Shades of Masculinity? The Gendering of Discourse about Tech Firms

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

21 February 2023

Abstract

Gender inequality at work has persisted despite decades of protest and regulation. We study one factor that can create and maintain gender inequality at work: cultural conceptions of work and organizations that frame organizational goals, workers, and jobs as male-typed. Male-typed cultural conceptions of organizations, workers, and jobs raise questions about how well women fit into organizations and whether they are competent in many jobs, making it harder for them to thrive and driving them to exit. We focus on the tech sector because employees, journalists, and academics have revealed that tech has a gender problem in terms of demographics, culture, and practices. We observe how cultural conceptions of tech firms are male-typed by applying natural-language processing techniques (word embeddings) to employees' descriptions of their firms and deriving a gender cultural axis (male to female). We find associations between the gender cultural axis, on the one hand, and cultural conceptions of ideal workers and bosses and gender stereotypes, on the other. Some associations are as expected, but there are several surprises. Finally, we find that outcomes that are important to tech firms – innovation, speed, and performance – are generally male-typed.

[1] We thank Glassdoor.com for access to data, and Sameer Srivastava for facilitating an introduction to the firm. For funding, we acknowledge the Berkeley Culture Initiative and the Center for Equality, Gender, and Leadership at the Haas School of Business. We thank Minori Jaggia and Emily Guo for their programming skills, as well as Lily Bhattarjee, Grace Kull, and Sandeep Sainath for work on earlier parts of the analysis. Finally, comments from the Macro Research Lunch group were helpful in fine-tuning the analysis and improving the write-up.

Gender equality at work has long been a goal for activists and governments. But despite legislators passing statutes mandating equal opportunity, regulators and courts enforcing those statutes, and employers developing new policies and procedures to comply with those statutes, gender equality at work is far from being realized. This is especially true in some sectors like trucking and tech, and in well compensated jobs like management. In 2018, women constituted just 14% of the top managerial ranks in transportation, logistics, and infrastructure, and between 11% and 17% in computer software, computer hardware, IT services, and telecommunications (McKinsey and Company 2018).

A common explanation for gender inequality at work is male-typed cultural conceptions of jobs, workers, and organizations, which can erect barriers to gender equality by gender-typing jobs and workers, as well as entire organizations: distinctions between male and female determine what work means, who workers are, and who has power over whom (Kanter 1977; Acker 1990; Britton 2000; Heilman 2012). In most organizations, decision makers assume the "ideal worker" is male when they formally divide up tasks into jobs, describe jobs in postings and recruiting sessions, design work groups and departments, write rules and develop standard operating procedures, create evaluation and promotion criteria, and devise compensation and benefits packages. The results are male-typed cultural conceptions of organizations, jobs, and employees. Male-typed cultural conceptions of organizations are reflected in and reinforced by the patterned activities that develop through everyday interaction and decision-making.

When cultural conceptions of organizations are male-typed, they raise questions about how well women fit into organizations and whether they are competent in many jobs. Women tend to be less interested in applying for male-typed jobs (Gorman 2005). Moreover, because male-typed cultural conceptions of organizations value women less than men, women are less likely to be hired when they do apply for jobs in such organizations (especially for the most male-typed positions), less likely to feel comfortable and supported by bosses and coworkers, and less likely to be evaluated positively, compensated well, and offered good career opportunities; as a result, women are more likely to exit such organizations (Blair-Loy 2003; Gorman 2005; Turco 2010; Wynn and Correll 2018).

We focus on the tech sector: firms in industries with computing technology at their core, including hardware, software, network systems, internet services, and video games. This sector offers high pay because many employees are in professional occupations like computer science and engineering where talent is scarce. This sector also offers good opportunities for upward mobility and pay increases because many tech firms grow rapidly. But it has a persistent gender problem, with women under-represented among technologists and managers, especially in the top ranks (McKinsey and Company 2018), and with complaints about corporate cultures and workplace practices that disadvantage women (e.g., Fowler 2017; Chang 2018; Wynn and Correll 2018; Luhr 2020).

To investigate the cultural gendering of tech firms, we analyze the language employees use to describe their firms. Language has shadings of beliefs and values baked into it (Barley 1983). Because language is a medium for expressing fundamental assumptions, values, and norms, the language employees use to describe their firms can reveal their conceptions of organizations. For example, studying the language of performance evaluations reveals how gender stereotypes infuse managers' perceptions of workers (Correll, Weisshaar, Wynn, and Wehner 2020). Concretely, we measure how much employees use gendered language; i.e., language that reflects the gender cultural binary, male vs. female. We analyze data on over 900,000 employee reviews of firms from 2014 to 2020 from Glassdoor.com, an online job-search platform, and deploy natural-language-processing (NLP) techniques to parse the data. Tapping into these data makes it possible to develop an unobtrusive indicator of organizational culture, which is especially useful for creating valid and unbiased measures of aspects of culture that, like gender bias, is difficult to achieve with interviews or standard surveys (Reader, Gillespie, Haid, and Patterson 2020).

Following growing lines of research in computer science and computational linguistics (Bolukbasi, Chang, Zou, Saligrama, and Kalai 2016; Caliskan, Bryson, and Narayan 2017; Garg, Schiebinger, Jurafsky, and Zou 2018) and sociology (Kozlowski, Taddy, and Evans 2019; Arseniev-Kohler and Foster 2022), we train a word embedding model to calculate the direction of a gender axis in semantic space, a dimension of meaning running from male to female. We validate this axis by reading hundreds of reviews and by assessing semantic associations between the gender axis, gender-denoting terms, and common words in our corpus. We then explore relationships between the gender axis and three sets of concepts: cultural conceptions of employees and managers, gender stereotypes, and things that firms value, such as innovation and performance.

Our findings are sometimes as expected and sometimes surprising. First, in contrast with expectations about men as ideal workers and leaders, we found that words denoting these concepts were generally female-typed. Second, associations between the gender cultural axis and most concepts denoting gender stereotypes were as expected, but for some words denoting those concepts, associations were weak or opposite to expectations. And for two male-stereotype concepts, independence and instrumental competence, associations were surprising. Independence appeared to be gender-neutral, while instrumental competence was female-typed. Finally, the concepts denoting key firm values (innovation, speed, and performance) were largely male-typed, as expected. We continue to dig into the data to make sense of the surprising findings and further validate our measures and analytical techniques.

Theory

Gender is the biggest divide in many societies, including the U.S. For over 40 years, sociologists and management scholars have demonstrated that employing organizations can both generate and reduce gender inequality (Kanter 1977; Bielby and Baron 1986; Gorman 2005; Castilla 2015; Kelly and Moen 2020). Such work has focused on organizational structures (e.g., Baron, Davis-Blake, and Bielby 1986; Reskin 1993) and policies (e.g., Kelly and Dobbin 2009; Dobbin, Schrage, and Kalev 2015; Castilla 2015). Less studied is culture: widely shared understandings of what organizations do, what matters, and what is normal (for exceptions, see

Gorman 2005; Turco 2010; Wynn and Correll 2018). Of particular interest to us is the extent to which organizations are "gendered," meaning that cultural schemas about differences between men and women determine what work means, who workers are, who has status and power over whom, who fits into the organization and who does not, and how work should get done (Acker 1990; Britton 2000; Ridgeway 2011).

We begin by discussing cultural conceptions of gender in general. Next, we discuss how cultural conceptions of gender are invoked in the workplace specifically. We then explain the importance of language in understanding cultural conceptions of gender at work. Last, we introduce three aspects of workplace culture that we examine in relation to gender: ideal-typical conceptions of workers and bosses, gender stereotypes about employee attributes and behavior, and corporate values.

Cultural conceptions of gender

The most widespread cultural understanding of gender is that it is a binary concept, contrasting male with female. The male/female axis is just one of the many binary cultural schemas we use to classify and order the social world; other examples include young/old, happy/sad, fast/slow, good/bad, rich/poor, success/failure, and politically right/left. The axes defined by such schemas have neutral positions in the middle. For example, political independents are in the middle of the politically left/right axis, while "indifferent" is in the middle of the good/bad axis.

All cultural schemas are inherently relational: they consist of clusters of mental associations about individuals, groups, events, objects, and organizations that develop with experience and that provide default associations about the characteristics of individuals, events, and so forth (Carley and Palmquist 1992; Emirbayer 1997; Hunzaker and Valentino 2019; Boutyline and Soter 2021). Concepts in cultural schemas and the relationships between them are represented in the ways we speak and write. These relationships have meaning (content and context, in terms of surrounding words and symbols), sign (positive or negative), strength (strong or weak), and directionality (one-way or two-way).

We can examine relationships both within and between cultural schemas. For example, cultural conceptions of wealth are positively associated with ideas about success and privilege (Kozlowski et al. 2019). And the schema about youth vs. old age is often aligned with the schema about health vs. sickness, even though not all youth are healthy and not all old people are sick. Sometimes relationships between binary cultural schemas are unexpected. For instance, gender is associated with body weight: fat is not just a feminist issue, as the saying goes, but a widely perceived attribute of femaleness (Arseniev-Kohler and Foster 2022).

Relationships between cultural schemas can vary within societies. For example, for U.S. liberals, the poverty end of the affluence/poverty axis is associated with structural factors like racism; for U.S. conservatives, it is associated with individual factors like laziness (Hunzaker and Valentino 2019). Relationships between cultural schemas can also vary over time. For instance, over the twentieth century, 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).

Like other cultural schemas, our ideas about gender are relational and historically situated. Structural linguists have long argued that we cannot understand the concept "male" without reference to the concept "female" because these concepts are defined by their oppositional relationship (Lévi-Strauss 1963 [1983]). Moreover, gender is related to cultural conceptions of home and family, work, education, and culture, etc. For instance, after the Industrial Revolution, women were relegated to the family sphere, men to the work and public spheres – even though many women (e.g., in the U.S., black women) worked outside the home. And, like poverty/affluence, our ideas about gender are historically embedded: differences between what is considered "male" and what is considered "female" (in terms of both content and strength) depend on time and place. For instance, the increase in American women working outside the home that started in the 1970s has sometimes blurred the boundaries between gender roles, while other times invoking cultural backlashes that have brightened those boundaries (Faludi 1991).

Gender cultural schemas at work

Cultural conceptions of employing organizations encompass expectations and value judgements about organizational goals, the tasks employees do to achieve those goals, and the attributes and backgrounds of the employees who are best suited to those tasks. For some organizations, cultural conceptions tilt toward the male end of the gender cultural axis because of perceptions that core tasks are male-typed and men are the "natural" people to do those tasks. Organizations with male-typed cultures tend to be unfriendly toward and unsupportive of women, especially those who occupy managerial, professional, and technical positions, where men have traditionally been numerically dominant (Gorman 2005; Phillips 2005; Baron, Hannan, Hsu, and Koçak 2007). For example, in elite law firms, selection criteria for new associates varied in how much they included stereotypically masculine characteristics such as decisiveness and assertiveness versus stereotypically feminine characteristics such as friendliness and willingness to cooperate (Gorman 2005). The more law firms' selection criteria included stereotypically masculine characteristics, the fewer women were hired. In essence, then, organization can be gendered (i.e., male-typed) by framing ideal workers as male and core tasks as male-typed, thereby raising questions about whether women are competent in many jobs (Kanter 1977; Acker 1990; Britton 2000). Yet, not all organizations are strongly male-typed; instead, there is likely to be considerable variation in the gendering of cultural conceptions of organizations, workers, and work.

Expectations of what work is, how it should be done, and who should do it drive the development and maintenance of workplace policies, procedures, and practices, including recruiting, promotion, compensation, training, and scheduling. Workplace norms and practices may seem, on their face, to be gender-neutral, but they can have disparate impacts on male and female workers. For example, holding work-related social gatherings (e.g., meetings with clients or recruits) in strip clubs or male-restricted golf courses can make female employees feel excluded and disrespected – even denigrated. Seemingly more benign, but also generating disparate impacts on men and women, are such expectations as employees being available around the clock and such practices as routinely sending messages on evenings and weekends; they have disproportionately negative effects on women because women do more domestic labor than men, so working evenings and weekends is more difficult for female employees (Hochschild 1989; Jacobs and Gerson 2004).

Using language to capture gendered cultural schemas at work

To probe the norms, values, and related practices that can sustain (or reduce) gender inequality at work, we focus on the language employees use to describe their firms. Such language reveals employees' understandings of their firms and their experiences at work – their mental models (Carley and Palmquist 1992) or cognitive schemas (DiMaggio 1997; Hunzaker and Valentino 2019). Language allows 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, studying the language used in job postings reveals 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 infuse managers' perceptions of workers (Correll et al. 2020). Most germane to this analysis is that because language is a medium for expressing fundamental assumptions about how firms can and should operate, as well as corporate norms and values, the language employees use to describe their firms can reveal their conceptions of workplace cultures (Barley 1983; Van Maanen 1991; Corritore, Goldberg, and Srivastava 2020).

Analyzing the language employees use to describe their firms is an excellent way to create an unobtrusive indicator of organizational culture, one that is not prompted by researchers' theoretical interests. Unobtrusive data are especially useful for capturing elements of organizational culture that are socially contested or sensitive, such as gender biases (Reader et al. 2020). It is very difficult to study such topics using interviews or surveys of employees because of social-desirability bias. In contrast, when employees volunteer descriptions of their workplaces, especially in an anonymous forum, they are more likely to "tell it like it is," even when they bring up socially contested or sensitive topics.

While studying how organizational cultures are associated with the cultural conception of gender, we take into consideration two aspects of language: content and context. The *content* of language indicates what matters and how: words and phrases that are common are associated with central cultural elements, while words and phrases that are rare (or missing) are associated with peripheral (or foreign) cultural elements (Whorf 1956; Sapir 1958). That is not to say that all central cultural elements are valued; indeed, some that are highly central to an organization's culture may be contested or disdained by employees. If so, they will be prevalent *and* take the form of complaints about the prevailing culture and associated practices. For example, in a living-with-covid world, the idea that workers need to be onsite, rather than remote, may be prominent among managers but may be contested loudly and clearly by many workers, especially those with long commutes.

The meaning of language also derives from its *context* – nearby grammatical elements such as words and punctuation marks (Harris 1954; Firth 1957). For example, the word "bank" means the edge of a river when it is surrounded by words like fish, rapids, boat, and swim. In contrast, "bank" means a financial institution when it is surrounded by words like deposit, borrow, collateral, and assets. Most germane to our analysis is that the proximity of concepts to words that denote gender (e.g., "he," "woman," "male") determine how gendered those concepts are, while their proximity to words like "good" or "best" determines their valence and relationships with employee evaluations.

Empirical focus 1: ideal-typical workers and managers

Workers. As explained above, employing organizations' cultures can be gendered (i.e., male-typed) if they frame core tasks as male-typed and ideal workers as male. Rooted in the gendered division of labor, with men responsible for work and women for home life, many organizations design jobs and employee selection and evaluation systems with the assumption that men can give their all at work because they have wives who can deal with life outside of work (Acker 1990). Therefore, men are preferred for many jobs. In general, cultural schemas about differences between men and women determine what work means, who workers are, who has status and power over whom, who fits into the organization and who does not, and how work should get done (Acker 1990; Britton 2000; Ridgeway 2011). Male-typed organizational expectations about employees and the jobs they do make it difficult to conceive of women as competent in many jobs (Kanter 1977; Acker 1990; Britton 2000).

Our research site is the tech sector. In tech, the assumption that the ideal worker is male is bolstered by the sector's emphasis on science and technology. In the U.S., mathematics, which is foundational for science and technology, is perceived as a male domain and men are perceived as being better at mathematics than women – even though gender differences in school mathematics performance have declined over time, are small in magnitude, and sometimes favor women (e.g., Fennema and Sherman 1977; Hyde, Fennema, and Lemon 1990). Even today, most college students believe men are better at mathematics than women (Nosek, Banaji, and Greenwald 2002). This "math=male" stereotype and the psychological threat it poses for women (Steele 1997) reduces women's motivation to pursue STEM studies and train for STEM jobs (Correll 2001). It may also make jobs in STEM-focused firms, like those in the tech sector, less attractive, as women presume they do not fit in. This presumption that women do not fit well in the tech sector is reinforced by experience; i.e., by men's domination of tech jobs (McKinsey and Company 2018).

Managers. Even more than the role of employee, the role of manager is likely to be male-typed. When people "think manager," they "think male" (Schein and Davidson 1993; Schein 2001; Eagly and Karau 2002). Increases in the representation of women in the managerial ranks may weaken this "manager=male" stereotype, but men are still less likely to perceive of women as successful managers (Dueher and Bono 2006). This persistent "manager=male" stereotype is reinforced by experience: men continue to dominate the managerial ranks (U.S. Bureau of Labor Statistics 2020).

Empirical focus 2: gender stereotypes

Gender stereotypes, like all stereotypes, are cultural schemas: widely shared generalizations about men and women that are applied to individuals. They are both descriptive (explicating what men versus women are and do) and prescriptive (explicating what men versus women should be and do). The masculine stereotype tends to emphasize *agency*, while the feminine stereotype tends to emphasize *communality* (Ridgeway 2011; Heilman 2012).

Although researchers have long recognized that each stereotype can be broken down into components – e.g., achievement orientation, willingness to take charge, rationality, and autonomy for agency; concern for others, connections to others, deference, and emotional sensitivity for communality (Heilman 2012) – most scholars have assumed these components blend together, so they have investigated agency versus communality overall. But different studies have used different terms to capture agency and communality, and those terms connote different components of those two overarching concepts. Recognizing the heterogeneity in how gender stereotypes have been conceived of and measured prompted an investigation of their components (Hentschel, Heilman, and Peus 2019). This study found that, based on experimental subjects' responses, the concept of communality can be broken down into *concern for others*, *sociability*, and *emotional sensitivity*, while the concept of agency can be broken down into *assertiveness*, *independence*, *instrumental competence*, and *leadership competence*. We adopt this seven-part categorization scheme to find associations between each component of each stereotype's overarching concept, considered independently, and the gender cultural axis.

Among the female stereotype components, concern for others involves understanding others, being kind to them, and being considerate of their interests and feelings. Emotional sensitivity involves intuition and attention to sentiment. Sociability involves a relationship orientation, yielding collaborative behavior and an emphasis on communication. Among the male stereotype components, assertiveness involves dominance, forcefulness, and boldness. Independence involves self-reliance, taking responsibility, and autonomy. Instrumental competence involves diligence, hard work, and productivity. Finally, leadership competence involves having an achievement orientation, taking charge, and being persuasive.

A priori, we expect, based on previous research, that concern for others, sociability, and emotional sensitivity will be associated with the female end of the gender axis, while assertiveness, independence, and both forms of competence will be associated with the female end. These expectations are tentative, however, for three reasons: there has been no prior work analyzing these semantic relationships that we could use to build strong predictions; moreover, gender stereotypes vary across contexts and over time. In different contexts, people hold very different stereotypes about men and women. Across countries, people's beliefs about gender differences in aptitude for science, engineering, and math (fields at the core of our research site, the tech sector), as captured in implicit association tests, vary greatly (Nosek et al. 2009). Over time, based on U.S. public-opinion polls, women's perceived advantage in communality has increased, while men's perceived advantage in agency has remained constant (Eagly, Nater, Miller, Kaufmann, and Sczeny 2019). Surprisingly, men's perceived advantage in competence has eroded, as poll respondents after 2000 have increasingly perceived women to be more competent than men. Yet the evidence is mixed, as poll respondents have, over the same time period, come to perceive men and women to be equally competent.

Eagly et al. (2019) did not distinguish between instrumental competence (getting the job done) and leadership competence (achievement orientation and influence over others). Social-role theory (Eagly and Steffen 1984; Koenig and Eagly 2014) holds that stereotypes derive from the roles different groups play: historically, men and women have held different kinds of jobs, and many women have not done paid work outside the home, leading to shared understandings that men's and women's natures vary, and so what kind of work suits them varies. Building on this theory, we expect that because women's labor-force participation has increased to near-parity with men's – in 2019, women constituted 47% of the U.S. civilian workforce (U.S. Department of Labor 2022) – instrumental competence may be perceived as (nearly) gender-equal. But because women remain underrepresented in managerial and executive roles – in 2019, women constituted only 40% of managers in private-sector workplaces and only 29% of computer- and information-systems managers (U.S. Bureau of Labor Statistics 2020) – we expect that leadership competence will be perceived as more male than equal or female. Moreover, we study the tech sector, where women are underrepresented in core jobs – only 15% of engineers and 27% of workers in computer-related occupations in 2019 were female (U.S. Department of Labor 2022). Therefore, core tasks are likely to be male-typed, and core workers are likely to be male by assumption (Correll et al. 2020; Luhr 2020). Overall then, we cannot be sure that instrumental competence will be male-typed.

We investigate one final gender stereotype concept: *rationality*. This involves being logical, methodical, and analytical; it has long been assumed to be a male-typed characteristic (e.g., Heilman 2012). But it is not part of the seven-fold categorization scheme developed by Hentschel et al. (2019). The core work done by firms in our research site, the tech sector, is based on science and engineering, fields that prize logic and reasoning, so it makes sense to also investigate the association between rationality and the gender cultural axis.

Empirical focus 3: firm values

Beyond gendered conceptions of workers, bosses, and gender stereotypes, the next natural question is whether (and if so, how) employee discourse reflects gendered understandings of important outcomes such as employee job satisfaction, commitment, and turnover; organizational values and norms; organizational policies, procedures, and practices; and ultimately, individual and organizational performance. Because gendered discourse is a cultural phenomenon, we began by analyzing three very general corporate values: innovation and speed, which tech firms especially prize, and performance, which all firms value.

Innovation is critical for tech firms to stay close to the state of the art and outcompete their rivals, as hundreds of research studies show (e.g., Kogan, Papanikolaou, Seru, and Stoffman 2017). Indeed, the tech sector is defined by its dependence on continually inventing new technologies or improving existing ones to develop new products and processes or update existing ones. Tech insiders recognize the central importance of innovation to their firms' futures. Many tech firms proclaim innovation is a goal; for example, OpenAI's charter declares the firm "must be on the cutting edge of AI capabilities" (https://openai.com/charter, viewed 2023-01-17), while Microsoft's hiring site 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). Following the lauded footsteps of Bell Labs (which brought us the transistor and the Unix programming language, among other things) and XEROX PARC (which developed laser printing and the graphical user interface, among other things), many large tech firms have created "innovation labs"; for example, Google[x] (which created Google Glass) and Amazon Lab126 (which developed the Kindle) are tasked with "moonshot" research on far-edge technologies.

Speed is a less obvious target for analysis. We chose to examine speed because it is connected to innovation: coming up with new products and processes is necessary for tech firms, but they have to do this quickly enough to keep up with or – better – stay ahead of rivals. Being first to market with a new or improved product is a common goal among tech firms. First movers can establish their products as industry standards and create strong brand recognition (Lieberman and Montgomery 1988). In addition, first movers can also gain advantages over rivals when there are high switching costs. Although scholars have shown that first-mover advantages are far from assured, many business people and engineers believe it is likely, based on popular management books (e.g., Moore 2013). Therefore, we expect speed to be a common topic among tech leaders. Many tech leaders believe that the faster their firms get new products and processes finished and to market, the better their positions vis-a-vis competitors.

If the cultures of tech firms are male-typed, and if innovation and speed are central to their goals and values, then we expect these concepts will also be male-typed. But as with the components of gender stereotypes, there is little research on the gendered semantics of innovation or speed, so our expectations are tentative.

Next, we consider one thing that all firms value: *performance*. Performance is obviously critical for long-run survival. We focus on financial or economic performance, meaning profits and, for private-sector firms, returns to stock price. There are other important dimensions of performance that we do not study, including technological performance and sustainability. Technological performance will be largely captured in the analysis of innovation and speed. Sustainability, especially its social and environmental components, are newer concerns for many firms. It is only in the past decade that the shareholder-value logic, which holds that returns to shareholders in the form of profits and stock-price returns, has been challenged, in the face of evidence about economic inequality and climate change. But sustainability remains a highly contested value in the corporate (and political) world, so for the sake of conceptual clarity, we focus solely on financial or economic performance.

We expect that performance will be more strongly associated with the male end of the gender cultural axis than the female end. But no-one, to our knowledge, has investigated the gendered semantics of performance, so we lack deep theory or empirical evidence with which to develop strong predictions. Therefore, our expectation is again tentative.

Methods

Research site: Tech firms

We study tech firms, meaning those operating in industries with computing technology at their core, such as computer hardware and software, networking systems, and video games. Tech firms have long had a gender problem. Most basically, women are underrepresented in the technical and managerial ranks. For example, based on anonymized EEO-1 reports for 2017 from a sample of large publicly traded tech firms in the Bay Area, women constituted 15% of executives, 24% of managers, 28% of professionals, and 28% of all employees in the typical (median) firm (author's calculation, based on Rangarajan 2018). These numbers are much lower than numbers for the U.S. private-sector workforce overall, where women constitute 31% of executives, 40% of managers, 54% of professionals, and 48% of employees (https://smartasset.com/checking-account/women-in-the-workforce-glass-ceiling-2020).

The problem goes beyond mere numbers. In tech firms, the ideal worker is male, so women find it difficult to fit into these firms (Wynn and Correll 2018; Luhr 2020). Women in tech have complained loudly, clearly, and frequently about discrimination and harassment, and the "bro" culture that permeates the tech sector (e.g., Fowler 2017; Kolhatkar 2017). In some tech firms, it goes further, with employees using crude and chauvinistic language to deride women's appearance and denigrate their career potential, encouraging alcohol and drug consumption at social events for both employees and customers, and condoning (or at least ignoring) sexual harassment (Fowler 2017; Chang 2018).

Data sources

Our primary data consist of 948,785 employee reviews of firms in the tech sector from Glassdoor.com between 2014 to 2020. Tech tends to be over-represented in this data source (Karabarbounis and Pinto 2018), which is fine for our purposes. Glassdoor reviews contain ratings on a 1-5 scale. Overall ratings are required. Employees also have the option to provide ratings on specific topics: culture and values, diversity and inclusion, work-life balance, senior management, and career opportunities. The meat of these reviews consists of verbal descriptions of pros and cons, plus advice to management. These descriptions are open-ended: employees present their own perceptions of and reactions to their workplaces. To provide a concrete example, Figure 1 shows a typical recent review for Salesforce, a San

Francisco-based tech firm that provides customer-relations-management software to large and small businesses. The bottom half of the figure shows the optional topical ratings.

[Figure 1 about here]

Glassdoor has several advantages for research on organizational culture and workplace practices. The company serves primarily as a job-search platform, so it attracts a large group of workers – about 64 million unique users every month. Half of job-seekers use Glassdoor (DeMers 2014), so these reviews serve as signals to potential employees how they might feel about working for reviewed companies. Glassdoor has excellent coverage of many firms, especially large ones. Importantly, Glassdoor's coverage ranges far beyond large, publicly traded firms to include many small and medium-sized firms, as well as many privately held firms. Glassdoor allows people to review their employers anonymously, so reviews are not susceptible to bias stemming from fear of employer retribution (Marinescu et al. 2021). Given these advantages, it is not surprising that Glassdoor reviews are beginning to be used in studies of organizational culture (e.g., Schmiedel, Müller, and vom Brocke 2019; Corritore et al. 2020) and employee satisfaction with compensation policies (e.g., Farhadi and Nanda 2021; Storer and Reich 2021).

Glassdoor reviews have two possible downsides, however. The first may derive from sampling bias: 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 emotional reactions to their employers, the data might be biased both positively *and* negatively. This would result in extreme bimodal distributions, with many one- and five-starred reviews. This is seen in review data from platforms like Yelp (retail businesses such as restaurants and plumbers) and Rotten Tomatoes (films and TV shows). But because Glassdoor has a "give-to-get" model that requires users to provide reviews before gaining unlimited access to data to aid their own job searches (such other people's reviews, CEO approval ratings, salaries, and interview questions), these reviews are less likely to be biased in this way (Marinescu et al. 2020). As a result of this behind-the-scenes feature, Glassdoor reviews are likely to be representative of the experiences and perceptions of all employees. Indeed, the distribution 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.

The second downside of Glassdoor data may involve gender bias: reviews may disproportionately represent the perceptions and experiences of men. In our data, 38% of reviewers self-identified as male, 19% percent self-identified as female, and 42% gave no answer. Thus, of those who volunteered their gender identity, 33% were female, slightly more than the 28% of employees in the IT sector (McKinsey 2018). This suggests that these data do *not* disproportionately represent what male employees think.

Measures

We used natural-language-processing (NLP) techniques to analyze employees' descriptions of their firms. This approach goes beyond most previous research using ethnographic methods (e.g., Barley 1983; Van Maanen 1991; Turco 2010), surveys (e.g., Hofstede 1980; O'Reilly, Chatman, and Caldwell 1991), or administrative data like formal statements in annual reports or web sites (e.g., Nguyen, Nguyen, and Sila 2019; Haber 2021). 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 by survey participants, which may not mirror on-the-ground reality. Administrative data can also cover many organizations but may be purely symbolic or capture aspirations rather than reality. In contrast, our approach captures on-the-ground reality volunteered by employees, does so in rich detail, allows for variation across organizations, and covers many employees in many firms.

Applying NLP techniques requires thinking through not just research questions and the theories that may help answer them, but also the complex nature of language – spelling, punctuation, grammar and syntax, word forms (morphology), semantics, symbolic representations (e.g., using all capitals or italics for emphasis), and special characters (e.g., $ or

@) (Jurafsky and Martin 2023). The Technical Appendix describes the steps we took to prepare the data for analysis, as well as details on the specific analytical techniques we used.

Capturing the gendering of organizational discourse. For this task, we used *word embeddings*, a set of unsupervised machine-learning techniques whose underlying assumption is relational – word context determines meaning (Harris 1954; Firth 1957). The relational conception of meaning is congruent with sociological theories of cultural schemas (Kozlowski et al. 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 between 100 and 300 (Mikolov, Sutskever, Chen, Corrado, and Dean 2013; Jurafsky and Martin 2023). Words that share many contexts, like "engineer" and "lawyer," are located near each other in semantic space (i.e., their word vectors point in similar directions), and words that do not, like "engineer" and "frog," are located far away from each other (i.e., their word vectors point in different directions). Reducing complexity from a very large number of unique words to a smaller number of dimensions facilitates uncovering both attributional similarity between words (synonymy) and relational similarity (analogy) (Mikolov, Sutskever, et al. 2013; Mikolov, Chen, Corrado, and Dean 2013).

Because word embeddings capture relational meanings, they can be used in analogical reasoning (Mikolov, Chen, et al. 2013). For example, the analogical query "bad:worse::hard:?" should yield the answer "harder." With word-embedding models, analogical reasoning takes the form of arithmetic operations on word vectors. To continue the following example, vector(bad) – vector(worse) + vector(hard) ≈ vector(harder). Thus word embeddings can be used to assess to what extent the meaning of one word is associated with the meaning of another, even when those words do not have the same meaning. For example, because some jobs are culturally associated with men (e.g., engineer) and others with women (e.g., nurse), vector(engineer) – vector(man) + vector(woman) ≈ vector(nurse) (Garg et al. 2018). Most germane to our purposes is that word embeddings can be used to investigate semantic relationships between words denoting gender and word denoting other concepts. This is true even if concepts do not seem to be logically connected to gender, like class (Kozlowski et al. 2019) or body weight (Arseniev-Kohler and Foster 2022). We can also investigate relationships between words denoting gender and words denoting such things as jobs and tasks, as well as organizational goals, values, and norms.

The semantic relationships that word embeddings reveal 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 explicitly gendered terms (e.g., he, she, woman, man) and word vectors for concepts that are, on the surface, not gendered (e.g., literature, math, executive, wedding) are strongly correlated with those concepts' gendered associations as measured by the implicit association test (Caliskan et al. 2017). Words related to art and family tend to be associated with women, while words related to science and work tend to be associated with men. Although such cultural associations have been couched by some researchers in terms of bias in word embeddings (e.g., Bolukbasi et al. 2016), we interpret them as revealing widely shared cultural schemas. These schemas may be morally problematic (black race associated with poverty, white race with affluence) or morally neutral (flowers with pleasant, insects with unpleasant). They may also reflect facts on the ground, such as the great prevalence of women's names among elementary-school teachers and men's names among electricians. Therefore, we do not seek to debias word embeddings, such as by removing gender bias in job advertisements (Bolukbasi et al. 2016), although that is a worthy endeavor. Instead, we seek to exploit the semantic associations reflected in word embeddings to study widespread cultural schemas and norms (for similar analyses, see Caliskan et al. 2017; Garg et al. 2018; Arseniev-Kohler and Foster 2022).

Our analysis leverages the ability of word-embedding models to reveal relational similarity, in our case how close in semantic space words are to the male or female end of the gender cultural axis. To do this, we start with a list of antonym pairs used in previous research (Kozlowski et al. 2019) that denote the gender binary, such as man-woman, he-she, him-her, and male-female. Subtracting the word vector for a male term from the word vector for its antonymic female term yields a new word vector (a resultant) that captures their semantic difference; i.e., the cultural schema for gender (Caliskan et al. 2017; Garg et al. 2018; Kozlowski et al. 2019; Arseniev-Kohler and Foster 2022). In other words, this arithmetic operation yields a vector that points toward the female end of the gender cultural axis in semantic space and away from the male end. Many different antonym pairs capture gender. The meaning for each pair will vary because different words tend to be used in different ways (e.g., noun vs. pronoun, singular vs. plural) and in different contexts, so their individual meanings differ. Therefore, subtracting each male term from its female antonym will yield a slightly different resultant – a slightly different vector. To capture the gender cultural axis robustly, we averaged the resultants for all antonym pairs.

Figure 2a shows this conceptually (i.e., with hypothetical data). The X axis (showing the overall gender cultural axis) is oriented from male on the left to female on the right. Gender neutrality is in the middle. Individual terms like "she," "he," "female," and "male" that explicitly denote gender are listed along the surface of the sphere, reflecting the fact that word-embedding algorithms normalize word vectors so they are all the same length. The solid lines running from the midpoint of the gender axis to the surface are the word vectors for each term. The dotted lines are the resultants for each pair of terms.

[Figure 2 about here]

Although this application of word embeddings has been validated by computer scientists and sociologists, it is still unclear what the gender cultural axis means. To probe this more, we compared vectors for gender-denoting words to the gender vector, using cosine similarity scores. We also analyzed words denoting workers and bosses, gender stereotypes at work, and corporate values. This process yields each word's gender valence – the degree to which it is related to (i.e., shares an orientation to the meaning of) the female end 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.” Again, word locations denote associations with the gender axis. The gender axis points in the direction from male to female, so smaller (or more negative) cosine similarity scores represent stronger associations with the male end of the gender axis, while larger (or more positive) scores represent stronger associations with the female end.

To capture the sensitivity of these associations to sampling (i.e., the specific documents under study), we used bootstrapping methods. We created 20 subsample datasets through random sampling with replacement, each containing 90% of the full dataset. We trained a different word2vec model on each subsample dataset. Using the 20 sets of word embeddings from training these models, we then calculated the gender cultural axis and cosine similarity scores between the gender axis and each word of interest. Across those 20 subsample datasets, we calculated means and confidence intervals for cosine similarities for each word.

Measuring ideal-typical employees and managers. To investigate whether the ideal-typical conception of the tech worker is male, we created a list of terms that denote workers, both single words (e.g., worker, employees, coworker) and two-word phrases (e.g., team mates). Then, to investigate whether the ideal-typical conception of the manager is male, we created a list of words that denote managers (e.g., boss, supervisors, executives).

Measuring gender stereotypes. We followed recent research on the seven components of both gender stereotypes that were identified in recent research (Hentschel et al. 2019) and created lexicons for seven gender stereotype concepts. As explained above, the female stereotype included concern for others, emotional sensitivity, and sociability; the male stereotype included assertiveness, independence, instrumental competence, and leadership competence. We also created a lexicon for rationality, which is especially relevant to the tech sector and has been found to be male-typed (e.g., Heilman 2012). For all eight concepts, we compiled lists of terms, both single words and two-word phrases. For the seven concepts identified by Hentschel et al. (2019), we started with the terms listed in their paper; for rationality, we started with terms mentioned by Heilman (2012). We then added synonyms of those terms, synonyms of those synonyms, and antonyms of the antonyms of key words and their synonyms. After that, we added morphemes (different word forms) of lexicon words (e.g., adamantly for adamant; cordially for cordial). Then, to ensure that lexicons were coherent – i.e., all the words in each lexicon were semantically similar – we "trimmed" each lexicon by iteratively comparing cosine similarities. We did this in stages, starting in each stage by eliminating the word with the lowest cosine similarity score when compared to all the other words in the focal lexicon. We stopped when all remaining words had cosine similarity scores of 0.50 or more, which corresponds to an angle of 60°. Some of these words were rare; in order to procure robust results, the lexicon was limited to words that appeared in at least 10 of the 20 subsamples we analyzed.

Measuring corporate values. To generate lexicons of terms associated with these concepts denoting innovation, speed, and performance, we followed the same procedure: using synonyms, synonyms of synonyms, and antonyms of antonyms, as well as morphemes. Again, we ensured that these lexicons were coherent by "trimming" each lexicon by comparing cosine similarities. For each lexicon, all remaining words had cosine similarity scores of 0.50 or more. Again, analysis was limited to words that appeared in at least 10 subsamples.

Results: Understanding the Gender Cultural Axis

Because the idea of using language to capture gender as a cultural axis is new, we conducted several analyses of this measure. We began by assessing how each term in the antonym pairs used to create the gender cultural axis is related to that axis, based on cosine similarity scores between individual words and the gender axis. Figure 3 shows the results of this analysis. For each word denoting the male or female gender, the dot represents the mean across 20 random subsample estimates while the ends of the bars represent the 99% confidence interval.

[Figure 3 about here]

As expected, all words denoting the cultural concept *female* had positive cosine similarity scores, while all words denoting the cultural concept *male*, except one ("male"), had negative scores. Interestingly, all words denoting *female* were all similarly situated relative to the female end of the gender axis: their cosine similarity scores were close, as the coefficient of variation (standard deviation/mean) was 0.34. In contrast, the locations of the words denoting *male* varied greatly, as the coefficient of variation was 0.92. Two male-denoting words, "men" and "he," had very small negative cosine similarity scores, while one, "male," had a very small positive cosine similarity score. All three word vectors were less than 2 degrees from the middle of the gender axis, suggesting they were gender-neutral. Other words denoting the cultural concept *male*, especially "boy" and "boys," had larger negative cosine similarity scores, indicating that they were much closer to the male end of the gender axis.

The gender-neutrality of "male," "he," and "men" suggests that men are the default employees. Most tech workers assumed that others in their workplace are male. So when tech workers said "he," they were talking in gender-neutral terms. This suggests that tech workers' expectations about who belongs in tech firms were gendered, skewed toward the expectation that men are "natural" or "normal" tech employees. In contrast, when tech workers said "she" or "her," they used female-typed terms, indicating that they expected that women to be in female-typed jobs doing female-typed tasks in female-typed ways. (Related to this point, we describe our analysis of the male ideal-worker concept in the next section.)

To explore why "boys" and "boy" are especially strongly male-typed, we read through a random sample of 100 reviews that contained one or the other of those words. Most of the time when tech workers used these words, they were talking about "good old boys" or the "old boys club." (We show some examples below, in Table 2.) By using these phrases, tech workers were choosing to remark on the tendency they observed for men, especially those in top positions, to help each other out and restrict opportunities to members of their group.

We also investigated which words – beyond the words that specifically denote gender – were closest to both ends of the gender score using cosine similarity. We limited this analysis to words that appeared over 1,000 times in the corpus because those contribute most to the analysis. The top 30 results are shown in Table 1. The columns on the left focus on words associated with the male end of the gender cultural axis; those on the right, on words associated with the female end.

[Table 1 about here]

The top-30 male-associated words included two of the terms used to create the gender cultural 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. This fits with our reading through entire reviews, where the phrases "good ole boys club" and "good ole boy" were common, as in the examples in Table 2, discussed below. This fits also with the cosine similarity scores for "boys" and "boy," which were the closest male-denoting words to the male end of the gender cultural axis, as discussed above. Some words on the top-30 list ("entrepreneurial," "trenches," "arrogance," and "analysis") fit with the male stereotype of initiative, logic, emotional coldness, and hard work. But other top-30 words were not obviously gendered ("shareholders," "kool," "legacy," and "mentality"). This is to be expected because cosine similarity scores find words whose vectors point in the same semantic direction, meaning they are used in similar contexts.

The top-30 female-associated words included two of the terms used to create the gender cultural axis ("her," "female," and "woman"); "women" also ranked in the top 100. Some top-30 words represent roles and relationships ("assistant" and "coworker"), which fits with the gender stereotype that women at work are more collaborative than men. But surprisingly, some words ("supervisor" and "supervisors") were more strongly associated with female than male. Other 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). But many of the top-30 words were not obviously gendered (“myself,” “unprofessional,” “degrees,” and “fortunate”). Again, this is to be expected. Taken together, these results indicate that the word-level gender cultural vector and the review gender score were indeed capturing the gender cultural binary.

Validating the Review Gender Score

We read hundreds of reviews to validate the review gender score. Table 2 shows reviews that scored among the lowest (most male-typed, Table 2a) and the highest (most female-typed, Table 2b). Reviews with the lowest gender scores invoked management in general or specific top-management roles or used male terms such as “boys,” “boy,” and “man”; terms associated with male stereotype concepts like assertiveness or instrumental competence, such as “competitive,” “can do,” and “mission driven”. In contrast, reviews with the lowest (most female-typed scores) generally invoked terms that reflect female stereotype concepts like sociability, such as “friendly,” “helpful,” and “nice.”

[Table 2 about here]

The Gender Axis and Ideal-Typical Workers and Bosses

Figure 4 shows these results. It consists of two parts: The figure on the left shows associations between the gender cultural axis and words denoting employee(s); the figure on the right shows associations between the gender cultural axis and words denoting boss(es). Dots represent the mean across the 20 subsamples; lines represent 95% confidence intervals.

[Figure 4 about here]

Surprisingly, all words denoting the concept of workers, shown on the left, had positive cosine similarity scores, indicating that they are female-typed. One of these – employees – was essentially gender-neutral, as it was extremely close to zero and the 95% confidence interval straddles the zero line. Also surprisingly, most words denoting the concept of bosses, shown on the right, had positive similarity scores, again indicating that they were female-typed. One of these – executive – was essentially gender-neutral. Only one term denoting bosses –

executives – was clearly male-typed, with a negative cosine score. These results are surprising because previous research has found the expected male gender stereotypes for words denoting work (Caliskan et al. 2017). They are especially surprising given that our research site is the tech sector, which is heavily male-dominated. Yet our data show that tech employees conceived of both employees and supervisors as generally female.

We dug into the raw data to figure out these puzzling findings. First, we examined the 50 words most closely associated with "manager," "boss," and "worker," based on cosine similarity scores. In the list for "manager," there were two female-denoting words were but no male-denoting words. In the list for "boss," there were four female-denoting words, along with three male-denoting words. And in the list for "worker," there were no female- or male-denoting words. Second, we compared the number of reviews containing "he" versus "she" and "boss" or "manager" within the six-word window that our word2vec model used. We found that he+boss and he+manager were more common than she+boss and she+manager. Third, we read through dozens of randomly sampled reviews where "she" was located near "manager" or "boss." We found several instances where tech workers described female managers as positive representations of a company's increasing commitment to improving gender representation. But there were even more examples of reviews with male-typed depictions of managers. These findings do not explain exactly what is going on in the word2vec model, but they do suggest that tech workers' discussions of workers and bosses is less gendered than many observers would have expected.

We will continue to investigate these surprising results in more depth (alas, after we submit the paper to ASA). We will assess semantic associations between each gender-denoting word and each word denoting workers or bosses. We will also compare the frequencies of reviews containing all combinations of gender-denoting words and words denoting workers and bosses within the six-word window that our word2vec model used. Finally, we may code (using MaxQDA, a commonly used qualitative coding software) a random sample of reviews containing male-denoting versus female-denoting words located near words denoting workers or bosses.

The Gender Axis and Gender Stereotypes

Figure 5 shows these results. Figure 5a shows associations between the gender cultural axis and feminine-stereotype words (three figures); Figure 5a shows associations between the gender cultural axis and the masculine-stereotype words (five figures). For each word denoting a component of a gender stereotype, the dot represents the mean across 20 random subsample estimates and the ends of the bars represent the 95% confidence interval.

[Figure 5 about here]

Let us begin with female stereotype concepts in Figure 5a. The top left panel shows that 20 out of 27 words denoting *concern for others* (74%) had positive cosine scores, so they were associated with the female end of the gender cultural axis. Three words ("insight." "thoughtfully," and "gracious") had negative average cosine scores but they are essentially gender-neutral because their 95% confidence intervals around 20-subsample averages straddled the zero line. The four remaining words denoting concern for others were male-typed. The top right panel shows that only seven out of 13 words denoting *emotional sensitivity* (54%) were female-typed. Three words were gender-neutral, while three others were male-typed. The bottom panel shows that 16 out of 24 words denoting *sociability* (67%) were female-typed, three were gender-neutral, and five were male-typed. Together, these results suggest that when tech employees discussed their jobs and firms, they generally invoked all three female stereotype concepts in female-typed ways. Yet there is evidence that these concepts were not entirely female-typed, as a sizable minority of words denoting these concepts were gender-neutral or male-typed, especially for emotional sensitivity.

Now let us turn to male stereotype concepts in Figure 5b. The top left panel shows that 19 out of 28 words denoting *assertiveness* (68%) had negative cosine scores, so they were associated with the male end of the gender cultural axis. Two words ("dominant" and "stronger") were gender-neutral, while seven were female-typed. Surprisingly, the top right panel shows that eight out of 15 words denoting *independence* (53%) were gender-neutral, not male-typed as previous research would predict. Only one word ("autonomy") was male-typed, and six words are female-typed. This suggests that when tech employees discussed aggressiveness in the context of their workplaces, they conceived of it as a trait that men generally (but not exclusively) exhibited. But when they discussed independence, they conceived of it in a gender-neutral way or sometimes as a trait that women exhibited.

The bottom left panel shows that only 15 out of 36 words denoting *leadership competence* (42%) were male-typed. Nine (25%) were gender-neutral, while 12 (33%) were female-typed. The bottom right panel shows that 25 out of 38 (66%) of words denoting *instrumental competence* were female-typed, not male-typed, as previous research would predict. Only six words (16%) were male-typed, while 7 (18%) were gender-neutral. Together, these results indicate that when tech employees discussed getting things done (instrumental competence) or success at leading teams (leadership competence), they conceived of the former as something that both men and women did, and the second as something that women were more likely to do than men. The finding on instrumental competence may seem surprising, given decades of analysis using experimental subjects to label terms as male- or female-typed. But it is congruent with recent research showing that that women are more likely to be judged by experimental subjects as competent (Eagly, Nater, Miller, Kaufmann, and Sczeny 2020; Bongiorno, Bain, Ryan, Kroonenberg, and Leach 2021).

We finish the analysis of gender stereotype concepts by turning to the concept of *rationality*, which is highly relevant to the tech sector. The panel on the next page shows that 11 out of 19 words (58%) denoting rationality were clearly male-typed. Among the other words denoting rationality, four were gender-neutral and four were female-typed (21% each). This set of results indicates that rationality was sometimes, but not always, conceived as a trait that men possessed.

Overall, these results suggest that there has been some progress in erasing some gender stereotypes from the discourse about jobs and workplaces – even in the tech sector, which has long been heavily male-dominated. Among the concepts we examined, even the most clearly gendered – *concern for others* – had only 74% of the expected associations with the gender cultural axis. Notably, *assertiveness*, which we expected would be male-typed, based on previous research, was instead gender-neutral. Moreover, the concept of *instrumental competence* was unexpectedly female-typed.

Of course, these results depend on the lexicons we developed. We started with core terms, added synonyms and synonyms of synonyms (and antonyms of antonyms), then trimmed the list by looking at semantic similarity overall. An alternative, which we will implement next, is to start with words that are core to the concept under study and add only those words that have strong semantic similarity to the core set. We will also validate our lexicons through human coding – the gold standard for dictionary methods in NLP (Grimmer and Stewart 2015; Nelson, Burk, Knudsen, and McCall 2021).

The Gender Axis and Firm Values

To investigate tech firms' values, we focused on three concepts: innovation, speed, and performance. The first two concepts are important determinants of performance and competitive position for tech firms in particular because they must constantly update products in order to stay ahead of their rivals. The third concept is important for all firms. Figure 6 shows the words associated with these concepts and their associations with the gender axis; it has three parts, one for each concept. In each part of this figure, words are arranged in ascending order of cosine similarity, from those closest to the male end of the gender axis to those closest to the female end. For each word, the dot represents the mean across 20 random subsample estimates while the ends of the bars represent the 95% confidence interval.

[Figure 6 about here]

The top left panel shows that all words denoting *innovation* except two (“creative” and “pioneer”) were closer to the male end of the gender axis than the female end. These results suggest that in tech firms, innovation was usually a male-typed value. Employees seemed to perceive that innovation is more likely to be what men do than what women do. The top right panel shows that just over half (53% or 8/15) of words that denote *speed* except one (“swift”) were more strongly associated with the male end of the gender axis, and just under half (47% or 7/15) were gender-neutral. These results indicate that as a corporate value, speed was sometimes male in connotation, although much less so than innovation. Men were conceived by tech employees as only a little more likely than women (or anyone) to achieve rapid results.

The bottom panel examines the gendering of the concept *performance*. It shows that 78% of words denoting performance (14/18) were clearly male-typed. One word denoting performance (“accomplishment”) was gender-neutral. And three (“achievements,” “accomplishments,” and “succeeded”) were female-typed. This suggests that this general corporate goal was somewhat male-typed: tech-firm employees usually conceived of performing well as something that men do more than women; they rarely conceived of it as something that women do more than men.

Discussion and Conclusion

We investigated one subtle but powerful attribute of employing organizations that can contribute to gender inequality at work: cultural conceptions of work and organizations that frame workers, jobs, and goals as male-typed. Male-typed cultural conceptions of work and organizations raise questions about how well women fit into organizations and whether they are competent in many jobs, which makes it harder for them to thrive and drives them to leave (Blair-Loy 2003; Gorman 2005; Turco 2010; Wynn and Correll 2018). Our analysis focused on the tech sector because employees, journalists, and academics have revealed that tech has a gender problem in terms of demographics, culture, and practices (e.g., Fowler 2017; Chang 2018; McKinsey and Company 2018; Wynn and Correll 2018; Luhr 2020).

To observe the extent to which cultural conceptions of tech firms are male-typed, we applied natural-language processing techniques, specifically word embeddings, to employees' descriptions of their firms from Glassdoor.com. Word-embedding models, which take both the content and context of language into account, map words onto points (represented as vectors) in a high-dimensional semantic space. Word-embedding models yield insights into employees' perceptions of their firms' cultures and the everyday practices that reflect and reinforce those cultures (Corritore et al. 2020).

We trained a word2vec model on the Glassdoor data. We then leveraged a well-known property of word vectors, that they capture both semantic and relational similarity, to derive a gender cultural axis (male to female) using vector arithmetic. Concretely, we subtracted word vectors for each word denoting "male" from the word vector for its antonym denoting "female," yielding a series of vectors mapping onto the binary concept of gender. We then calculated the centroid (mean) of these gender vectors to end up with a robust vector: the gender cultural axis, which runs from the location of the male end of the gender cultural binary in meaning space to the female end. To validate this measure, we probed relationships between the gender cultural axis and words denoting gender; we also investigated associations between the gender cultural axis and common words in our corpus.

We then turned to investigate the gendering of three empirical phenomena: cultural ideal types for workers and bosses, components of gender stereotypes, and things that tech firms value. Surprisingly, we found that cultural conceptions of both workers and bosses to be female-typed. Probing these findings suggested that tech workers' discussions of workers and bosses is less gendered than many observers would have expected. But so far, we have not figured out exactly why we see these patterns. We also found a mix of surprising and expected associations between the gender cultural axis and gender stereotypes. Among the surprising associations, the concept of "independence" was conceived of as gender-neutral, rather than male-typed. Finally, we find that things that are important to tech firms – innovation, speed, and performance – are generally (although not always strongly) male-typed.

Next steps. Before submitting this paper to a journal for review, we will do more validation checks of the gender axis and of the lexicons for words of theoretical interest, especially those for gender stereotype concepts. We will also investigate further the unexpected female-typed associations with words denoting workers and bosses. Beyond what is discussed here, we also plan to analyze the gendering of discourse for *entire reviews*. This will involve aggregating data from the level of the word to the level of the document. For this, we have identified a new technique that handles binary concepts like gender very well. A review-level measure will allow us to predict which tech firms, in which locations and times, have more or less gendered cultures, as evidenced by the way their employees describe their jobs and workplaces.

References

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

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., 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.

Baron, James N., Michael T. Hannan, Greta Hsu, and Özgecan Koçak. 2007. In the company of women: Gender inequality and the logic of bureaucracy in start-up firms. *Work and Occupations*, 34: 35-66.

Bielby, William T., and James N. Baron. 1986. Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology*, 91: 759-799.

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

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *30th Conference on Neural Information Processing Systems*, Barcelona, Spain.

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. (available at 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.

Castilla, Emilio J. 2015. Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science*, 26 (2): 311-333.

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

Correll, Shelley J. 2001. Gender and the career choice process: The role of biased self-assessments. American Journal of Sociology, 106: 691-730.

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.

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.

Duehr, Emily E., and Joyce E. Bono. 2006. Men, women, and managers: Are stereotypes finally changing? *Personnel Psychology*, 59: 815-846.

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.

Eagly, Alice H., and Valerie J. Steffen. 1984. Gender stereotypes derive from the distribution of men and women into social roles. *Journal of Personality and Social Psychology*, 46 (4): 735-764.

Emirbayer, Mustafa. 1997. Manifesto for a relational sociology. *American Journal of Sociology*, 103: 281-312.

Faludi, Susan. 1991. *Backlash: The Undeclared War Against American Women*. New York: Crown Publishing.

Farhadi, Reza, and Vikram Nanda. 2021. What do employees know? Quality perception and "over-satisfaction" in firms going public. *Journal of Finance*, in press.

Fennema, Elizabeth, and Julia A. Sherman. 1977. Sexual stereotyping and mathematics learning. *Arithmetic Teacher*, 24: 369-372.

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.

Fiske, Susan T., Jun Xu, Amy J.C Cuddy, and Peter S Glick. 1999. (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, 55: 473-489.

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, and Peter Glick. 2007. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Science*, 11 (2): 77-83.

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.

Grimmer, Justin, and Brandon Stewart. 2013. Text as data: The promises and pitfalls of content analysis methods for political texts. *Political Analysis*, 21 (3): 267-297.

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.

Haber, Jaren R. 2021. Sorting schools: A computational analysis of charter school identities and stratification. *Sociology of Education*, 94 (1): 43-64.

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

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.

Hochschild, Arlie Russell. 1989. *The Second Shift: Working Parents and the Revolution at Home*. New York: Avon Books.

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

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

Hyde, Jane Shibley, Elizabeth Fennema, and Susan J. Lamon. 1990. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107 (2): 139-155.

Jacobs, Jerry A., and Kathleen Gerson. 2004. *The Time Divide: Work, Family, and Gender Inequality*. Cambridge, MA: Harvard University Press.

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

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

Karabarbounis, Marios, and Santiago Pinto. 2018. What can we learn from online job postings? Evidence from Glassdoor. *Economic Quarterly*, 104 (4): 173-189.

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.

Koenig, Anne M., and Alice H. Eagly. 2014. Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, 107 (3): 371-392.

Kolhatkar, Sheela. 2017. The tech industry's gender-discrimination problem. *New Yorker*, Nov. 13.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics*, 132 (2): 655-712.

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.

Lagos, Danya. 2022. Has there been a transgender tipping point? Gender identification differences in U.S. cohorts born between 1935 and 2001. *American Journal of Sociology*, 128 (1): 94-143.

Lagos, Danya, and D'Lane Compton. 2021. Evaluating the use of a two-step gender identity measure in the 2018 General Social Survey. *Demography*, 58 (2): 763-771.

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.

Luhr, Sigrid W. 2020. Engineering inequality: Rethinking Inclusion in the San Francisco Bay Area Tech Industry. Unpublished dissertation, U.C. Berkeley.

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.

McKinsey and Company. 2018. *Women in the Workplace 2018*. (https://womenintheworkplace.com/2018; retrieved 2020-10-22)

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

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. 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.

Moore, Geoffrey A. 2013. *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Consumers, 3rd Ed*. New York: Harper Business.

Nelson, Laura K., Derek Burk, Marcel Knudsen, and Leslie McCall. 2021. The future of coding: A comparison of hand-coding and three types of computer-assisted text-analysis methods. *Sociological Methods and Research*, 50 (1): 202-237.

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.

Nosek, Bryan A., Mahzarin R. Banaji, and Anthony G. Greenwald. 2002. Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, 83: 44-59.

Nosek, Bryan A., et al. 2009. National differences in gender-science stereotypes predict national differences in science and math achievement. *Proceedings of the National Academy of Science*, 106 (26): 10,593-10,597.

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.

Rangarajan, Sinduja. 2018. Here's the clearest picture of Silicon Valley diversity yet: It's bad. But some companies are doing less bad. (https://www.salon.com/2018/06/29/heres-the-clearest-picture-of-silicon-valleys-diversity-yet-it-is-bad_partner/; viewed 13 June, 2020)

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.

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

Schein, Virginia E., and Marilyn J. Davidson. 1993. Think manager, think male. *Management Development Review*, 6 (3): 24-28.

Schein, Virginia E. 2001. A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, 57: 675-688.

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.

Steele, Claude M. 1997. A threat is in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52: 613-629.

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.

Turco, Catherine J. 2010. Cultural foundations of tokenism: Evidence from the leveraged buyout industry. *American Sociological Review*, 75: 894-913.

U.S. Bureau of Labor Statistics. 2020. Women in the labor force: A databook. (https://www.bls.gov/opub/reports/womens-databook/2020/home.htm; viewed 2023-01-16)

U.S. Department of Labor. 2022. Women in the labor force. (https://www.dol.gov/agencies/wb/data/facts-over-time/women-in-the-labor-force; viewed 2023-01-16)

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.

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.

Figure 1: Example Employee Review from Glassdoor.com

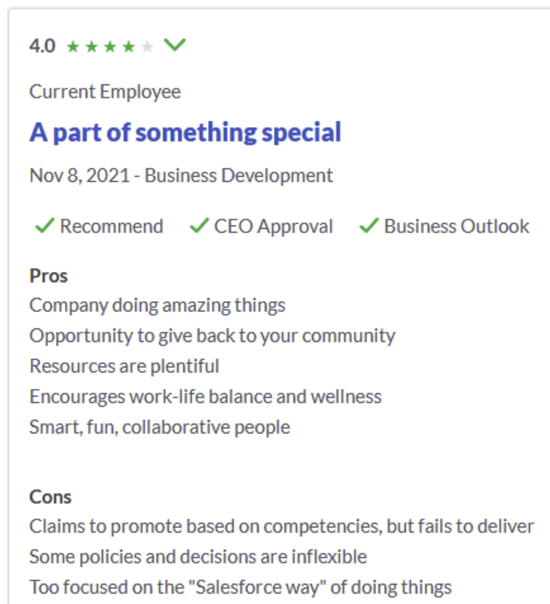


Figure 1a: Example Employee Review from Glassdoor.com: Optional Ratings on Topics

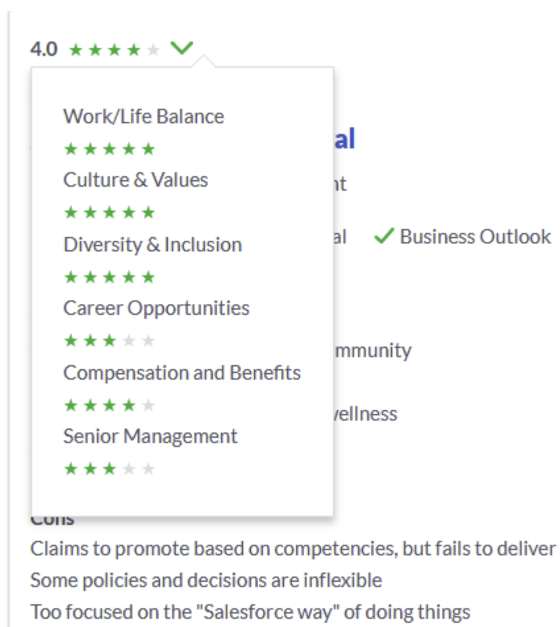


Figure 2a: Conceptual Diagram of Gender Word Embeddings and Resultants

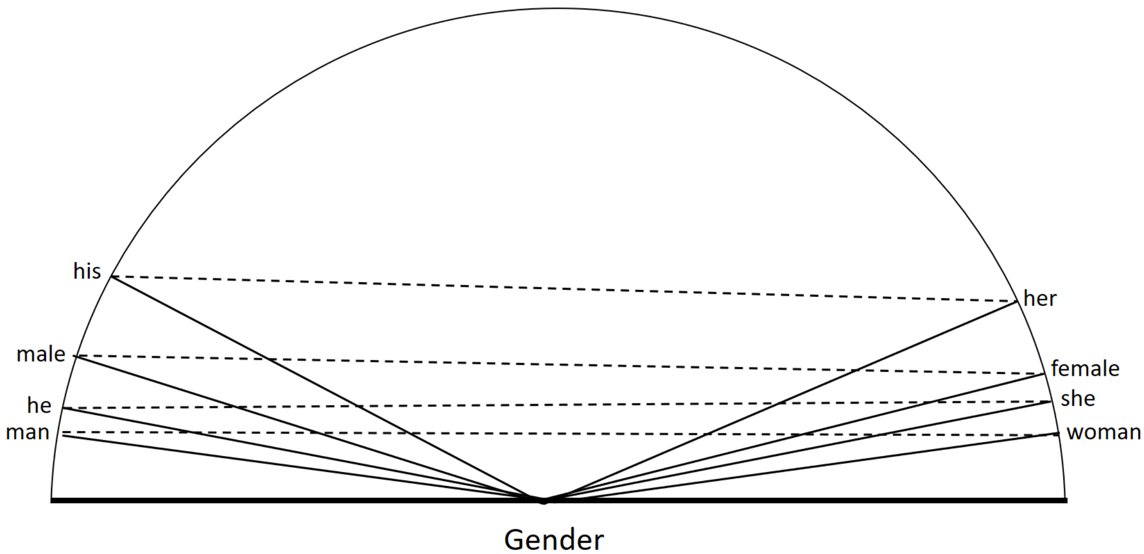
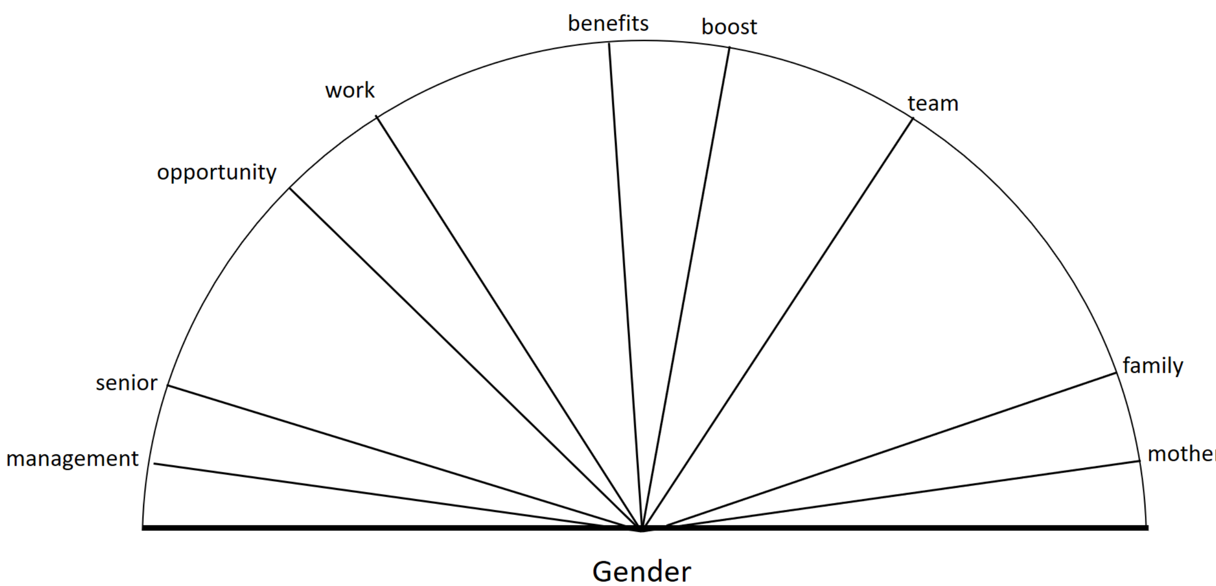
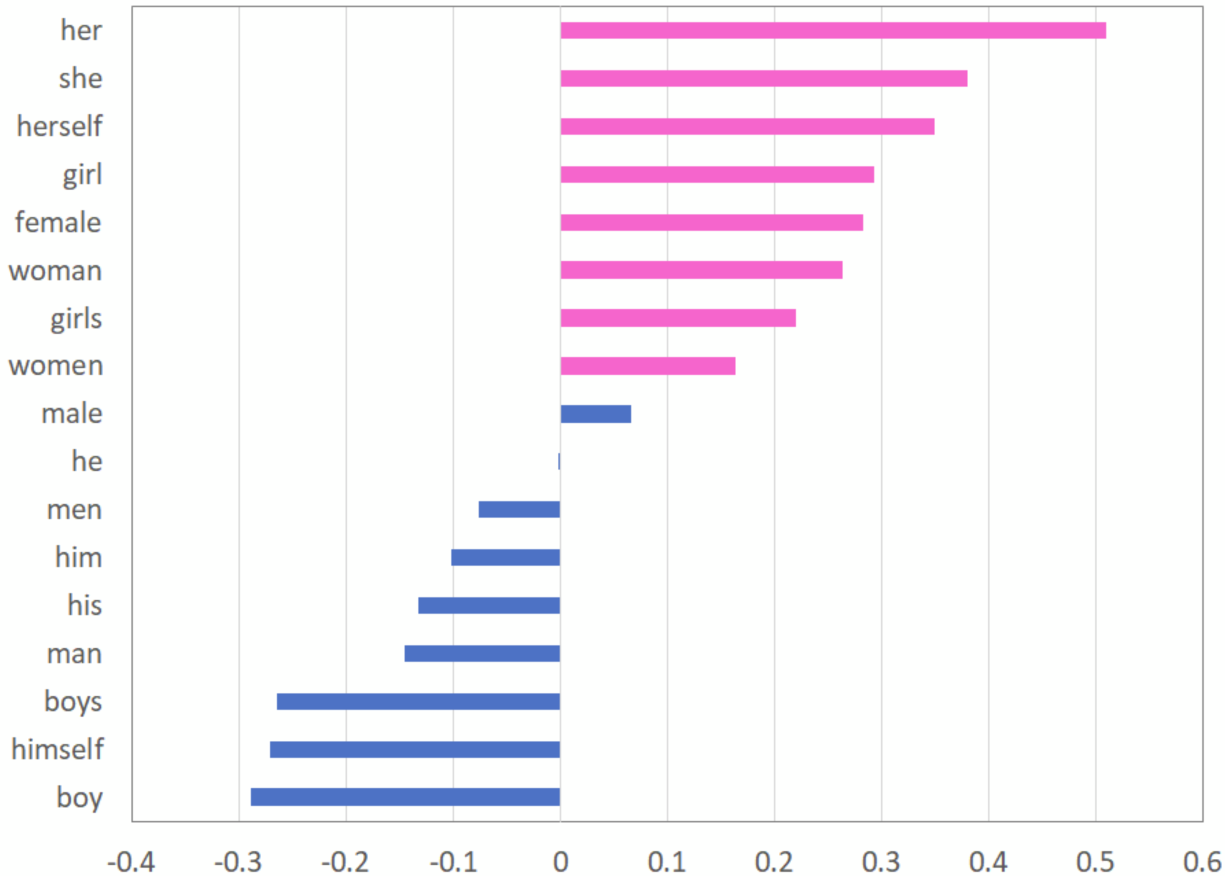


Figure 2b: Conceptual Diagram of Gender Cultural Axis and Words Describing Tech Firms



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.
The mean for “male” is 0.066; the 99% confidence interval runs from 0.039 to 0.093.

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

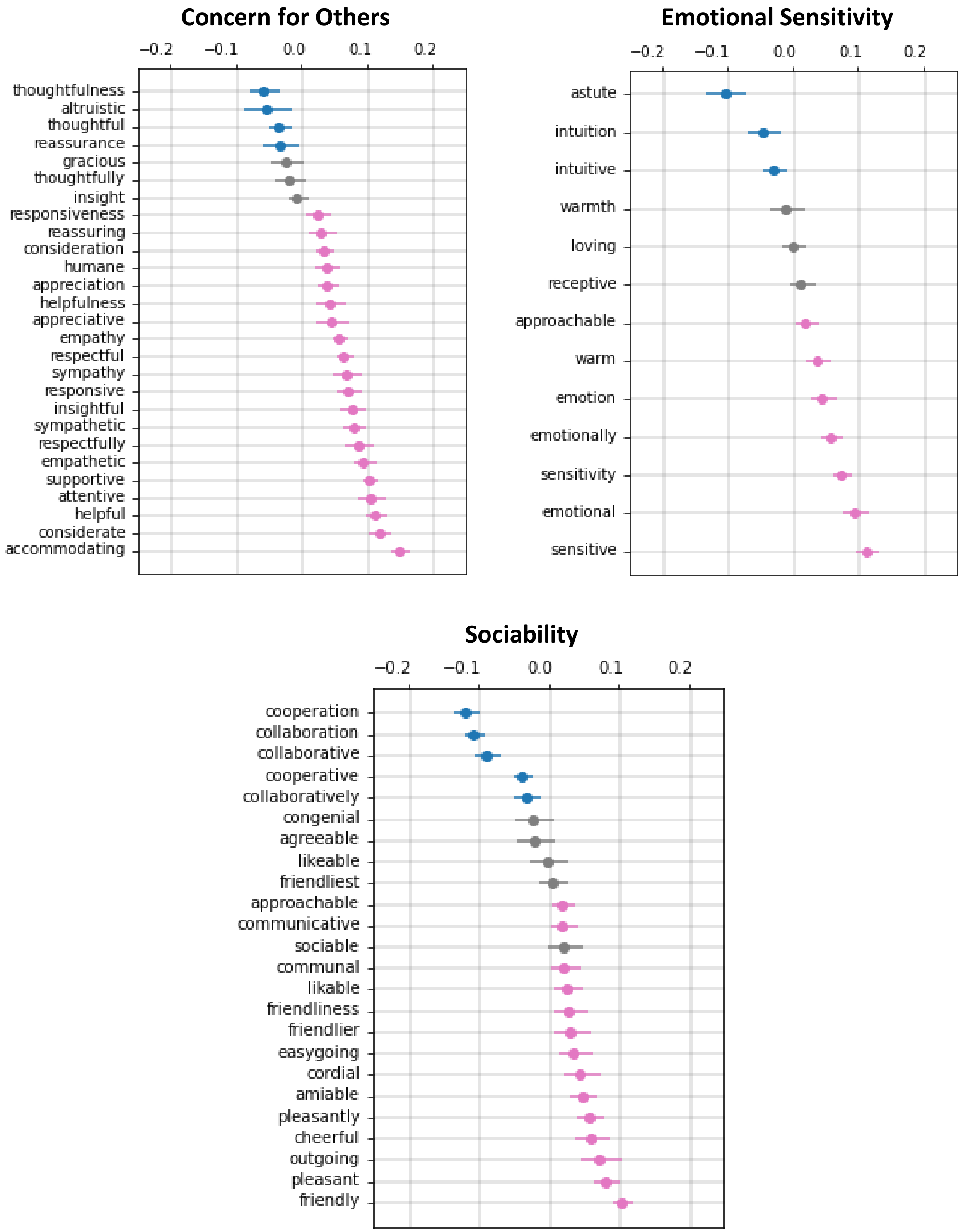
Table 2a: Reviews with the Highest Gender Score (Most Male-typed)
CEO/CFO are out from themselves. CEO doesn't care about stakeholders other then himself. You are phonies !
mission driven, passionate, and competitive comp Senior management appears to be an 'ole boys club'
Fun place to work at Management is a good ol boys club
The management had great vision for BMC products Need more innovation to integrate all BMC products
Good manufacturer in hw and disk drives slow adapting in software and innovation adapt and innovate
Startup attitude with can do mentality Need money, former CEO was a disaster Change management
If your a yes man and a good ole boy, your in good. If your not a yes man and not a good ole boy, your in trouble. The dinosaurs died out a long time ago!!
Great people, many with deep experttise Entrenched cronyism, old boy network Shape up !
Good manufacturer in hw and disk drives slow adapting in software and innovation adapt and innovate

Table 2b: Reviews with the Lowest Gender Score (Most Female-typed)
great coworkers all age levels and experience training is handled by coworkers
Atmosphere was calm treated like a full time employee even though I was a intern lunch voucher pay was great as a intern contract ended short no available position to become a full time employee
Pretty chill job with great coworkers This position was a contracted position
My supervisors were incredibly helpful, and the work I got to do was interesting. There were times when I felt like I was overqualified for the work, as I got assigned some administrative tasks.
Flexible work time, salary, friendly co-workers health benefits, personal assistant job instead of Executive Assistant
my boss was nice to me and my coworkers nothing, really, it was all great
Consist work, and friendly staff. Found it difficult to aquire information about myself from my contractor and the location i was working at.
Easy interview questions asked during interview. Interviewer is very friendly and nice.
The staff is friendly and helpful. The staff is small but helpful.

Figure 4: Associations between the Gender Axis and Words Denoting Workers and Bosses

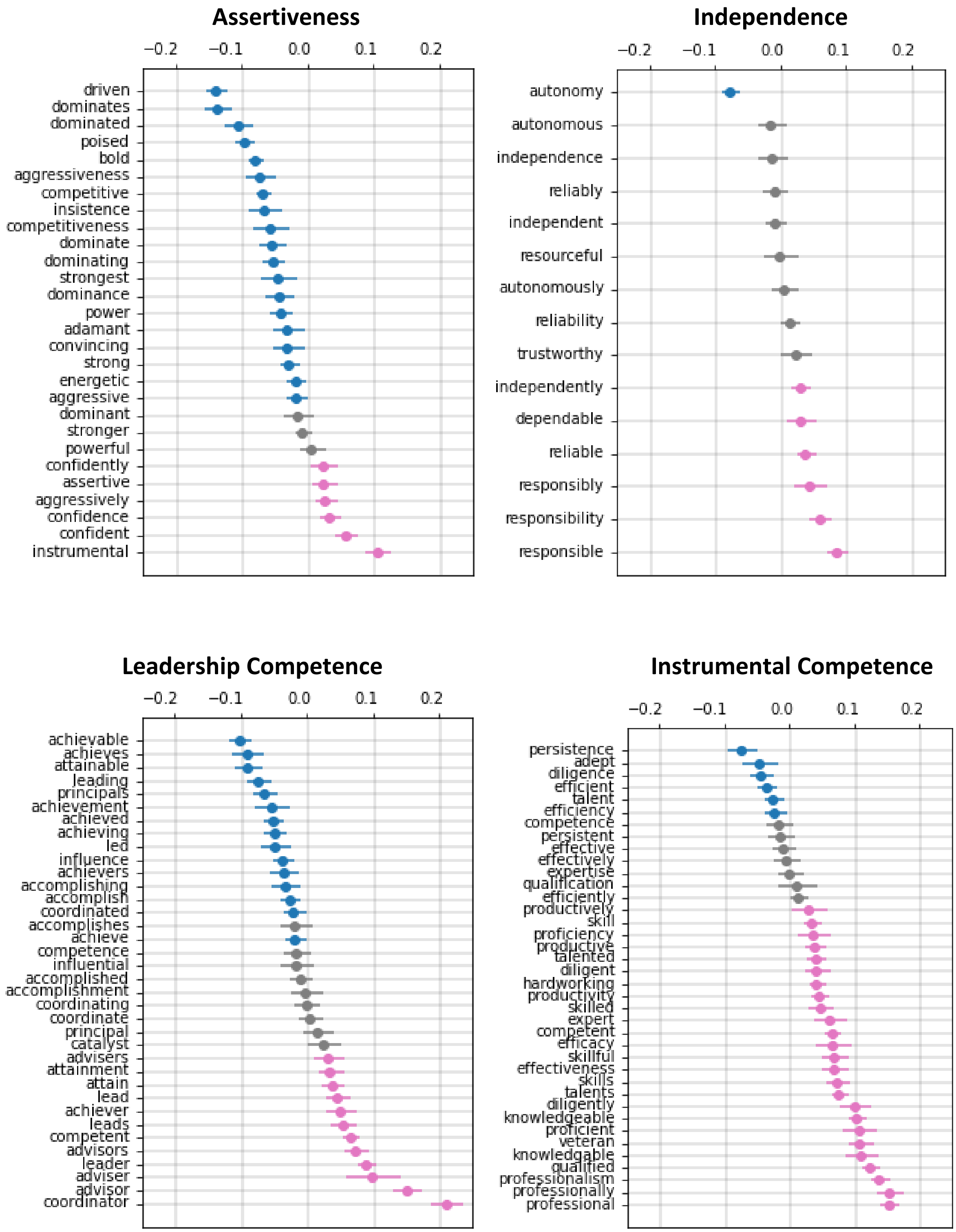
In these figures, the dots represent means across 20 subsamples, while the ends of the bars represent 95% confidence intervals. Blue indicates words that are male-typed, gray indicates words that are gender-neutral, and pink indicates words that are female-typed.

Figure 5a: Associations between the Gender Axis and Words Denoting Female Stereotypes



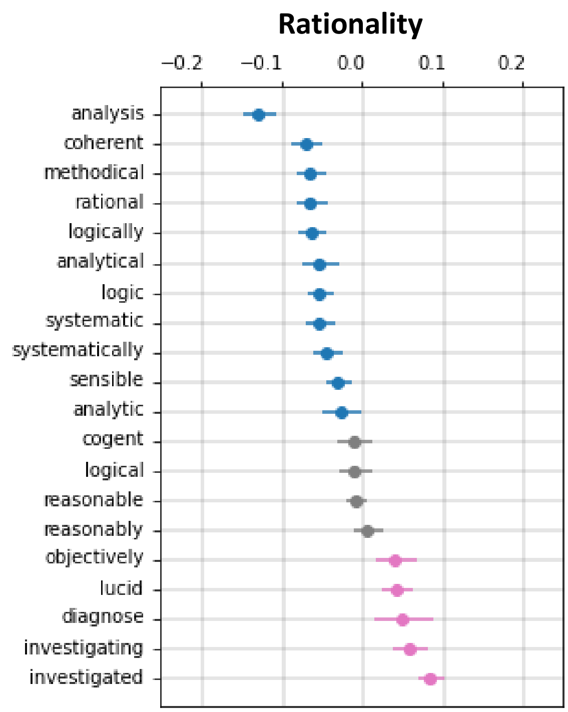
In these figures, the dots represent means across 20 subsamples, while the ends of the bars represent 95% confidence intervals. Blue indicates words that are male-typed, gray indicates words that are gender-neutral, and pink indicates words that are female-typed.

Figure 5b: Associations between the Gender Axis and Words Denoting Male Stereotypes



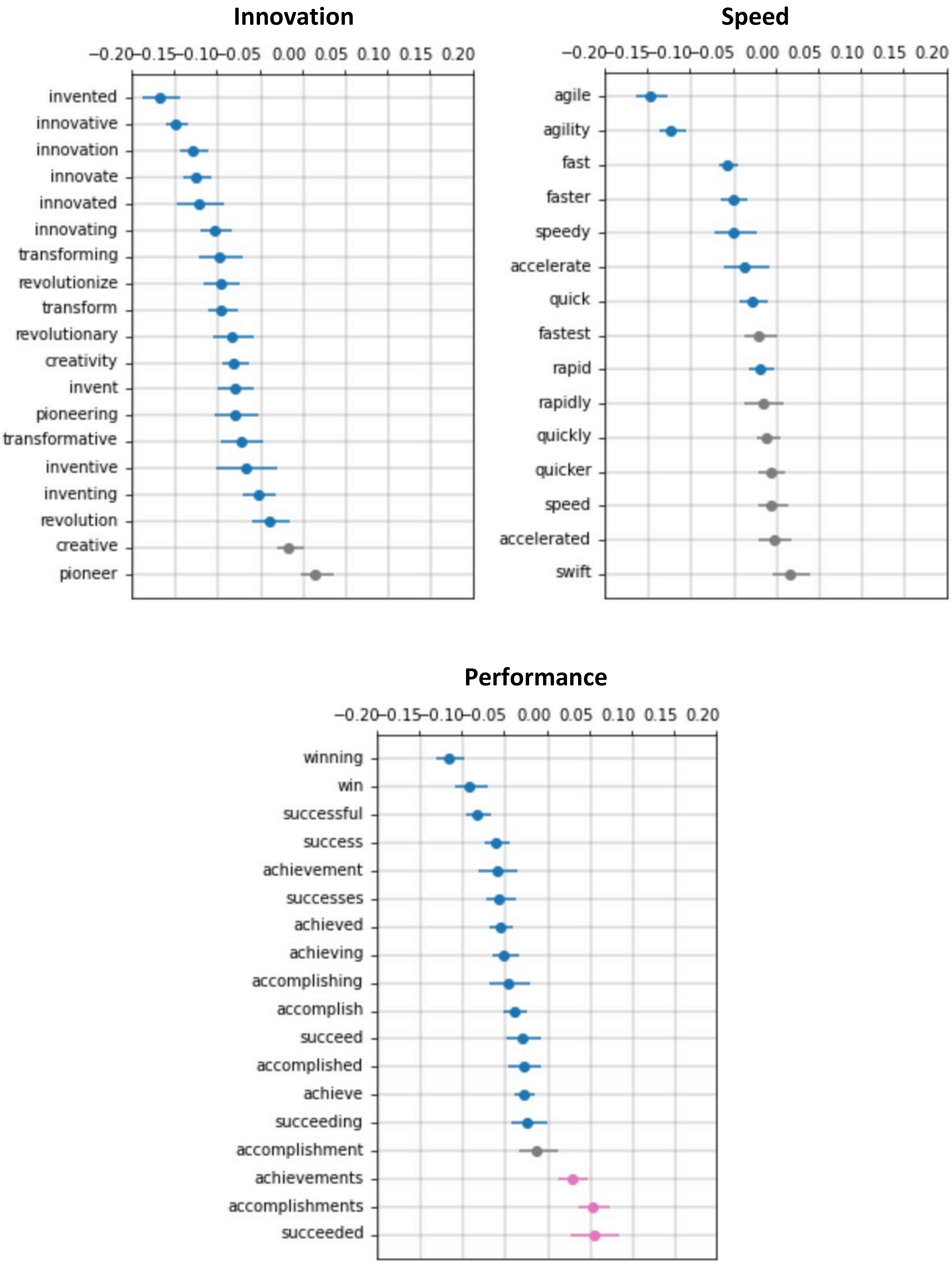
In these figures, the dots represent means across 20 subsamples, while the ends of the bars represent 95% confidence intervals. Blue indicates words that are male-typed, gray indicates words that are gender-neutral, and pink indicates words that are female-typed.

Figure 5b (cont'd):
Associations between the Gender Axis and Words Denoting Male Stereotypes



In these figures, the dots represent means across 20 subsamples, while the ends of the bars represent 95% confidence intervals. Blue indicates words that are male-typed, gray indicates words that are gender-neutral, and pink indicates words that are female-typed.

Figure 6: Associations Between Gender Axis and Words Denoting Corporate Values



Note: In these figures, the dots represent means across 20 subsamples, while the ends of the bars represent 95% confidence intervals. Blue indicates words that are male-typed, while pink indicates words that are female-typed.

Technical Appendix: Data Preparation and Analysis

Here we describe how we prepared the Glassdoor data for analysis and provide more details on the NLP techniques we used.

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. We 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 (corpus) originally contained 1,128,375 reviews of firms in the tech sector (nine industries defined above) from January 2008 to October 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 deeply 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. When we limit the analysis to reviews posted 2014 onward, we are left with 948,785 reviews. There are 210,422 unique words in this dataset; many of these are misspellings (e.g., thinkful, peopke, brincando) or acronyms and jargon (e.g., powershell, saas), but they are rare (the vast majority occur fewer than 10 times in the corpus), so they do not enter our word-embedding analysis. As we explain below, in order to derive robust word embeddings we eliminated from the training dataset for the word2vec algorithm any word that occurred fewer than 75 times.

We removed stop words from the remaining corpus only if they did not reduce the informational content, such as "the," "and," "or," and "if." We began with the union of stop-word lists from three popular Python libraries, nltk, scikit-learn, and spaCy. After removing words on this union list, we compared the filtered strings to the original strings in two ways. First, we conducted visual analysis, examining each filtered, original string pair to determine if the current stop-word list was removing essential meaning. For this, we made value judgments based on the important keywords and information persisting after filtering. In particular, we retained words denoting degrees of intensity (e.g., "more", "any," "very," and "always"), negations (e.g., "not," and "without"), and negative contractions (e.g., "don't" and "can't"). Second, we conducted data-driven analysis, generating a table to compare stop words with their frequency of occurrence in the entire dataset and the average shortening of each review when they were removed. These analyses led us to drop about 30 words from the union stop-word list. We then added some non-specific location words (e.g., "here" and "there") and pronoun contractions (e.g., "I'm" and "we're") to the stop-word list. The end result was a list containing 242 words; it is available from the first author upon request. After eliminating stop words, 210,422 unique words remain in this dataset.

Word-embedding analysis

We used word2vec, the most popular word-embedding algorithm (Mikolov, Sutskever, et al. 2013; Mikolov, Chen et al. 2013). 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. They are complementary: the CBOW architecture uses context words to predict a focal word; the skip-gram architecture uses a focal word to predict context words. Each architecture 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 skip-gram, 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 the context words are used as weights, so words that are closer to the focal word will have more influence than words that are farther away.

We custom-trained the word2vec algorithm on our corpus.[2] This algorithm begins by randomly assigning word vectors. After this first iteration, the model's predictions tend to be far from correct. After each iteration, the correct words are revealed and the algorithm updates the word vectors to reduce prediction error. Over many iterations, when predictions reach a predetermined level of accuracy, that yields the final set of word vectors to use in research.

As input for training models on our corpus, we separated out the three sections of Glassdoor reviews – pros, cons, and advice to management – in order to prevent contamination across the sections. The three sections of Glassdoor reviews are distinct documents, even though they were written by the same person at the same time. If we had concatenated these sections, 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 (we explain window sizes below), the window around the last word in the

[2] The most common pre-trained word2vec embeddings were trained on three million words in the Google news crawl. They 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 (after removing stop words). The corpus includes words that are less common in news stories, such as "environment," "hours," and "pay."

pros section would have been compared with the four words preceding it in that section 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.

We experimented with three different hyperparameters for the word2vec model: the number of dimensions for the semantic space (100, 200, 300), window size (6, 8, 10), and architecture (skip-gram vs. continuous bag of words [CBOW]). We selected hyperparameter combinations based on model performance (Levy, Goldberg, and Dagan 2015). Following previous research (e.g., Mikolov, Sutskever, et al. 2013), we used two performance standards to compare models: analogies and most similar terms. We began with the Google analogy test, which contains 8,869 semantic word pairs and 10,675 syntactical (word form) word pairs. This test assesses how well comparing two word embeddings (two word vectors) produced by a model reveals relationships between the words in a word pair. 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 this analogy test works, see Gladkova, Drozd, and Matsuova (2016).

The results of the Google analogies test comparison are shown in Table A1. To summarize, we found that (i) models with more dimensions were often (but not always) better; (ii) smaller windows were better, probably because the texts we study are relatively short; (iii) which architecture performed better depended on other parameters; and (iv) all models were close in terms of performance. We also experimented with excluding rare words[3] because associations between rare words and surrounding words tend to be very noisy, so excluding rare words improves the robustness of word-embedding models. 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 has 200 dimensions, a 6-word window, the skip-gram architecture, and a minimum word frequency of 75. It is bolded in Table A1.

[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 Mimno 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, 50 million tokens, and 210,000 unique tokens (16,000 after eliminating words that occurred in the corpus less than 75 times) – we assessed variability in model results (after settling on the hyperparameters, as described above). To generate robust estimates of word embeddings, we ran models repeatedly ("epochs" in NLP parlance). The gensim implementation of word2vec in Python allows researchers to train models iteratively. The first epoch is initialized with random values for parameters. The output for the first epoch is used as input to the second epoch, the output for the second epoch is used as input to the third, and so on. We continued this cycle for 50 epochs. We then compared word embeddings across each successive epoch for the words used to construct the gender cultural binary, which are central to our analysis. Concretely, we calculated cosine similarity scores for each gender term and each other word in the focal review; we then averaged cosine similarity scores across each review. There is no clear

[3] Many rare words were misspellings (e.g., "managament"), typographical errors (e.g., "stiw2" for "sit"), acronyms (e.g., "saas" for "software as a service"), or numbers. Others included numbers (e.g., "12th").

standard for judging when a custom-trained word2vec model is robust, so we followed previous research (e.g., Kulkarni, Al-Rfhou, Perozzi, and Skiena 2015; Hellrich and Hahn 2016a) and used 0.990 as a threshold. This model reached that threshold at 35 epochs.

We validated our model by reading reviews, as explained in the main text, comparing reviews with high and low gender scores (male-typed vs. female-typed). We also conducted most-similar queries; i.e., searches for words that are closest in semantic space to a focal word. We then manually assessed whether it makes sense for associated words to be in close proximity to words denoting gender. The results are shown in Table A2. The top half (Table A2a) demonstrates 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]

For the bootstrapped subsamples (20 subsamples, each covering 90% of the dataset), we followed the same epoch approach for each training effort. We judged that each of the 20 subsample models converged when the average cosine similarity score between the gender terms from the focal epoch and the subsequent epoch reached or exceeded 0.990. For all 20 subsamples, the model converged after 42 iterations; i.e., between the 42nd and 43rd iteration, the average cosine similarity score for the gender terms was above 0.990.

Word-embedding models calculate distances between terms in semantic spaces that typically have 100-300 dimensions; as explained above, our best-performing word2vec model has 200 dimensions. Given the high number of dimensions, Euclidean or straight-line distances are inappropriate for comparing semantic locations. Instead, analysts use cosine similarity to measure distances between word vectors; i.e., by calculating the cosine of the angle between two word vectors (Jurafsky and Martin 2023). This measure ranges from 1 to -1. A score of 1 means that words have the same meaning (the angle between their word vectors is 0°). A score of 0 means that they have orthogonal or independent meanings (the angle is 90°) because they are used in very different ways. A score of -1 means that they have opposite meanings (the angle is 180°). In the word2vec algorithm, word vectors are normalized to unit length before calculating similarity, so they all lie along the surface of a hypersphere – i.e., the generalization of a sphere into more than three dimensions (Levy, Goldberg, and Dagan 2015; Xing, Wang, Liu, and Lin 2015). The cosine of the angle between the vectors for two words then equals the dot (inner) product of their vectors.

When we calculated the gender axis, we used the average (centroid) of the set of gender vectors created by subtracting the vectors of male terms from the vectors of their female antonyms. Following other scholars (Bolukbasi et al. 2016), we tried an alternative method of aggregating data on multiple gender vectors – principal-component analysis (PCA). But the first component from a PCA of the set of gender vectors accounted for only 45% of their variance. This indicates that gender has multiple meanings, which is consistent with recent research on multiple components of gender stereotypes (Hentschel et al. 2019). Retaining only the first component would yield a weak signal of the concept of gender, so we rejected this alternative method.

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

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

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.

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

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

Xing, Chao, Dong Wang, Chao Liu, and Yiye Lin. 2015. Normalized word embedding and orthogonal transform for bilingual word translation. *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, Denver, CO, May 31–June 5: 1006-1011.

Table A1: Varying Hyperparameters: Results of Google Analogies Test

# Dimensions	Window Size	Min Word Freq	Accuracy: CBOW	Accuracy: Skip Gram
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 bolded.

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