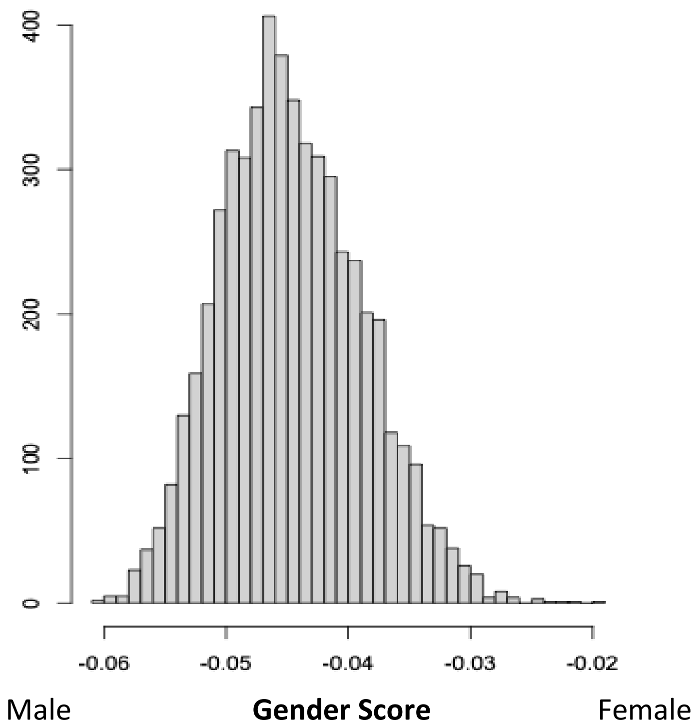


Table 2a: Reviews with the Lowest Gender Score (Most Male-shaded)

Living the values of the company. Matrix org can slow down execution.
The solution, the people, the vision, the culture. The Board that ultimately calls the shots
good vision by the founder / CEO poor execution by the middle management team
Good money when you can get past the obstacles Challenges selling with the uncertainty of the merger
Strategic business goals have improved. Old guard divisional leadership hinders innovation and growth.
Innovation, vision, breadth of offerings Still settling down after EMC merger Keep the vision going!
Strength of Innovation, Disruptive, Convergence and Interconnection Just have Short term milestones Drive employees to have entrepreneurship and product commercialization
The people, the leadership, the benefits Bureaucracy can sometimes get in the way.
The money can be good. Churn and burn mentality, upper management is cut throat.
You get the freedom to operate Too many silos in the company

Table 2b: Reviews with the Highest Gender Score (Most Female-shaded)

Great for college student. Hours and Pay Max 28 hrs for PT.
Housing stipend Friendly office environment Intern program activities Flexible hours Low structure for some intern positions
Friendly coworkers. Kept busy during summer Sent home or called off a lot during spring
Nice environment, Great coworkers, good supervisor temp position, monotonous work, call job
Helpful supervisors and friendly coworkers Not enough training for an internship position
Flexible work time, salary, friendly co-workers health benefits, personal assistant job instead of Executive Assistant
Needed employment to support myself Long hours and physically very demanding especially for female employees. No comment.
Pay, Benefits, UPT/ PTO/ VTO, Hours, Vacation time horrible work schedule for college student
Treated like family, friendly helpful people, opportunity for advancement Slow at times, Saturday work
Frequent food days Helpful managers Friendly coworkers High stress call center environment

Table 3: OLS Regression Analysis of the Review-Level Gender Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (Female = 1)	0.00336***				0.00421***	0.00314***	0.00285***	0.00370***
	(0.000053)				(0.000137)	(0.000076)	(0.000095)	(0.000164)
Satisfaction		0.000382***			0.000499***			0.000554***
		(0.000018)			(0.000022)			(0.000022)
Ownership Public			0.00294***			0.00281***		0.000085
			(0.000053)			(0.000064)		(0.000088)
Ownership Other			0.00236***			0.00231***		0.00108***
			(0.000104)			(0.000126)		(0.000130)
Firm Size Medium				0.00206***			0.00177***	0.00186***
				(0.000062)			(0.000076)	(0.000083)
Firm Size Large				0.00504***			0.00478***	0.00480***
				(0.000064)			(0.000077)	(0.000106)
Gender * Satisfaction					-0.000235***			-0.000227***
					(0.000037)			(0.000037)
Gender * Ownership Public						0.000582***		0.000256
						(0.000108)		(0.000150)
Gender * Ownership Other						0.000215		0.000132
						(0.000215)		(0.000220)
Gender * Firm Size Medium							0.000732***	0.000571***
							(0.000129)	(0.000140)
Gender * Firm Size Large							0.000936***	0.000675***
							(0.000130)	(0.000181)
Constant	-0.0439***	-0.0440***	-0.0439***	-0.0448***	-0.0457***	-0.0451***	-0.0458***	-0.0479***
	(0.00012)	(0.00013)	(0.00012)	(0.00012)	(0.00014)	(0.00012)	(0.00013)	(0.00015)

Notes: N=545,519 reviews from January 2014 to September 2020. These results are net of fixed effects for industry, region, and year, which are shown in Table A3 in the Appendix. * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table 4: Hybrid FE-RE Regression Analysis of the Review-Level Gender Score

	(1)	(2)	(3)	(4)
Firm Mean Gender (Female = 1)	0.00566***			
	(0.000801)			
FE Gender (Female = 1)	0.00345***			
	(0.000061)			
Firm Mean Satisfaction		-0.00189***		
		(0.000169)		
FE Satisfaction		0.000823***		
		(0.000022)		
Ownership Public			0.00211***	
			(0.000244)	
Ownership Other			0.00325***	
			(0.000384)	
Firm Size Medium				0.00288***
				(0.000211)
Firm Size Large				0.00619***
				(0.000385)
Constant	-0.0450***	-0.0363***	-0.0437***	-0.0450***
	(0.000321)	(0.000608)	(0.000181)	(0.000206)

Notes: N=414,029 reviews from January 2014 to September 2020. These results are net of fixed effects for region and year, which are not shown to save space. Coefficients labeled “FE” are those where values for factors that vary greatly over time within firms (gender and satisfaction) are deviated from their firm-mean values. Those coefficients give us within-firm estimates of the effects of gender and satisfaction. * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table 4: Hybrid FE-RE Regression Analysis of the Review-Level Gender Score (continued)

	(5)	(6)	(7)	(8)
Firm Mean Gender (Female = 1)	0.00963***	0.00676***	0.00853***	0.0137***
	(0.002780)	(0.000908)	(0.00110)	(0.00296)
FE Gender (Female = 1)	0.00403***	0.00299***	0.00268***	0.00323***
	(0.000160)	(0.000102)	(0.000166)	(0.000229)
Firm Mean Satisfaction	-0.00147***			-0.000568
	(0.000317)			(0.000310)
FE Satisfaction	0.000909***			0.000907***
	(0.000027)			(0.000027)
Ownership Public		0.00311***		0.000297
		(0.000752)		(0.000796)
Ownership Other		0.00312**		0.00135
		(0.000996)		(0.000957)
Firm Size Medium			0.00374***	0.00314***
			(0.000575)	(0.000604)
Firm Size Large			0.00883***	0.00831***
			(0.00126)	(0.001382)
Firm Mean Gender * Satisfaction	-0.00121			-0.00168*
	(0.000786)			(0.000760)
FE Gender * Satisfaction	-0.000152***			-0.000145***
	(0.000043)			(0.000043)
Firm Mean Gender * Ownership Public		-0.00249		-0.000139
		(0.00220)		(0.00229)
FE Gender * Ownership Public		0.000728***		0.000478**
		(0.000130)		(0.000166)
Firm Mean Gender * Ownership Other		0.000515		0.00230
		(0.00269)		(0.00257)
FE Gender * Ownership Other		0.000571*		0.000393
		(0.000254)		(0.000260)
Firm Mean * Firm Size Medium			-0.00201	-0.00182
			(0.00155)	(0.00159)
FE Gender * Firm Size Medium			0.000696***	0.000443*
			(0.000191)	(0.000201)
Firm Mean Gender * Firm Size Large			-0.00722	-0.00776
			(0.003830)	(0.00413)
FE Gender * Firm Size Large			0.00107***	0.000631**
			(0.000189)	(0.000232)
Constant	-0.0398***	-0.0462***	-0.0481***	-0.0458***
	(0.001140)	(0.000366)	(0.000444)	(0.00123)

Notes: N=414,029 reviews from January 2014 to September 2020. These results are net of fixed effects for region and year, which are not shown to save space. Coefficients labeled "FE" are those where values for factors that vary greatly over time within firms (gender and satisfaction) are deviated from their firm-mean values. Those coefficients give us within-firm estimates of the effects of gender and satisfaction.
* indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Appendix: Data Preparation and Technical Details of Analysis

Here we describe how we prepared the Glassdoor data for analysis and provide more details on the NLP techniques we used. We also discuss the robustness tests.

Data preprocessing

We used standard NLP techniques (i.e., regular expressions) to pre-process the text data. We transformed some symbols into words: "&" became "and," "%" became "percent," and "@" became "at." We removed most punctuation marks and symbols but retained the symbols "/" and ":" when they were surrounded by numbers, and the symbol "$" when it was followed by numbers. We replaced hyphens with "to" when surrounded by numbers and retained them otherwise. We also normalized the text – i.e., we turned upper-case letters into lower-case ones – which increased the frequencies of rare words. Finally, we removed excess white space (blanks).

The text dataset originally contained 1,128,375 reviews of firms in the tech sector (nine industries defined above) from January 2008 to September 2020. We eliminated 7,313 reviews written in other languages by comparing words in all reviews with words in lists of common words in English texts, such as "the," "too," "and," and "of." These are known as stop words; they are commonly removed from text analysis because they add little informational value to the texts in which they appear. For this, we used the intersection of three commonly used stop-word lists (from the popular Python libraries NLTK, scikit-learn, and SpaCy) and targeted reviews that contained none of those words.

We then eliminated short reviews (e.g., "Great firm! Excellent benefits"), which are often perfunctory and so not informative about organizational culture and practices. Specifically, we eliminated 41,444 reviews containing fewer than 13 words (the fifth percentile of reviews by length). The final dataset contains 1,079,978 reviews from 2008 to 2020. Limiting the analysis to reviews posted 2014 onward left 948,785 reviews. Finally, for the regression analysis, we deleted reviews from employees in U.S. territories (e.g., Puerto Rico, Guam, Samoa), reviews that reported firm size as zero (probably errors), and reviews without data on year, leaving 946,653 reviews (545,519 when restricted to reviews with data on gender). There are 210,422 unique words in the dataset (apart from stop words); the rarest are mostly misspellings (e.g., thinkful, peopke) or acronyms and jargon (e.g., powershell, saas). Because they are rare, they do not enter our word-embedding analysis: as we explain below, we eliminated from the training dataset for the word2vec algorithm any word that occurred fewer than 75 times.

Indexing firms for the fixed-effects analysis. We used firm names, as recorded in the data, to match reviews for the same firm. Glassdoor's system offers a drop-down list of employer names when workers post reviews. To ensure that employer names were standardized, we used the record linkage package in Python (documentation: https://recordlinkage.readthedocs.io/en/latest/index.html) and two string-matching algorithms (Damerau-Levenshtein and Jaro-Winkler) to pull out possible matches between pairs of texts (for us, employer names). We checked all possible matches using internet searches for information on both employer names to classify matches as true or false. Many employer names were very similar – e.g., Advanced Computer Technology and Advanced Computer Technologies. Among the 36,957 employer names, we discovered matches for about 200 pairs of names; only 31 were true matches.

Word-embedding analysis

We used word2vec, the most popular word-embedding algorithm (Mikolov et al. 2013a; Mikolov et al. 2013b), which has been used frequently in social-science research (e.g., Kozlowski et al. 2019; Arseniev-Kohler and Foster 2022). Computationally, word2vec learns word vectors using a neural network with a single hidden layer to predict relationships between words using a sample of observations of word relationships. In any document under analysis, a word's context is a "window" of terms surrounding it. The algorithm has two architectures, continuous bag of words (CBOW) and skip-gram (SG), that use context in different but complementary ways. The CBOW architecture uses context words to predict a focal word; the SG architecture uses a focal word to predict context words, as shown in Figure A1. Each begins with a series of words w_1, w_2, w_3, … , w_T. For CBOW, the goal is to maximize the average log probability of a focal word, w_t, given a series of context words (w_{t-c}, w_{t-c-1}, w_{t-c-2}, … w_{t+c-2}, w_{t+c-1}, w_{t+c}), where c is the size of the word context window (e.g., c=6 if the window is ±6 words). For SG, the goal is to maximize the average log probability of a series of context words (w_{t-c}, w_{t-c-1}, w_{t-c-2}, … w_{t+c-2}, w_{t+c-1}, w_{t+c}) given the focal word, w_t. Distances between w_t and context words are used as weights, so context words that are closer to the focal word will have more influence than those that are farther away.

[Figure A1 about here]

We custom-trained the word2vec algorithm on our corpus.[15] The algorithm begins by randomly assigning word vectors, generating high prediction errors. After each iteration, the correct words are revealed and the algorithm updates word vectors to reduce prediction errors. Over many iterations, when predictions reach a predetermined level of accuracy, the algorithm produces a stable set of word vectors.[16]

As input for training models on our corpus, we separated out the three sections of Glassdoor reviews – pros, cons, and advice to management – to prevent contamination across sections. These sections are distinct documents, even though they were written by the same person at the same time. If we had concatenated them, the windows around words might have contained terms from a different section, feeding noise into the algorithm and worsening its performance. For example, assuming a window size of four, the window around the last word in the pros section would have been compared with the four words preceding it in that section

[15] The most commonly used pre-trained word2vec embeddings performed poorly because our corpus includes very different words than that model's training corpus. For instance, "work," "company," and "management" are among the most common words in our corpus (ignoring stop words).

[16] Technically, the optimization procedure sets vector probabilities by estimating them with a softmax function that is learned via a recurrent neural-net algorithm and that maximizes the probability of a window of words given each focal word repeatedly across chunks of text. For an intuitive understanding of this optimization procedure, see Arseniev-Kohler and Foster (2022).

and to the first four words of the cons section. Similarly, the first word in the cons section would have been compared with the four words preceding it in the cons section and the last four words in the pros section. In a preliminary analysis, we concatenated the three sections, but the best model from this analysis performed worse (in terms of the analogies test we describe below) than the best model from the analysis using three separate documents per review as input. So we present only the analysis based on training with three separate input documents per review. After model training was complete, we combined embeddings for words used in all three sections into a single observation.

We experimented with three different hyperparameters for the model: the number of dimensions of the semantic space (100, 200, 300), window size (6, 8, 10), and architecture (skip-gram vs. continuous bag of words). We selected hyperparameter combinations based on model performance (Levy et al. 2015). Following previous research (e.g., Mikolov et al. 2013a), we used two performance standards to compare models: analogies and most-similar terms. We began with the Google analogies test, which contains 8,869 semantic word pairs and 10,675 syntactical (word form) word pairs. This test assesses how well calculating relationships between pairs of word embeddings (word vectors). The relationships being tested always have to do with meaning (semantics), such as synonyms, antonyms, parts-to-wholes, categorical memberships, and degree. For example, "big:biggest :: small:___" tests a degree relationship, while "acceptable:unacceptable :: aware:___" tests an antonymic relationship. The test lexicon is available at this site; for more information about how the analogies test works, see Gladkova et al. (2016).

The results of the Google analogies test are shown in Table A1. To summarize, we found that (i) models with more dimensions were often better; (ii) smaller windows were always better, probably because the texts we study were relatively short; (iii) which architecture performed better depended on the other parameters; and (iv) all models were close in terms of performance. We also experimented with excluding rare words[17] because associations between rare words and surrounding words tended to be very noisy, so excluding rare words improved model robustness. We tried several minimum thresholds for word frequency (2, 5, 10, 25, 50, and 75). Accuracy improved as minimum frequency increased, albeit at a decreasing rate. The model with the best accuracy score, highlighted in Table A1, used the skip-gram architecture with 200 dimensions, a 6-word window, and a minimum word frequency of 75.

[Table A1 about here]

Because word-embedding models are stochastic – they are initialized with randomly chosen values – results can vary greatly across implementations, even when based on a single corpus (Tian et al. 2016; Hellrich and Hahn 2016b; Antoniak and Minmo 2018). This is especially likely when the corpus is small, in which case individual documents may have a large impact on the results. Even though our corpus is large – almost one million reviews – we assessed variability in model results (after settling on model features, as described above). To generate robust estimates of word embeddings, we ran models repeatedly ("epochs" in NLP parlance). The first epoch was initialized with random values for parameters. The output for the first epoch was used as input to the second epoch, the output for the second epoch as input to the third, and so on. We continued until the average cosine similarity score between successive epochs across all gender-denoting words reached 0.990.[18] Figure A2 shows the accuracy score across training epochs for the chosen model.

[Figure A2 about here]

As explained in the main text, we validated the model by reading reviews, comparing reviews with high and low gender scores (male-shaded vs. female-shaded). We also conducted most-similar queries; i.e., searches for words that are closest in semantic space to a focal word and assessed whether it made sense for associated words to be in close proximity to individual

---

[17] Many rare words were misspellings (e.g., "managament"), typographical errors (e.g., "sti" for "sit"), acronyms (e.g., "saas" for "software as a service"), or numbers. Others included numbers (e.g., "12th").
[18] There is no clear standard for judging when a word2vec model is robust, so we followed previous research (e.g., Kulkarni et al. 2015; Hellrich and Hahn 2016a) and set the threshold at 0.990.

words denoting gender. The results for that validity check are shown in Table A2. The top half (Table A2a) shows that female words tend to be most strongly associated with other female words, such as *woman* with women, female, females; and *female* with women, females, woman. These terms are also strongly associated with terms that tend to occupy similar semantic positions, such as male, minority, and minorities. Similarly, the bottom half (Table A2b) shows that male words tend to be most strongly associated with other male words, such as *men* with man, guy, and sir; and *boy* with boys, boy's, and buddies. These terms are also strongly associated with terms that tend to occupy similar semantic positions, such as woman, females, and minority.

[Table A2 about here]

Measuring distance from the gender axis for entire documents

To measure this distance for entire documents, rather than individual words, we used a measure called *concept-mover's distance* (Stoltz and Taylor 2019; Taylor and Stoltz 2021a, 2021b). This was adapted from word-mover's distance, a measure conceived by computer scientists (Kusner et al. 2015), which in turn was adapted from earth-mover's distance, a measure used in transportation research (Rubner et al. 1998). Let us unpack this intellectual genealogy in order of invention. *Earth-mover's distance* is the energy cost of moving some material (e.g., dirt or gravel) based on its weight and the physical distance it is to be moved, using Euclidean distance. Analogously, *word-mover's distance* is the semantic "cost" of "moving" some word to the location of another word – in semantic space rather than physical space – based on the distance it is to be moved. Semantic distance is defined by the embedding vectors the words being compared. For word-mover's distance, the total cost of moving all words in one document to the location of the words in another document is a function of semantic distances for all the words being compared.

However, the objective with word-mover's distance is not to measure semantic distances between *words*, but rather between entire *documents*. Two documents that share many semantically similar words should be closer together on this metric than two that contain many semantically dissimilar words. The word-mover's distance algorithm measures distance between documents by solving for the shortest overall path between each pair of word vectors d and d' in documents D and D' based on the words' cosine similarities, accounting for the number of times each word occurs in each document. When D and D' are the same length, this is simply the sum of the cosine distances between the nearest-neighbor word vectors for each pair of words in D and D' multiplied by the frequency of each word. But when the documents D and D' differ in length, it is possible to generate a minimal-distance path by distributing a fractional "portion" of a word vector's motion along multiple paths. This procedure is computationally intractable at scale, so a linear-time approximation that reduces computational effort, *relaxed word-mover's distance*, can be used instead (Atasu et al. 2017).[19]

Relaxed word-mover's distance involves two steps. First, the algorithm finds the total distance between all nearest-neighbor word vectors from document D to D', weighting each nearest-neighbor pair by the frequency of each word in D. Second, the distance calculation is repeated, this time for the distance from D' to D, weighted by the frequency of each word in D'. The larger of the two values is then used as word-mover's distance. Formally, this is defined as:

$$\max\left[\min \textstyle\sum_{i,j} d_i c(i,j)\,,\, \min \textstyle\sum_{i,j} d'_j c(i,j)\right],$$

where d and d' are the frequencies of words i and j in documents D and D', and c(i, j) is the cosine distance between words i and j. The larger of the two distance estimates is an upper-bound approximation of word-mover's distance, the smaller is a lower-bound approximation, and the correct distance lies somewhere in-between the two.

Figure A3 illustrates how word-mover's distance works by comparing two sentences to a target sentence. In the figure's top panel, the target sentence, D2, is in the middle. On the top

[19] The original word-mover's distance algorithm involved searching for the shortest overall path between all word vectors in D and D', where a given word vector's frequency-weighted distance could be distributed across many paths. This yields the best measure of word-mover's distance but it is computationally intractable, especially for researchers with limited computing resources.

of this panel is sentence D1, which has a similar meaning as D2, even though it doesn't contain any words that are in D2, except for two common (stop) words, "the" and "in." On the bottom of this panel is sentence D3, which has a very different meaning than D2. The arrows show the distance in terms of cosine score between each word (apart from stop words) in D1 and D3, on the one hand, and D2, on the other. The bottom panel of this figure shows calculations of the word-mover's distance measures for D1 vs. D2 (1.07) and D3 vs. D2 (1.63). Thus, D3 is farther away semantically from D2 than D1.

[Figure A3 about here]

Importantly, word-mover's distance is *not* a bag-of-words measure. Instead, it recognizes context: each word's location in semantic space is based on word embeddings. Its context-specific nature is why word-mover's distance yields lower classification errors than distance measures that ignore context (i.e., those that assume each document is simply a bag of words), including measures involving term frequency/inverse document frequency weights, latent semantic indexing, and topic modeling (Kusner et al. 2015).

Concept-mover's distance adapts word-mover's distance to determine a document's distance from a *concept* rather than a *document*. The target concept (e.g., music," "death," or "gender") can be denoted by a one or more words (e.g., "melody," "musical," and "song" for the concept "music"). The concept-mover's distance algorithm replaces the target document analyzed in the word-mover's distance algorithm with a "pseudo document" consisting entirely of words that denote the focal concept (Stoltz and Taylor 2019). Figure A4 illustrates this for the single-word concept "music," which contains just that word and constitutes the pseudo (target) document, D2, and the same sentences, D1 and D3, as in Figure A3. This time, D3 is closer to the focal concept than D1, with a concept-mover's score of 0.64 vs. 0.78.

[Figure A4 about here]

The concept-mover's algorithm does not work well for binary concepts; e.g., death/life; male/female (Taylor and Stoltz 2021a, 2021b) because words denoting opposing concepts (e.g., "death" and "life") may be closer to each other in the semantic space created by word-embedding models than to unrelated words (e.g., "death" and "music"). The simplest way around this problem is to use the word embeddings of antonym words denoting binaries (e.g., male/female or him/her for the binary "gender") to extract a vector in semantic space pointing from one concept to the other (e.g., from male to female) (Taylor and Stotz 2021a, 2021b), as we did in creating the gender axis. This vector is used as the pseudo-document in the implementation of the relaxed word-mover's distance algorithm.

To reduce computational costs, we removed stop words before implementing the word-mover's distance algorithm, as suggested. We then used the *text2map* package in R written by Stoltz and Taylor (2019; Taylor and Stoltz 2021a), which is described in their paper and which builds on code written by Kusner et al. (2015) and Selivanov (2019). Note that this package generates a measure of proximity (rather than distance) between real documents and the pseudo document containing terms denoting the concept of interest.

Robustness checks for regression models

As explained above, the models in Table 3 contain fixed effects for industry (the omitted category is computer hardware and software, the most common category), region (unknown region is the omitted category), and year (2020 is the omitted category). Table A3 shows those coefficients. There are substantial differences in the gendering of language between industries and some between regions. For example, employees in the electrical and electronic manufacturing and telecommunications industries used more female-shaded language (less male-shaded language) than those in the computer hardware and software industry (the reference category), while those in the enterprise software and network solutions industry used more male-shaded language (less female-shaded language). Employees in the Far West and Rocky Mountain regions used the most male-shaded language, while those in the Southeast region used the least. Over time there is a slight trend toward less male-shaded language. This time trend correlates with the rise of the #MeToo movement, which erupted in

October 2017, so it may reflect a broad cultural shift in consciousness about gender (in)equality in workplaces.

[Table A3 about here]

We conducted additional analyses to check the robustness of our results to sampling, model specification, and measurement, which are shown in Tables A4 to A7. First, we analyzed all reviews, including those without data on gender. To do this, we re-estimated the regression models shown in Table 3 and added a dummy variable to indicate gender unknown. These results, shown in Table A4, are very similar to the results in Table 3: most coefficients are in the same direction and have similar magnitudes and levels of statistical significance. The results again support all hypotheses except 3a. The language used by employees who did not report their gender was 0.3% less male-shaded than the language used by employees who reported their gender as male and 7.7% more male-shaded than the language used by employees who reported their gender as female.

[Tables A4-A7 about here]

Second, we re-estimated models shown in Table 3 after dropping reviews where firm size was in the lowest category (≤200 employees) because people working in small firms were less likely to report gender than those in medium-sized or large firms. These results, reported in Table A5, are very similar to those in Table 3: in the same direction and at the same level of statistical significance. The coefficient on satisfaction was 2.4 times what it was in Table 3 and the coefficients on ownership were half of what they were in Table 3. The difference between large and medium-sized firms (now the reference category) is about the same size.

Third, we dropped reviews with the highest satisfaction ratings (5 stars) because those reviews are more likely than reviews with lower ratings to be coached by managers. The results, shown in Table A6, are again very similar to those in Table 3: in the same direction and at the same level of statistical significance. But magnitudes vary for many variables: the coefficient on satisfaction is 4.5 times the size of the coefficient in Table 3, while those on ownership and firm size are about two-thirds of what they were in Table 3.

Fourth, the spread of Covid disrupted workplaces around the world. Our data included reviews posted up to September 2022. We re-estimated our regression models after dropping all reviews posted after the end of February 2022. The results on that temporal subsample, shown in Table A7, are almost identical to those on the entire temporal sample. Together, these results indicate that our main results are robust to sampling, possible coaching by managers, and reporting bias.

References for the Appendix (those not in the main paper's reference list)

Atasu, Kubilay, Thomas Parnell, Celestine Dunner, Manolis Sifalakis, Haralampos Pozidis, Vasileios Vasileiadis, Michail Vlachos, Cesar Berrospi, and Abdel Labbi. 2017. Linear-complexity relaxed word mover's distance with GPU acceleration. (Arxiv:1711.07227; retrieved 2023-10-13)

Gladkova, Anna, Aleksandr Drozd, and Satoshi Matsuoka. 2016. Analogy-based detection of morphological and semantic relations with word embeddings: What works and what doesn't. *Proceedings of the NAACL-HLT 2016*: 8-15.

Hellrich, Johannes, and Udo Hahn. 2016a. An assessment of experimental protocols for tracing changes in word semantics relative to accuracy and reliability. *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*: 111-117.

Hellrich, Johannes, and Udo Hahn. 2016b. Bad company—Neighborhoods in neural embedding spaces considered harmful. *Proceedings of the 26th Annual Conference on Computational Linguistics*: 2785-2796.

Kulkarni, Vivek, Rami El-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of language change. *Proceedings of the 24th International Conference on the World Wide Web*: 625-635.

Kusner, Matt J., Yu Sun, Nicholas I. Kolkin, and Killian Q. Weinberger. 2015. From word embeddings to document distances. *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France: n.p.

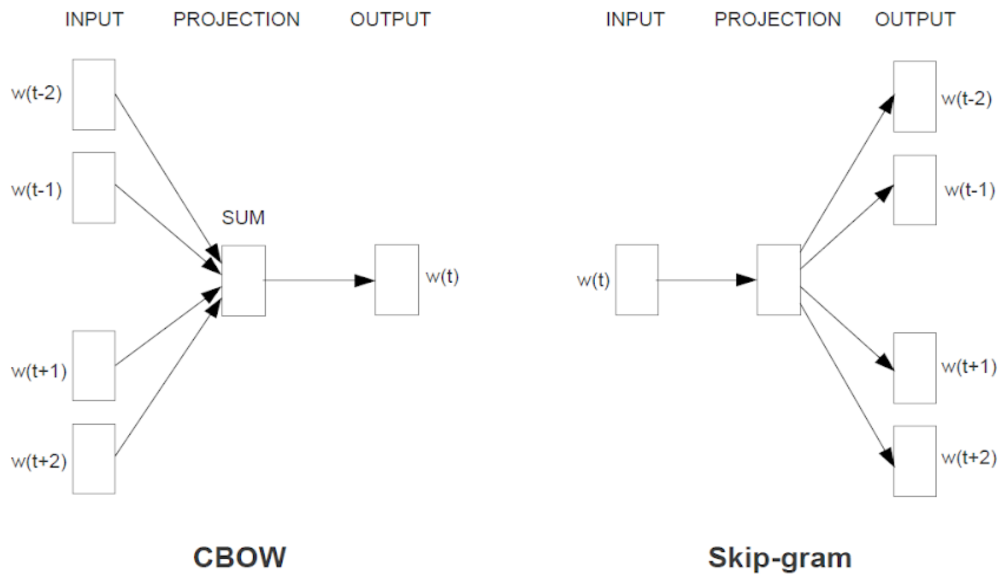
Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the ACL*, 3: 211-225.

Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. 1998. A metric for distributions with applications to image databases. *Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India*: n.p.

Selivanov, Dimtry. 2019. https://github.com/dselivanov/text2vec/blob/master/R/distance_RWMD.R; retrieved 2023-04-22.

Tian, Yingtao, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2016. On the convergent properties of word embedding methods. (https://arxiv.org/abs/1605.03956)

Figure A1: Word2vec Architectures: Continuous Bag of Words (CBOW) vs. Skip-gram



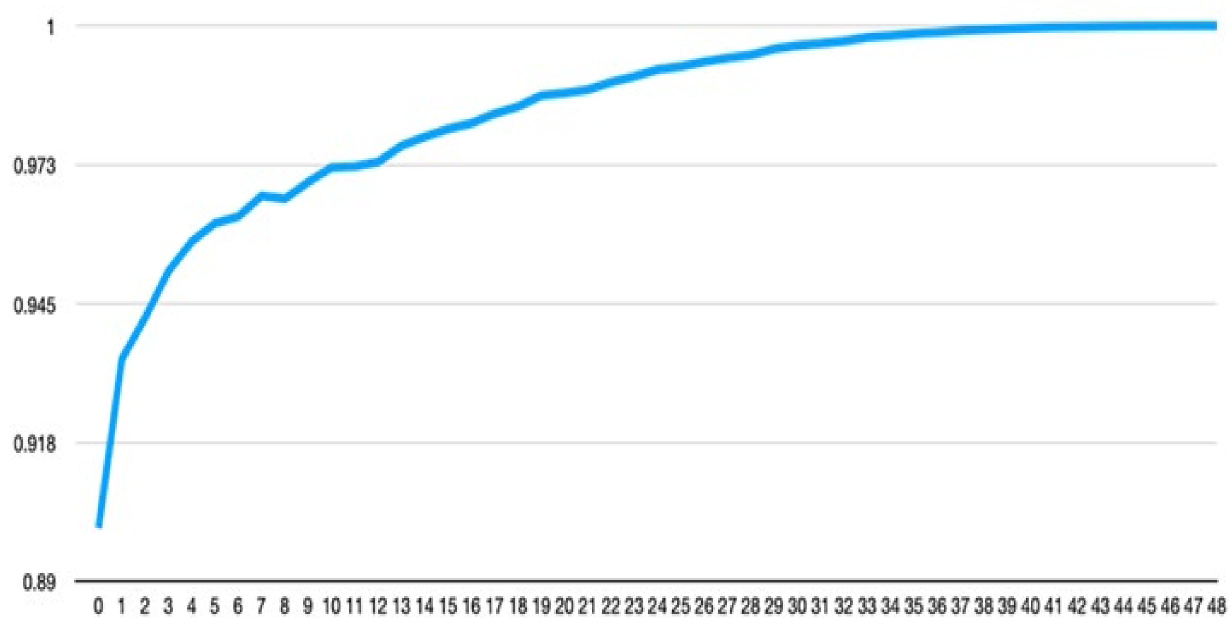
Source: Mikolov et al. (2013a: 4)

Table A1:
The Performance on the Google Analogies Test of Word2Vec Models with Different Hyperparameters

# Dimensions	Window Size	Min Word Freq	Accuracy: CBOW	Accuracy: SG
100	6	0	0.242977626	0.260611975
100	6	10	0.296775876	0.3021639
100	6	25	0.338502895	0.357733664
100	6	50	0.378998264	0.401314158
100	6	75	0.412927427	0.432837974
100	8	0	0.241374503	0.251341744
100	8	10	0.287564091	0.299817502
100	8	25	0.346050455	0.353701406
100	8	50	0.364492933	0.384453261
100	8	75	0.394315395	0.425624008
100	10	0	0.229943542	0.246671778
100	10	10	0.277830886	0.294690189
100	10	25	0.324234905	0.340674111
100	10	50	0.366476568	0.372551451
100	10	75	0.385658635	0.415235897
200	6	0	0.266885063	0.270997421
200	6	10	0.329625445	0.327887373
200	6	25	0.3691067	0.378411911
200	6	50	0.410860402	0.434044136
200	6	75	0.453614197	0.464146588
200	8	0	0.262772705	0.266397156
200	8	10	0.315547058	0.328756409
200	8	25	0.372415219	0.373759305
200	8	50	0.4065212	0.414951649
200	8	75	0.441061896	0.456932622
200	10	0	0.260054367	0.263190911
200	10	10	0.311462588	0.31511254
200	10	25	0.360628619	0.370347395
200	10	50	0.401810067	0.414951649
200	10	75	0.437887751	0.445245996
300	6	0	0.265839548	0.261308984
300	6	10	0.329625445	0.311810202
300	6	25	0.376447477	0.370244003
300	6	50	0.412224151	0.418794942
300	6	75	0.453902756	0.441206175
300	8	0	0.269812504	0.259357357
300	8	10	0.324498132	0.309550708
300	8	25	0.364040529	0.355769231
300	8	50	0.411108356	0.406273246
300	8	75	0.447265907	0.434425047
300	10	0	0.261936293	0.254060082
300	10	10	0.323281481	0.310419745
300	10	25	0.365074442	0.355562448
300	10	50	0.406645177	0.392635755
300	10	75	0.44236041	0.430240946

Notes: CBOW stands for the continuous-bag-of-words model architecture; SG for the skip-gram model architecture. The accuracy score for the best-fitting model is highlighted in yellow.

Figure A2: Accuracy Score from Training the Best Word2Vec Model over Multiple Epochs



Note: The best-performing word2vec model, according to the parameter turning we conducted using the Google analogies test, has a 6-word window around the focal word, a semantic space of 200 dimensions, and uses the skip-gram architecture. All words with frequency less than 75 were dropped.

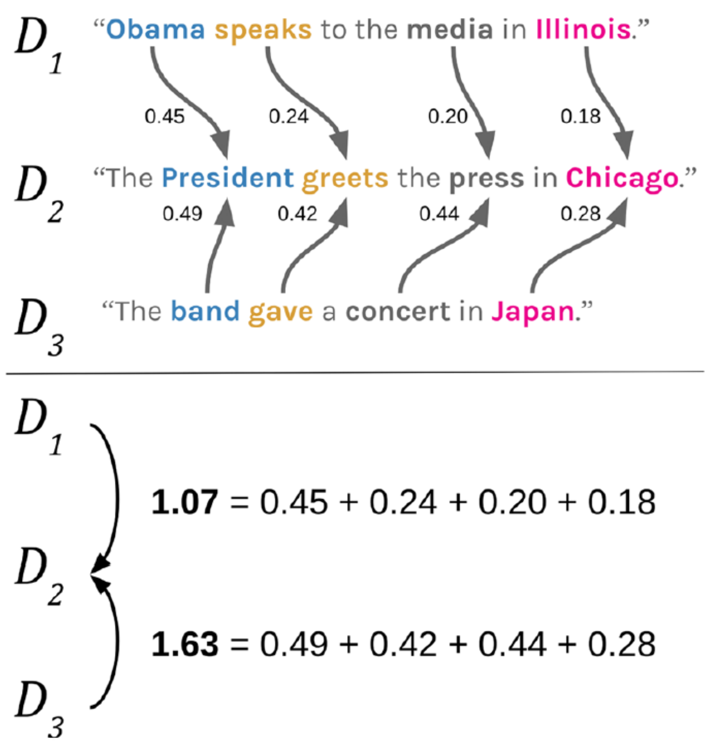
Table A2a: Most Similar Words for Female Terms

woman	**women**	**she**	**her**	**hers**	**girl**	**girls**	**female**
women	females	her	herself	her	girls	girl	women
female	minorities	he	she's	his	lady	bros	females
females	female	she's	she	vendetta	guy	hipsters	woman
minority	woman	his	wife	manager's	ladies	frat	male
male	minority	apologized	boss	my	kid	sorority	minority
men	males	who	his	she'll	woman	kids	minorities
minorities	men	gentleman	supervisor	retaliated	sorority	ladies	caucasian
caucasian	caucasian	replied	he	she	her	flirting	hispanic
misogynist	male	guy	him	she's	daughter	lady	white
males	women's	girlfriend	manager	him	dude	fraternity	sexist

Table A2b: Most Similar Words for Male Terms

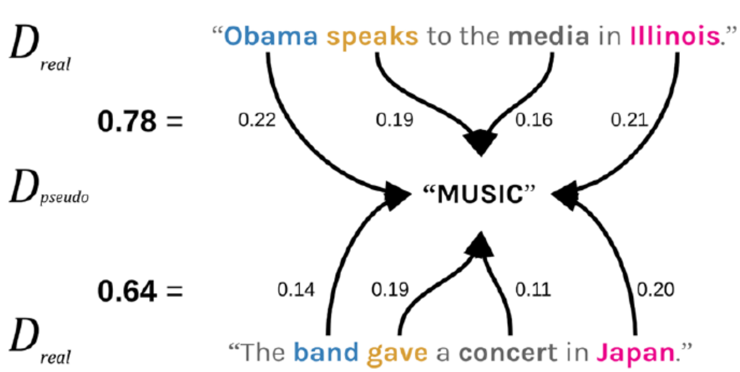
man	**men**	**he**	**his**	**boy**	**boys**	**male**
men	males	he's	her	boys	boy	males
guy	male	she	she	boy's	boy's	female
sir	man	guy	she's	good	good	men
woman	females	her	he	buddy	insiders	females
he's	women	him	owner	fashioned	bro	minority
dictator	woman	owner	herself	buddies	fashioned	caucasian
owner	minorities	she's	disagrees	sorority	buddies	woman
man's	female	ceo	micromanager	bro	men's	women
himself	sycophants	who	him	insiders	buddy	asian
person	sir	girlfriend	chairman	cronyism	adage	minorities

Figure A3: Word Mover's Distance Illustrated



Source: Stoltz and Taylor (2019: 296)

Figure A4: Concept Mover's Distance Illustrated



Source: Stoltz and Taylor (2019: 296)

Table A3: Coefficients for Industry, Region, and Year Fixed Effects for Table 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry: Electric/Electronic	0.00293***	0.00287***	0.00233***	0.00244***	0.00304***	0.00249***	0.00262***	0.00264***
Manufacturing	(0.000116)	(0.000117)	(0.000117)	(0.000116)	(0.000116)	(0.000117)	(0.000116)	(0.000116)
Industry: Enterprise Software	-0.00283***	-0.00278***	-0.00255***	-0.00212***	-0.00283***	-0.00260***	-0.00217***	-0.00216***
Network Solutions	(0.000077)	(0.000077)	(0.000077)	(0.000077)	(0.000077)	(0.000077)	(0.000077)	(0.000077)
Industry: IT Services	0.00221***	0.00225***	0.00229***	0.00234***	0.00230***	0.00232***	0.00237***	0.00244***
	(0.000072)	(0.000072)	(0.000072)	(0.000072)	(0.000072)	(0.000072)	(0.000072)	(0.000072)
Industry: Internet	0.00177***	0.00205***	0.00185***	0.00205***	0.00174***	0.00155***	0.00174***	0.00169***
	(0.000081)	(0.000081)	(0.000081)	(0.000080)	(0.000081)	(0.000081)	(0.000080)	(0.000080)
Industry: Cable Internet	0.00219***	0.00241***	0.00199***	0.00159***	0.00241***	0.00196***	0.00155***	0.00158***
Telephone Provider	(0.000133)	(0.000134)	(0.000135)	(0.000133)	(0.000134)	(0.000135)	(0.000133)	(0.000135)
Industry: Telecommunications	-0.000122	-0.000121	-0.0001505	0.000441	0.0000668	0.0000093	0.000632*	0.000818**
Manufacturing	(0.000292)	(0.000293)	(0.000292)	(0.000292)	(0.000292)	(0.000291)	(0.000291)	(0.000291)
Industry: Telecommunications	0.00299***	0.00316***	0.00225***	0.00182***	0.00308***	0.00213***	0.00170***	0.00177***
Services	(0.000095)	(0.000096)	(0.000097)	(0.000096)	(0.000095)	(0.000096)	(0.000096)	(0.000096)
Industry: Video Games	0.000229	0.000064	0.000095	0.00124***	0.000230	0.000260	0.00140***	0.00113***
	(0.000258)	(0.000259)	(0.000259)	(0.000258)	(0.000258)	(0.000258)	(0.000257)	(0.000258)
New England	-0.00180***	-0.00203***	-0.00191***	-0.00166***	-0.00186***	-0.00178***	-0.00150***	-0.00156***
	(0.000142)	(0.000143)	(0.000142)	(0.000142)	(0.000142)	(0.000142)	(0.000141)	(0.000141)
Mideast	-0.000460***	-0.000676***	-0.000514***	-0.000534***	-0.000502***	-0.000339***	-0.000363***	-0.000408***
	(0.000094)	(0.000095)	(0.000094)	(0.000094)	(0.000094)	(0.000094)	(0.000094)	(0.000094)
Southeast	0.000110	0.000043	0.000147	0.000035	0.000079	0.000183*	0.000064	0.000025
	(0.000087)	(0.000088)	(0.000087)	(0.000087)	(0.000087)	(0.000087)	(0.000087)	(0.000087)
Great Lakes	-0.000393***	-0.000545***	-0.000380***	-0.000377***	-0.000406***	-0.000244*	-0.000238*	-0.000255*
	(0.000109)	(0.000110)	(0.000109)	(0.000109)	(0.000109)	(0.000109)	(0.000109)	(0.000109)
The Plains	-0.0000341	-0.000153	-0.0001377	-0.000304	-0.0000644	-0.0000508	-0.000222	-0.000259
	(0.000161)	(0.000161)	(0.000161)	(0.000161)	(0.000161)	(0.000160)	(0.000160)	(0.000160)
Southwest	-0.000515***	-0.000667***	-0.000774***	-0.000769***	-0.000534***	-0.000644***	-0.000639***	-0.000663***
	(0.000097)	(0.000098)	(0.000098)	(0.000097)	(0.000097)	(0.000097)	(0.000097)	(0.000097)
Rocky Mountains	-0.00272***	-0.00303***	-0.00283***	-0.00268***	-0.00278***	-0.00257***	-0.00242***	-0.00246***
	(0.000142)	(0.000142)	(0.000142)	(0.000142)	(0.000142)	(0.000141)	(0.000141)	(0.000141)

Table A3: Coefficients for Industry, Region, and Year Fixed Effects for Table 3 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Far West	-0.00202***	-0.00231***	-0.00236***	-0.00234***	-0.00211***	-0.00216***	-0.00214***	-0.00224***
	(0.000070)	(0.000070)	(0.000070)	(0.000070)	(0.000070)	(0.000070)	(0.000070)	(0.000070)
2014	-0.00119***	-0.00126***	-0.0016***	-0.00189***	-0.00114***	-0.00155***	-0.00178***	-0.00175***
	(0.000139)	(0.000139)	(0.000139)	(0.000139)	(0.000139)	(0.000138)	(0.000138)	(0.000138)
2015	-0.0000925	-0.000146	-0.000485***	-0.000703***	-0.0000626	-0.000410**	-0.000628***	-0.000619***
	(0.000129)	(0.000129)	(0.000129)	(0.000128)	(0.000128)	(0.000128)	(0.000128)	(0.000128)
2016	0.000216	0.000193	(0.000119)	-0.000354**	0.000216	(0.000105)	-0.000342**	-0.000365**
	(0.000122)	(0.000122)	(0.000122)	(0.000122)	(0.000122)	(0.000122)	(0.000121)	(0.000121)
2017	0.000957***	0.000926***	0.000669***	0.000463***	0.000944***	0.000681***	0.000472***	0.000440***
	(0.000121)	(0.000121)	(0.000121)	(0.000121)	(0.000121)	(0.000120)	(0.000120)	(0.000120)
2018	0.00149***	0.00147***	0.00134***	0.00121***	0.00146***	0.00132***	0.00119***	0.00115***
	(0.000122)	(0.000123)	(0.000123)	(0.000122)	(0.000122)	(0.000122)	(0.000122)	(0.000122)
2019	0.00144***	0.00146***	0.00137***	0.00130***	0.001431***	0.00134***	0.00127***	0.00125***
	(0.000126)	(0.000127)	(0.000126)	(0.000126)	(0.000126)	(0.000126)	(0.000125)	(0.000125)

Table A4: OLS Regression Analysis of the Review-Level Gender Score, Including Reviews with No Reported Data on Employee Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (Female = 1)	0.00336***				0.00421***	0.00314***	0.00285***	0.00370***
	(0.000053)				(0.000137)	(0.000076)	(0.000095)	(0.000164)
Gender Missing	0.000165***				0.000840***	0.000282***	0.000301***	0.000835***
	(0.000043)				(0.000112)	(0.000061)	(0.000075)	(0.000133)
Satisfaction		0.000337***			0.000497***			0.000551***
		(0.000013)			(0.000022)			(0.000022)
Ownership Public			0.00297***			0.00284***		0.000088
			(0.000040)			(0.000064)		(0.000088)
Ownership Other			0.00233***			0.00235***		0.00111***
			(0.000079)			(0.000126)		(0.000129)
Firm Size Medium				0.00198***			0.00181***	0.00189***
				(0.000046)			(0.000076)	(0.000083)
Firm Size Large				0.00514***			0.00483***	0.00485***
				(0.000048)			(0.000076)	(0.000106)
Gender * Satisfaction					-0.000235***			-0.000228***
					(0.000037)			(0.000037)
Gender Missing * Satisfaction					-0.000196***			-0.000154***
					(0.0000294)			(0.0000294)
Gender*Ownership Public						0.000583***		0.000252
						(0.000109)		(0.000150)
Gender*Ownership Other						0.000217		0.000133
						(0.000215)		(0.000220)
Gender Missing*Ownership Public						0.000050		(0.000083)
						(0.000087)		(0.000121)
Gender Missing*Ownership Other						-0.000148		-0.0000814
						(0.000172)		(0.000177)
Gender * Firm Size Medium							0.000729***	0.000570***
							(0.000129)	(0.000140)
Gender Missing * Firm Size Medium							-0.0000363	-0.0000133
							(0.000102)	(0.000112)
Gender * Firm Size Large							0.000942***	0.000686***
							(0.000131)	(0.000181)
Gender Missing * Firm Size Large							0.000287**	0.000380**
							(0.000104)	(0.000145)
Constant	-0.0417***	-0.0424***	-0.0425***	-0.0434***	-0.0434***	-0.0430***	-0.0438***	-0.0459***
	(0.000079)	(0.000086)	(0.000073)	(0.000075)	(0.000110)	(0.000085)	(0.000091)	(0.000121)

Notes: N=946,653 reviews from January 2014 to September 2020. * indicates p < 0.05, ** p < 0.01, and *** p < 0.001.

Table A5: OLS Regression Analysis of the Review-Level Gender Score, Excluding Reviews from Employees in the Smallest Firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (Female = 1)	0.00356***				0.00434***	0.00345***	0.00356***	0.00420***
	(0.000066)				(0.000177)	(0.000121)	(0.000092)	(0.000202)
Satisfaction		0.000812***			0.000924***			0.000893***
		(0.000024)			(0.000029)			(0.000029)
Ownership Public			0.001616***			0.001632***		-0.000135
			(0.000071)			(0.000087)		(0.000101)
Ownership Other			0.001256***			0.00138***		0.000589***
			(0.000130)			(0.000158)		(0.000159)
Firm Size Large				0.00296***			0.00299***	0.00300***
				(0.000064)			(0.000078)	(0.000090)
Gender * Satisfaction					-0.000208***			-0.000200***
					(0.000049)			(0.000049)
Gender * Ownership Public						0.000270		0.000251
						(0.000147)		(0.000171)
Gender * Ownership Other						-0.000103		-0.0000858
						(0.000270)		(0.000273)
Gender * Firm Size Large							0.000178	0.000078
							(0.000132)	(0.000154)
Constant	-0.0413***	-0.0429***	-0.0411***	-0.0414***	-0.0446***	-0.0423***	-0.0427***	-0.0458***
	(0.000419)	(0.000428)	(0.000422)	(0.000420)	(0.000431)	(0.000423)	(0.000420)	(0.000433)

Notes: N=365,931 reviews from January 2014 to September 2020 from employees in large and medium-sized firms only (>10,001 employees and 201-10,000 employees, respectively). * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table A6: OLS Regression Analysis of the Review-Level Gender Score, Excluding Reviews with the Highest Satisfaction Ratings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (Female = 1)	0.00354***				0.00409***	0.00336***	0.00312***	0.00366***
	(0.000065)				(0.000164)	(0.000098)	(0.000131)	(0.000190)
Satisfaction		0.00175***			0.00186***			0.00170***
		(0.000027)			(0.000033)			(0.000033)
Ownership Public			0.00190***			0.00184***		-0.000400***
			(0.000065)			(0.000079)		(0.000107)
Ownership Other			0.00161***			0.00160***		0.000516***
			(0.000123)			(0.000150)		(0.000153)
Firm Size Medium				0.00119***			0.000991***	0.000754***
				(0.000081)			(0.000099)	(0.000106)
Firm Size Large				0.00359***			0.00342***	0.00285***
				(0.000081)			(0.000099)	(0.000132)
Gender * Satisfaction					-0.000134*			-0.000168**
					(0.000056)			(0.000056)
Gender * Ownership Public						0.000444***		0.000207
						(0.000134)		(0.000181)
Gender * Ownership Other						0.000131		0.000095
						(0.000254)		(0.000260)
Gender * Firm Size Medium							0.000547**	0.000491**
							(0.000167)	(0.000178)
Gender * Firm Size Large							0.000738***	0.000688**
							(0.000166)	(0.000223)
Constant	-0.0436***	-0.0468***	-0.0430***	-0.0437***	-0.0485***	-0.0443***	-0.0448***	-0.0489***
	(0.000155)	(0.000168)	(0.000156)	(0.000160)	(0.000177)	(0.000159)	(0.000166)	(0.000184)

Notes: N=367,000 reviews from January 2014 to September 2020. * indicates p < 0.05, ** p < 0.01, and *** p < 0.001. This table excludes reviews with the highest rating on satisfaction.

Table A7: OLS Regression Analysis of the Review-Level Gender Score, Excluding Reviews after February 2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (Female = 1)	0.00339***				0.00425***	0.00320***	0.00290***	0.00377***
	(0.000054)				(0.000141)	(0.000078)	(0.000098)	(0.000169)
Satisfaction		0.000393***			0.000512***			0.000569***
		(0.000018)			(0.000022)			(0.000022)
Ownership Public			0.00288***			0.00276***		0.000081
			(0.000054)			(0.000066)		(0.000090)
Ownership Other			0.00233***			0.00231***		0.00110***
			(0.000106)			(0.000129)		(0.000132)
Firm Size Medium				0.00202***			0.00173***	0.00182***
				(0.000064)			(0.000078)	(0.000085)
Firm Size Large				0.00494***			0.00470***	0.00474***
				(0.000065)			(0.000079)	(0.000109)
Gender * Satisfaction					-0.000236***			-0.000226***
					(0.000038)			(0.000038)
Gender * Ownership: Public						0.000536***		0.000241
						(0.000111)		(0.000154)
Gender * Ownership: Other						0.000129		0.000053
						(0.000219)		(0.000225)
Gender * Firm Size Medium							0.000720***	0.000567***
							(0.000132)	(0.000144)
Gender * Firm Size Large							0.000869***	0.000619***
							(0.000134)	(0.000185)
Constant	-0.0432***	-0.0433***	-0.0431***	-0.0439***	-0.0450***	-0.0443***	-0.0450***	-0.0471***
	(0.000314)	(0.000320)	(0.000314)	(0.000314)	(0.000323)	(0.000314)	(0.000315)	(0.000325)

Notes: N=520,282 reviews from January 2014 to February 2020. * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. This table excludes reviews from March 2020 onward (the covid era).