Gender, Risk, and Leadership

by

Nur Yaldiz

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

University
of Stavanger

PhD programme in Social Sciences
UiS Business School
2022

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

ISBN:978-82-8439-076-5
ISSN:1890-1387
PhD: Thesis UiS No.639

Acknowledgments

First and foremost, I would like to thank my supervisors, Ola Kvaløy and Yuko Onozaka.

I sincerely thank my main supervisor, Ola Kvaløy, for his guidance, patience, and the freedom I was granted during my studies.

I am also immensely grateful to my supervisor Yuko Onozaka. I couldn't ask for a better academic mentor and a role model who has always been genuinely dedicated to our collaboration and always pushed me to do better. I wholeheartedly thank Yuko for her time, patience, and guidance.

I would like to thank our Ph.D. program coordinator, Mari Rege, for her support in my decisions. I also extend my gratitude to Elham Ghazimatin for her academic support and giving me multiple opportunities to practice my teaching skills. I am grateful to Niaz Bashiri Behmiri for being like a sister. I think no other word could define the support that Niaz has given me over the years. I am thankful to Simone Häckl for her academic insights and moral support. I wish to thank Rune Todnem By for sharing his insights on the leadership literature. I also thank my 50% seminar opponent, Kristoffer Wigestrand Eriksen, and 90% seminar opponent, Venke Furre Haaland, for their feedback on my research.

I am so lucky to be surrounded by amazing colleague-friends. I am thankful to Andreas for his understanding of my odd and nonexisting sleep schedule and for being the best friend of the road from the first day of our journey. I also thank Kristoffer, Kwadwo, and Utku, who let me be part of their circle. I thank Linda for her moral support and for giving me the best gift that one could ever ask for. Many thanks to Andreea for making time in her hectic schedule to read my long emails about my existential crisis. Special thanks to Max for his patience with my antics, and thanks to Erika for reaching out at much-needed times.

I am grateful to my dearest friends Anusha and Kutay for their heartfelt support throughout the years. I also wish to thank Mariana, Ines, Åse, and Monica for their kindness and understanding.

Many thanks to Prakhar for his wonderful presence.

Finally, my mother, Sabiha Doğanay, and my father, Asım Yaldız, made me the person I am today. Without resilience and love for reading, developed from a very early age, I would not be able to come this far. I cannot express enough gratitude.

Nur Yaldiz

Olavskleivå, Stavanger
January 2022

Dedication

To my dear Naz,
I know, sometimes I can be a bit much… but please hear me,
You are the only person who can genuinely make me laugh, whose wit and sense of humor I always find genuinely amusing.
Your presence makes me feel better, I know that Naz is always there…
As cliché as it sounds, I always want what is best for my sister
But now I realize that my sister is a fully adult who can make the best decisions for herself.
I dedicate this work to you, hoping that it may bring you luck and success in the new career path you have chosen for yourself.
Stay resilient.

Table of Contents

Table of Figures

List of Tables

Introduction

1 Introduction

Daily human interactions are likely to bear the weight of society's gendered expectations. Gender stereotypes (i.e., descriptive beliefs) manifested through such gendered expectations, garnered throughout history across cultures, define ideal male and female characteristics and dictate how exemplary men and women should behave. Gender roles are the sum of these stereotypes. Agentic (e.g., competitive, aggressive) characteristics are commonly considered typical male traits, whereas communal (e.g., warm, kind) characteristics are associated with the female. Without even consciously knowing, human minds learn how to view members of society based on these gendered traits, even though people usually have a mix of both agentic and communal characteristics regardless of their gender (e.g., Hyde, 2005; Larsen & Seidman, 1986).

What happens when women do not comply with these deep-rooted gendered beliefs? The three essays comprising this doctoral dissertation explore women's deviations from these shared expectations. Women are seen as defying societal expectations by acting agentic, such as taking risks and being in a top leadership position. Drawing upon behavioral economics, management, and applied psychology literature, this dissertation investigates women who do not fit into stereotypes.

The first essay investigates risk-taking women. Stereotypically, people tend to associate risk-taking behavior with men and risk aversion with women. When women defy these expectations, they may experience adverse results. This essay explores attitudes toward risk-taking women by asking whether people evaluate the outcomes of risky decisions differently for men and women. More specifically, the essay asks whether society holds risk-taking women responsible for unfavorable outcomes by not compensating for their losses. Moreover, the essay explores whether society considers either lucky men or lucky women to be more entitled to their gains.

The first essay employs a survey experiment to answer these questions. In the survey experiment, participants read and answered a set of questions within a given timeframe. In this setting, we introduced a hypothetical case describing a pair of either same-sex or mixed-sex workers who completed a similar individual effort task. Between subjects (i.e., each experimental subject only evaluated one pair of workers), we randomly varied the gender of the lucky and unlucky workers. In addition to their work compensation, these two workers either picked a safe sum or entered a lottery. Based on the lottery's outcome, one worker ended up worse off, while the other worker ended up better off.

This hypothetical case was a modified dictator game. In this setting, a third-party individual, namely a spectator, evaluated two hypothetical workers engaging in risk-taking decisions. We asked spectators to redistribute earnings between workers without revealing any information to the workers. When making this redistribution decision, spectators considered the workers' risk attitudes. Furthermore, we employed an implicit association test to explore whether stereotypical beliefs motivated spectators' decisions.

We find no evidence on whether men or women are considered to be more responsible for unfortunate outcomes. However, male spectators are less willing to redistribute earnings from lucky male workers than from lucky female workers. Furthermore, male spectators leave unlucky male workers with no earnings more often than they disenfranchise unlucky female workers. These findings may suggest alternative implications for the lack of female presence in nonstereotypical professions. Women may refrain from entering specific fields due to a lack of support for risky decisions with better outcomes. Alternatively, men might consider women as more likely to fail at nonstereotypical tasks. We find no indication of implicit biases in spectators' decisions.

The second essay focuses on the public image of female leaders. The role congruency theory (Eagly & Karau, 2002) shows that stereotypical leader characteristics are congruent with male gender norms. Thus, creating a public image becomes relatively easier for male leaders. However, descriptive norms associated with women do not match with stereotypical leader characteristics. Displaying an agentic image signals competency. However, female leaders are likely to face pushback (Rudman, Moss-Racusin, Phelan, & Nauts, 2012). The incongruity between gender norms and stereotypical leader characteristics may create a significant challenge for women (Connor & Fiske, 2018; Eagly & Karau, 2002; Rudman & Phelan, 2008). This study contributes to the literature by exploring the verbal impression management strategies of female leaders.

It is difficult, if not impossible, to interview a large number of high-end executives and leaders from different domains. However, publicly available data (e.g., interviews and conversations) provide an alternative. Essay 2 utilizes large unique linguistic samples, based on Adam Bryant's "Corner Office" column in the *New York Times* (*NYT*) that includes text samples from 522 US-based executives, as well as newspapers and lifestyle magazines with a global readership, popular podcasts, and talks held in various environments of these executives.

The linguistic methodology combines a dictionary approach, linear regression models, feature selection, and a supervised learning algorithm. The dictionary approach is a top-down process in which the software sorts the text data into predetermined categories. Linguistic Inquiry and Word Count (LIWC) software were used. From the LIWC categories, we utilized specific word groups to profile linguistic styles. The second essay employs these scores in the linear models. Feature selection involves counting the most frequently occurring words and word groups within the text to explore the content. Among supervised learning algorithms, this essay employs random forest classification, with the aim of extracting the most distinguishing features (i.e., words and word groups) of a model with satisfactory accuracy.

The analysis shows that female leaders use a less agentic and more communal style to shape their public image. Among female leaders, consumer discretionary executives employ a more agentic linguistic style than their counterparts. The consumer discretionary sector has a relatively higher number of female executives who are not chief executive officers (CEOs) (Desilver, 2018). The literature supports that nonstereotypical women evoke less judgment and criticism in female-dominated settings (e.g., Dasgupta & Asgari, 2004). Content-related findings may indicate that female leaders project a maternal image to ease backlash. A "strong mother" image may not evoke bias if it is a female leader's sole image.

We also compared these results to those of the general public to obtain a baseline. Female leaders appear to have a stronger masculine linguistic style than women in the general population. When we extend this analysis to male leaders, we find that male leaders use a more communal linguistic style than general public men. An implication could be that male and female leaders tend to draw more of an androgynous image, combining agentic and communal rather than following stereotypical gendered expectations.

The third essay builds on the second essay by employing a portion of the same textual data. The main aim of this essay is to explore the personality expressions of business leaders through publicly available text data. It is not possible to observe the genuine personality traits of senior executives. However, methodological advancements that come with text analysis and machine learning algorithms have made it possible to have a general picture of these personality traits (Pennebaker, 2011). Although public speech samples do not provide genuine personality traits (Mehl et al., 2006), linguistic samples need to be close to one's actual self to present an authentic image (Pennebaker, 2011). Hence, personality expression may not be too apart from genuine personality traits.

Findings show that female leaders are more extraverted and agreeable but less conscientious, open, and emotionally stable than male leaders. Extraversion, agreeableness, and lack of emotional stability are more associated with women. In contrast, previous survey studies show that male and female leaders do not dramatically differ in their personality traits. These results may indicate that the linguistic personality expression of female leaders matches gender prescriptions.

The structure of this chapter is as follows. Section 2 discusses the focus of the overall dissertation by seeking an answer to the following question: Why does gender matter? Section 3 explores women's willingness to compete and take risks and includes a discussion about nature versus nurture. Section 4 introduces the main characteristics of dictator games and background on the survey experiments by reviewing studies that have a hypothetical design. Section 5 provides a broad background on human language, including neural, cognitive, and social qualities. Moreover, this section discusses the implications of employing diverse text data (e.g., written and spoken samples, and public and private speech) in linguistic analyses. The conceptual framework of this dissertation is presented in Section 6, wherein Section 7 provides general information about the overall methodology. Section 8 discusses the motivation behind using a US sample in this dissertation. Detailed abstracts for each essay are provided in Section 9. Section 10 provides an overall discussion.

2 Why does gender matter?

Technological developments have triggered shifts in societal dynamics. Each industrial revolution brought about changes in men's and women's roles. The First Industrial Revolution (1760–1840) created a stronger middle class in which men and women followed the stereotypical labor division. However, there was also a shift in women's place. Specifically, women who migrated to cities from rural areas started working in factories. As female labor is dramatically cheaper than male labor, the former primarily supplied the growing textile industry in New England, the United States (Dublin, 1979).

The Second Industrial Revolution (1871–1914) was an era of technological revolution when electricity, as well as railroad and telegraph networks, began transporting people and ideas faster, machines started replacing humans, and mass production brought cheaper products to the market. Growth also resulted in a surge in unemployment. Similar to in the previous era, women from lower income groups maintained factory jobs (Mokyr, 1992). This labor participation also triggered issues related to women's voting rights. It was crucial to have (female) representatives who could understand the struggles of working-class women (Mead, 2006).

After the two world wars, the Digital Revolution marked the Third Industrial Revolution (1950– 2013). During this time, stagnation in industrial and technological development expanded significantly with the emergence of computerized technologies (Taalbi, 2019). This era also witnessed compelling progress for women, starting in the late 1960s.

The spread of the birth control pill unequivocally changed women's lives (Goldin, 2006). Changes in state laws allowed young and single women to access contraceptives. The power to delay pregnancy gave women more time to seek an education. From the beginning of the 1970s, women began to earn tertiary degrees in male-dominated fields such as medicine, law, and business administration (Goldin & Katz, 2000).[1] Corporate leadership positions also began to see a female presence. In 1972, the Fortune 500 list had its first female CEO of *Washington Post*, Katherine Graham (Carpenter, 2017).

Women's desire to reach the top levels produced the glass ceiling metaphor in the mid-1980s. The term first appeared in *Working Woman* magazine in 1984 to describe women's inability to rise above middle management. A year later, the National Organization for Women used the metaphor to refer to the same phenomenon (Boyd, 2008). Journalists Carol Hymowitz and Timothy Schellhardt used the glass ceiling[2] to describe women's ascent to a certain level in a *Wall Street Journal* article on March 24, 1986 (Hymowitz & Timothy, 1986). Although it might seem as though women can reach executive levels, having the relevant skills and background may not lead to senior executive positions.

Invisible artificial barriers prevent women from reaching upper positions; these barriers arise from hierarchical male-dominated cultures in corporate America. *Glass* indicates that women may approach opportunities to advance, but *ceiling* implies that there is a concrete maximum limit to women's achievement. Although there are visible opportunities on the other side of the barrier, it is impossible for women to obtain them (Boyd, 2008).

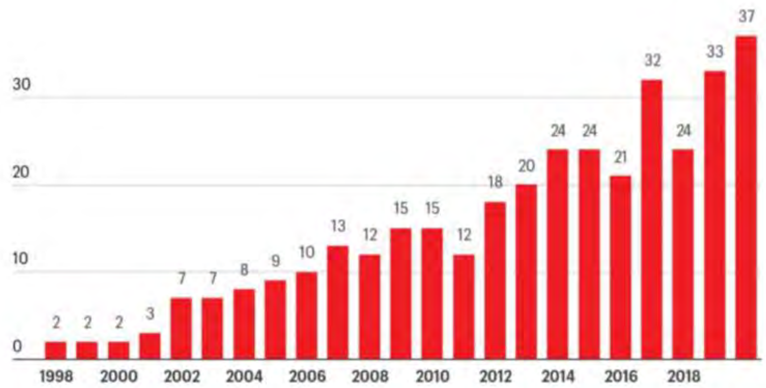
The 1991 Glass Ceiling Initiative report from the US Department of Labor confirmed that the glass ceiling had caused highly-skilled and qualified labor loss due to existing barriers erected against women and minorities. The numbers also revealed that women made up 46% of the

[1] The pill is the main catalyst behind the change in women's place. The late 1960s and the beginning of the 1970s involved a shift in social dynamics, triggered by various events. For example, the Civil Rights movements triggered the reemergence of feminism, which influenced affirmative action policies.

[2] The metaphor started with women and soon became a blanket term to describe the impediments that ethnic and racial minorities face. However, due to the scope of this dissertation, the researcher primarily discusses women's challenges overall.

workforce and earned more than half of the master's degrees at that time. Nevertheless, men occupied 95% of the senior positions and had a relatively higher income than their female counterparts. President George H. W. Bush approved the Glass Ceiling Act in the same year to prevent discrimination against minorities in the corporate world (Boyd, 2008).

These improvements brought some promising and dramatic albeit small-scale changes in CEO positions for women. In 1999, a Fortune 500 tech company, Hewlett-Packard, had its first female CEO. In 2009, the first female CEO succession occurred in the Fortune 500. This transition also produced the first female African American CEO on the list. Furthermore, since 2013, the automotive industry has had its first female CEO (Carpenter, 2017). Compared to a couple of female CEOs in the late 1990s, 7.4% of Fortune 500 CEOs are now women (Hinchliffe, 2020).



Source: https://fortune.com/2020/05/18/women-ceos-fortune-500-2020/

Figure 1 Fortune 500 female leaders from 1998 to 2020

Building on these improvements from the digital era, the Fourth Industrial Revolution has seen more significant female leadership in politics. However, this shift may not imply that people have become more tolerant of nonstereotypical persons. Women are still likely to struggle to achieve leadership positions in various domains. While their contributions might still be subject to bias and discrimination in specific fields, female and minority contributions in this era appear to be highly crucial.

Following the technological developments of the Digital Revolution, the Fourth Industrial Revolution mainly builds on artificial intelligence [AI] advancements and the Internet of Things (IoT). These technological developments shape products and services that have become an integral part of human life. Social media and many Internet applications have transformed people's lives and disrupted societal dynamics (Schwab, 2016).

Although the changes that have come about amidst the Digital Revolution are somewhat promising, the Fourth Industrial Revolution would benefit from a more diverse labor pool. Globally, 22% of AI (World Economic Forum, 2018) and 30% of science, technology, engineering, and mathematics [STEM] researchers are women (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2019). This ratio has remained in the same range in recent years and is unlikely to become a positive trend in the near future (World Economic Forum, 2018). The numbers in the United States do not differ dramatically from global indicators: 35% of STEM students (Madgavkar et al., 2019) and 22% of AI professionals are women (World Economic Forum, 2018)

When there is only one type of person, there can be no diversity of opinions. Diverse groups approach technological developments from different perspectives. Furthermore, social psychology studies show that compared to non-diverse teams, diverse teams produce more creative solutions to problems. This outcome is associated with paying more attention to non-similar group members (e.g., ethnicity, opinion, gender). The human mind assumes that non-similar people are likely to have different types of information (Phillips, 2014).

Non-diverse groups only approach existing problems from their perspective. If someone is not experiencing a particular problem, they are not likely to generate a solution to it. This situation occurs in AI applications, such as stereotyping and profiling. For example, big tech companies' (e.g., Microsoft, IBM) facial recognition algorithms perform better on male faces with lighter skin tones (Simonite, 2018). These AI systems fail to recognize female faces, especially those with darker skin tones (Buolamwini, 2019).

The flaws in AI applications appear to reflect developers' stereotypical thinking. In addition to facial recognition, language has emerged as another problem. In gender-neutral languages, Google's translation application assigns a gender to word groups based on stereotypes. For example, the software assumes that cleverness is associated with men and that beauty describes female subjects (Ullmann & Saunders, 2021). Similarly, "doctor" takes a male subject, while "nurse" takes a female subject. Voice recognition algorithms assume that a deep male voice belongs to a large, strong man (Cox, 2018), whereas most virtual assistants and companions have a female voice, complying with nurturing female stereotypes (Fung, 2019).

The influence of stereotypes goes as far as loan applications, job recommendations, and visual recognition. Research suggests that AI algorithms used in the finance industry tend to reject loan applications from single women. These systems' developers are likely to classify single women as having a limited income without additional support to strengthen their commitment to loan installments. Furthermore, Google and LinkedIn tend to display high-paying jobs less to female users. Similarly, Facebook's and Microsoft's image recognition system software mostly matches stereotypically female tasks (e.g., cooking and shopping) with women (Büchel, 2018). These examples illustrate the importance of diversity in the field of technology. Otherwise, the singlemindedness behind these newly developed systems may reinforce stereotypes and social biases.

These types of problems are not relegated to the AI and STEM fields. Academia appears to be in urgent need of diversity, especially in stereotypically male-dominated fields. Economics is no exception to the lack of diversity. In 2014, women received 30% of doctorates, and 8% were awarded to minorities in the United States. Faculty numbers are also not very promising. Women occupy 23.5% of the tenured and tenure-track positions, 31% of the assistant professor positions, and 15% of the full professor positions (Bayer & Rouse, 2016).

In addition to the gender gap in doctorates and faculty positions, studies show that the academic environment may not see women as being as competent as men (Moss-Racusin et al., 2012; Wu, 2018) and also may not consider women’s contributions to be equally valuable (Sarsons, 2017a). A recent study found that when giving presentations, female economists were 12% more likely to be interrupted and receive hostile questions (Dupas et al., 2021). Although 12% does not seem like a huge gap, this finding illustrates attitudes arising from different views. Indeed, the American Economic Association has focused on stimulating more egalitarian practices in recent years (e.g., Casselman & Tankersley, 2019; Rosalsky, 2020).

Women appear to struggle in the political leadership arena as well. In 2007, the US presidential race had its first female major-party nominee, Hillary Clinton. Although people seemed to accept a female candidate in the presidential race, they did not believe that a woman could win (Eagly, 2007). In 2006, a *NYT* opinion column frankly suggested that gender could be a significant drawback for Clinton and may prevent her from winning (Herbert, 2006).

Heeding these warnings, Clinton maintained a masculine linguistic style at the beginning of the first campaign. This image did not bode well in the context of societal expectations. Clinton appeared to be an angry, unlikeable candidate. Her efforts to maintain a public image that could mitigate the gender handicap seemed to backfire. Thereafter, it was replaced by a more feminine image. This time, however, the strategy failed to create a genuine impression. The campaign ended in crisis, with fluctuation between masculine and feminine linguistic styles (Jones, 2016).

After serving as secretary of state during the Obama administration, Clinton reran for the presidency in 2016. At that time, Clinton appeared to be the most qualified candidate, given her immense experience in Washington and her status as a well-known public figure. Despite all these accolades and efforts, Clinton could not manage to distance herself from being a female candidate. During her first presidential race, gender was the core of the Clinton campaign; however, highlighting gender likely harmed Clinton's image (Chozick, 2017).

As a comparison, Jenet Yellen managed her image by distancing herself from gender issues and de-emphasizing her status as the first female federal reserve chief in the stereotypically male-dominated finance world. Chozick (2017) suggests that Yellen's subtle move with respect to gender is a reasonable strategy for navigating a traditionally non-diverse environment.

In addition, before the social media ban, Donald Trump's Twitter posts were the subject of academic studies (e.g., Ahmadian et al., 2017; Jordan & Pennebaker, 2017) and widespread media attention. Findings show that Donald Trump has a solid feminine linguistic style (as determined by Jones for POLITICO) (Sedivy, 2016). A *NYT* article makes similar arguments based on research from a big data analytics company (Miller, 2016). These findings indicate that having an emotional and social tone with self-focus has become a selling point for Trump. Unlike Hillary

Clinton, who initially employed a masculine style, Trump's feminine style does not seem to stir backlash. Rather, it appears to portray a warm, likable image.

Male professionals may not endure this type of struggle when setting a public image. People still refer to a stereotypical male image when researchers ask them to describe a leader (Murphy, 2018). From their language to their appearance, female leaders might be subject to different societal judgments. It might not only be about whether women and men have comparable qualifications. Rather, it might also be about how female leaders reconcile with the norms of being a woman. Being soft-spoken and good-looking are seemingly trivial issues that society and professional circles may find crucial in the image of a female leader (Baird, 2016; Glass & Cook, 2016).

Furthermore, women's professional identity may be smaller in nonstereotypical contexts. In professional sports, women are likely to receive more gender-biased questions than their male counterparts, with 70% of the questions being unrelated to the profession (Mullainathan, 2017). These findings come from a machine learning algorithm trained on journalists' questions. Analyses show that male tennis players receive more professional and game-related questions in press conferences (Fu et al., 2016).

These details and struggles to maintain a well-received image as a nonstereotypical woman can discourage qualified candidates in different domains. Indeed, women may choose not to be part of a competitive environment, despite having the necessary skills and capabilities (e.g., Niederle & Vesterlund, 2007). In return, incompetent men may occupy crucially important positions and are likely to receive ample support (Chamorro-Premuzic & Gallop, 2020).

However, Hillary Clinton's efforts may not all have been in vain, as she has likely inspired other women (e.g., Beaman et al., 2009). In 2018, women won more seats than ever in the US Congress (Zhou, 2018) to a total of 112 women in congress. This was followed by the swearing in of the first female vice president, Kamala Harris, in US history (Lerer & Ember, 2020). Although these examples are from the political domain, having a female example, in general, can positively influence other women who have suitable backgrounds to pursue top positions (Ely, 1994). Whether in a local government context (e.g., Arvate et al., 2018) or a corporate setting (e.g., Dezsö & Ross, 2012), female senior leaders can motivate junior and middle managers to compete for top positions.

Women's presence in different settings are likely to reduce the impact of stereotypes. Female leaders in local government can positively influence adolescents' educational attainment and ease the burden of gender roles (e.g., household chores) in developing countries (Beaman et al., 2012). Furthermore, female mentors and role models can increase women's attainment in nonstereotypical academic fields during university education. At engineering schools, female mentors strengthen female undergraduate students' belonging, motivation, and aspirations. The first year appears to be especially crucial, since women tend to question their capabilities and decisions at this stage (Dennehy & Dasgupta, 2017). Similarly, Porter and Serra (2020) suggest that female role models can influence female undergraduate students majoring in economics.

Studies suggest that women's benevolence also promotes influential female leaders (e.g., Bass, 1999). Stereotypically communal characteristics (e.g., empathy, kindness, and humility) tend to fare better than competitiveness, assertiveness, and aggressiveness. Female leaders generally prioritize groups rather than individuals (Eagly et al., 2003). This approach is likely to be critical in leadership when establishing ties with stakeholders (Lemoine & Blum, 2021). Female leaders adopt an employee-centered approach and support fewer layoffs (Matsa & Miller, 2013). This style strengthens team identity by prioritizing a group rather than an individual (Eagly et al., 2003).

Although it is unclear whether a female leader's benevolent style is an outcome of stereotypical expectations (e.g., Vinkenburg et al., 2011), Adams and Funk (2012) suggest that benevolence is part of personality. In other words, benevolence creates the main difference between male and female leaders who otherwise share comparable characteristics. Thus, having both feminine and masculine features is also crucial. Various studies argue that followers may prefer a more androgynous (i.e., balanced feminine and masculine features) image over a masculine one (e.g., Kark et al., 2012; Kent & Moss, 1994).

Indeed, it is likely that female leaders may integrate masculine and feminine characteristics into their leadership style. A recent study has found that female-led US states have performed better against the novel coronavirus (Covid-19) pandemic and have had fewer deaths (Sergent & Stajkovic, 2020). These findings indicate that female leaders mix assertiveness with benevolence to implement a strict and protective plan. Furthermore, female leaders may have strong resilience in the face of adversity.

Such resilience might also explain why female leaders' strengths could paradoxically contribute to obstacles. The corporate world considers women to be a safeguard in times of crisis. The glass cliff metaphor illustrates that female leaders are likely to obtain positions that are prone to failure. Ryan and Haslam (2007) describe this situation as a drawback in female leaders' careers. Positioning women as saviors might seem to be a strengthening characteristic that is likely to open leadership positions (e.g., Stevenson & Orr, 2017). However, female leaders' strong people skills may not resolve gender discrimination in the long run and may even strengthen it. In other words, corporations are likely to select female candidates to absorb the fall and blame (e.g., Babcock et al., 2017; Ryan et al., 2011).

Partially in line with the glass cliff arguments, a recent study analyzing public opinion polls from 1946 to 2018 shows that stereotypical beliefs about leaders with masculine features appear to be changing gradually in women's favor, depending on the setting (Eagly et al., 2020). For example, in the military, women are less favored as leaders. However, in nonprofit organizations, educational settings, and social services, subordinates consider female leaders to be slightly better than male leaders (Eagly, 2007). This not-so-spectacular shift may also explain why female leaders in senior positions are likely to face dislike and disapproval (e.g., Koenig et al., 2011). Regarding leadership ranks, women are more associated with junior and middle managerial positions (e.g., Eagly, 2007; Eagly et al., 2020). Consequently, the proportion of female middle managers remains relatively high among Standard and Poor's (S&P) 500 companies (36.9%), while the number falls drastically for female-held executive positions (5.8%) (Catalyst, 2020).

Although women's stronger focus on building a career and achieving leadership positions has triggered changes in the numbers, this shift is a prolonged process. Highly-educated women are likely to bear a significant share of domestic chores and child-rearing duties (Bertrand et al., 2010). Even in the most egalitarian countries, women appear to contribute more to unpaid labor (Kleven et al., 2019). Consequently, women who wish to start a family and maintain their careers may eventually need to decide between one or the other (Goldin, 2014). These concerns might start from graduate school for highly-skilled women (Bursztyn et al., 2017). Among equally-skilled career-oriented opposite-sex couples, women are more likely to withdraw from the labor force after starting a family (Cha, 2010).

Bertrand (2013) suggests that college-educated women with a career and a family may be more dissatisfied than women who only raise their families. Work-induced stress and exhaustion may contribute to this unhappiness. Ryan et al. (2007) explain that detrimental effects on well-being might arise from taking over positions amid a crisis. This pressure from the work environment is more likely to lead to a lack of organizational belonging for women. In a way, these glass cliff positions may explain women's higher turnover. In general, supporting women's overall labor force participation and providing grounded opportunities for advancement are likely to ease the burden on women's psychological well-being. Anxiety might start when women divide their time between family and career (Bertrand, 2011). This burden may be amplified when navigating leadership and senior executive positions (Eagly & Carli, 2007).

Although there are many advantages of having higher female participation in specific fields and having women take on leadership roles, the actual effects of executive women's presence are likely to become apparent in the long run (Adams, 2016). Studies exploring Norwegian law on gender quotas (2003) have uncovered disheartening evidence due to young, inexperienced female board members (Ahern & Dittmar, 2012) or no benefit for women's labor market outcomes (Bertrand et al., 2019). There are mixed findings regarding financial performance and female senior executives in boardrooms (Eagly, 2007). Furthermore, innovation outcomes (Apesteguia et al., 2012) and employee diversity policies (Cook & Glass, 2016) can be more closely linked to board gender diversity than the sole presence of female directors. Instead of being influential role models and paving the way for other women, females in top leadership positions might assimilate to the dominant male culture and avoid associating with junior and midlevel employees (Derks et al., 2016). Correspondingly, women may not consider female leaders to be as competent as their male counterparts (e.g., Goldberg, 1968) and may overestimate men's capabilities (e.g., Bagues & Esteve-Volart, 2010). Finally, organizations may use diversity as a reputation enhancing tool and only hire women from certain ethnicities. It would be misleading to assume that challenges and discrimination may disappear as the number of women increases. In other words, mere numbers may not always give an accurate account of the process (e.g., Wilton et al., 2019).

This section covered a wide range of information regarding the importance of gender diversity. Technological improvements that emerged in the Third and Fourth Industrial Revolution have catalyzed societal change. The research discussed in this section shows that it is crucial to have people from diverse backgrounds in STEM, AI, and academia, as diminishing singlemindedness in these critical fields is likely to produce more effective outcomes for society as a whole. Furthermore, political and corporate leadership positions are likely to benefit from a diverse leader pool. Female leaders set a motivational example for their followers and prove to be influential leaders through their leadership style.

Finally, this section has provided insights into nonstereotypical women's struggles. These struggles primarily arise from third parties (e.g., the glass ceiling and the glass cliff) and are also defined as demand-side obstacles (e.g., Barbulescu & Bidwell, 2013). Section 3 below discusses the so-called obstacles that arise from the supply side, namely women's tendency to avoid risk and competition for leadership positions. This is accompanied by a nature versus nurture discussion.

3 Nature versus nurture

The nature argument arises from whether women avoid taking risks and entering competitive environments due to biological and evolutionary impetuses. In contrast, the nurture side suggests that women learn to refrain from these particular tendencies. Fields grounded in evolution and biology may support the argument that women's inclination is an outcome of their nature.

Evolutionary psychology argues that men and women have cognitive and behavioral differences due to their biological makeup (Buss, 2015). Women's limited resources to have and raise offspring have contributed to female evolution. Arising from the obligation to be around for their young offspring, women primarily gathered food and settled in a particular place. Due to these constraints, women might have developed instincts to be cautious and may also have evolved as caretakers and nurturers with a risk-averse attitude.

In contrast, men can have multiple offspring with different partners simultaneously. Thus, men may not need to be as cautious as women to maintain their chances of reproducing. Men may even need to take risks in order to increase their possibilities. In addition to continuously seeking opportunities to procreate, men were also hunters. While searching for a big hunt, men likely did not set up the conditions for a settled life centered around their offspring. Due to these tendencies, men have evolved as risk-seekers and competitors.

The evolutionary differences between men and women explain their nature. According to this view, evolutionary heritage seemingly impacts the sexes' decisions. A general assumption is that women have displayed specific tendencies throughout human history. These tendencies have built stereotypes that became an exaggerated version of reality. For example, women's maternal instincts to look after their offspring might be the origin of unpaid labor at home, and a lower income in professional life.

In general, the female biological setup may not work in favor of women during the most crucial years of their careers. Improvements in medicine are most likely to eliminate some of these struggles (e.g., Bertrand, 2011). As discussed in the previous section, women's significant participation in higher education was observed after the pill became available to single women. This development further increased female presence in nonstereotypical fields and likely triggered women's pursuit of leadership positions. In sum, exercising reproductive control has given women a much needed edge in terms of taking risks and competing. To some extent, biological differences shape both men's and women's decisions.

Instead of focusing on biological differences, nurture approaches gender differences as an outcome of social beliefs (e.g., Eagly & Carli, 2007). In this context, gender is a social identity defined by beliefs about men and women. People learn how to fulfill these expectations from their immediate environment. These traits are generally dynamic, and they change with age. In other words, the circumstances surrounding people's lives from the moment of birth contribute to the formation of these traits. Being a risk seeker or being benevolent may be more pronounced in a specific context than traits arising from evolutionary differences.

More specifically, women can adapt to the challenges presented in different contexts (Akerlof & Kranton, 2000). In professional settings, women's attitudes toward risk are comparable to those of men. However, to comply with societal expectations, women appear to be risk-averse outside their work domain (Drupp et al., 2020). Indeed, highly-skilled and competitive women may struggle to take risks and enter competition when their family identity is intact (Cadsby et al., 2013). These findings indicate that social beliefs within the context may substantially shape highly-skilled women's decisions.

Moreover, women may avoid specific occupational domains. As highly-skilled women, Master of Business Administration (MBA) students are less likely to apply for competitive jobs in the financial sector, as they prefer general management positions. Based on their pre-MBA experience, women are aware that they are most likely to work much harder to legitimize themselves in nonstereotypical contexts (Barbulescu & Bidwell, 2013).

Consequently, women may not consider investing considerable effort in an environment where they are outsiders. Women's lack of a sense of belonging can strengthen this avoidance. In top echelons, women may not consider applying to similar senior executive positions after a rejection (Brands & Fernandez-Mateo, 2017). Furthermore, the lack of female representation in senior leadership positions can discourage women (Niederle et al., 2013). Indeed, there is a limited number of available female candidates in the executive job market (Fernandez-Mateo & Fernandez, 2016). In addition, highly-skilled women are likely to consider fields that may allow them to have a work–life balance (Barbulescu & Bidwell, 2013).

Although search firms contribute somewhat to the candidate gap (Fernandez-Mateo & Fernandez, 2016), gender segregation in job offers may not occur toward members of a highly-skilled labor pool. Barbulescu and Bidwell (2013) find no evidence that women are less likely to receive an offer after applying to a job in a nonstereotypical domain.

These findings may indicate that women may not see themselves as part of the stereotypically masculine domains. This counter-motivation may arise from women considering themselves to be less competent than men. Alternatively, women may not want to experience the struggles of joining a nonstereotypical domain. Witnessing backlash against female managers may prevent women from competing for higher-ranking positions (Chakraborty & Serra, 2018). Similarly, among highly-skilled women, married candidates' decisions tend to resemble their male counterparts'. Overcoming multiple gender biases at the early career stage may strengthen competitiveness at subsequent stages (Barbulescu & Bidwell, 2013).

The literature shows that women tend to avoid competition, risk, and specific positions. However, this unwillingness likely arises from learned traits and societal beliefs rather than from biological differences between men and women.

4 Background on dictator games and survey experiments

In economics, the ultimatum game and its derivative dictator game seek to answer the question of whether it is selfish or altruistic motives that drive human behavior. In a simple ultimatum game, a proposer decides how to divide a specific sum between themselves and a recipient. The recipient can accept or reject the proposer's allocation decision. In the case of rejection, neither party gains any earnings. Otherwise, both parties receive the allocated amount, and this concludes the game. A subgame perfect Nash equilibrium occurs when the proposer makes a minimum amount offer, which the recipient accepts. Here, the assumption is that the recipient's best option is to take the minimum amount in order to avoid walking away with nothing. Experimental studies have shown that this may not be the case most of the time. Recipients are inclined to reject minimum offers equal to or lower than 20% (Camerer & Thaler, 1995). Proposers mostly share 40% of their total sum with their recipients (Camerer, 2003).

The design of ultimatum games can influence both parties' decisions. In this setting, proposers may wish to be fair by sharing a larger amount than they prefer. In contrast, responders may reject offers because they consider them to be unfair or insufficient. The literature explains that proposers share more either due to inequality aversion or fear of rejection. The former motivation aims to create an even distribution between parties. The latter can arise from selfishness and fear of ending up with nothing (Güth & Kocher, 2014).

Dictator games eliminate this vagueness by changing the proposer's role to that of a dictator who is empowered to make decisions. Since recipients cannot reject offers, dictators' fair allocation decisions may imply altruistic motives (e.g., Fehr & Schmidt, 1999). Substantial literature shows that there are three main types of dictator decisions.

Firstly, dictators may act selfishly and opt not to share (this inclination may be more prevalent under anonymity; Charness & Gneezy, 2008; Franzen & Pointner, 2012). Second, dictators may choose an equal distribution. The first known implementation of the dictator game in a classroom setting shows that more than half of the students split the amount evenly between themselves and the recipient (Kahneman et al., 1986). Third, dictators may share modestly (Camerer, 2003; Oxoby & Spraggon, 2008).

Broadly, researchers employ dictator games to investigate the role of gender in decision making. Generally, women are more inclined toward equality, while men prefer to make more selfish, and sometimes also more efficient allocations (e.g., Croson & Gneezy, 2009). In a double-blind dictator game, female dictators are likely to act more generously than male dictators (Eckel & Grossman, 1998). Bolton and Katok (1995) found no difference between men and women in a less anonymous setting. In a modified dictator game that focuses solely on dictators' giving behavior, men are likely to behave selfishly and share less when giving is expensive. This inclination reverses for men when sharing is cheaper. Unlike men, women are likely to choose an equal distribution in either case (Andreoni & Vesterlund, 2001). In a gender-paired setting, women share less with other females than they do with men and anonymous people (Ben-Ner et al., 2004). A meta-analysis suggests a slight difference between men and women regarding dictator generosity (Engel, 2011 and Bilén et al., 2020 also indicate comparable findings).

Furthermore, the current literature implements modified versions of dictator games to investigate the influence of stereotypes on people's decisions. Aguiar et al. (2009) sought to answer the question of whether people consider women to be more altruistic than men using a design involving active decision making from recipients only.

As an alternative to focusing on recipients, it is also possible to focus solely on dictators. In the spectator design, dictators represent impersonal third-party agents called spectators. In simple terms, these people determine a pair of workers' final earnings. Cappelen et al. (2013) utilized the spectator design with stakeholders (i.e., recipients who share their stakes with other recipients) in a risk environment. In another study, Cappelen et al. (2019) focus solely on spectator behavior when the cause of inequality between male and female workers is merit.

Regarding design, the first essay partially builds on Cappelen et al. (2019), that is, utilizing a hypothetical case and spectators without actual workers and real stakes. As impartial third-party agents, spectators only evaluated hypothetical workers' decisions. In addition to their income, workers could enter a lottery for a chance at additional earnings, or they could collect a smaller amount. The experiment consisted of the following simple steps. Firstly, spectators read hypothetical cases. Then, given the information about risk attitudes (i.e., entering the lottery or accepting a certain amount) and gender, spectators decided on hypothetical workers' final earnings.

Survey experiments do not serve as a standard instrument in economics as in psychology and sociology. Studies in the latter fields show that survey experiments produce reliable results (e.g., De Dreu et al., 2001). A sociological survey experiment concerning labor market outcomes for men and women found that both groups thought that women should earn lower wages than men with similar qualifications. Such beliefs among both women and men confirm stereotypes by viewing male workers as higher-performing individuals in a work environment (Auspurg et al., 2017). This finding also partially overlaps with economics studies that employ laboratory experiments (e.g., Reuben et al., 2014) and field studies (e.g., Bohren et al., 2019).

Furthermore, economics research employing hypothetical scenarios also suggests that survey experiments can produce reliable results (e.g., Baert & De Pauw, 2014; Finseraas et al., 2016; Stephan et al., 2014). Economists may question the generalizability and external validity of the results in an abstract context. Researchers supporting the use of survey experiments suggest that participants can relate to the presented situation without the distractions of a restricted artificial laboratory environment (Baert & De Pauw, 2014; Colquitt, 2008; Mook, 1983).

Though they are a nonmainstream tool, there are examples in the economics literature that employ hypothetical instruments. Daruvala (2007) utilizes hypothetical questions to explore whether people consider women to be more risk-averse than men. In this experiment, although there is no real impact on others' earnings, the experimental agent's decision impacts their earnings. The participants were 71 male and 61 female undergraduate students. The findings suggest that individual risk preferences and stereotypes shape predictions about others. This tendency reflects stereotypes, while failing to recognize the actual differences between men and women. Furthermore, risk-averse individuals mostly consider men to be more risk-neutral.

Another experiment asked participants to predict hypothetical male and female students' risk-taking decisions through hypothetical gambles. The participants were 30 male and 61 female students ranging from 18 to 28 years old. These findings partially overlap with those of a previous study. Men and women tend to overestimate men's risk preferences, while accurately predicting female risk-taking behavior. In this study, women relied on stereotypes in their predictions, explaining why women overestimate male risk-taking behavior more than men (Siegrist et al., 2002).

Ben-Ner et al. (2008) find that hypothetical and real stakes do not create differences in dictator allocation decisions. However, experimental agents' personalities might be the main deviation point, rather than experimental design. In another study, Ben-Ner et al. (2009) employed two different experiments to investigate the impact of in- and out-group identity while interacting with others in various settings such as work, commuting, and a dictator game. Although the first study has a hypothetical design and the second study is a dictator game with actual stakes and experimental agents, the two studies reveal comparable findings. The hypothetical case likely does not impact experimental agents' behavior toward others.

Being aware of a real experimental subject may impact decisions. Bohnet and Frey (1999) suggest that agents act more generously when they see other experimental participants. In contrast, subjects can be more selfish when researchers do not directly expose experimental agents to each other (Solnick, 2001). Women might settle for less when they sit face-to-face with their proposers (Eckel & Grossman, 2001). In other words, seeing the other experimental agent in person may influence the likelihood of an equal distribution. Participants may act less selfishly and follow stereotypical expectations. There can be pressure to act altruistically in front of others and maintain a good self-image when there are real stakes. Knowing that their decisions do not impact anybody's earnings might ease the burden of fairness concerns and reveal participants' genuine views (Bruttel & Stolley, 2018). A hypothetical experiment may allow participants to avoid facing the consequences of their decisions (Wiseman & Levin, 1996).

Furthermore, participants may not relate to hypothetical people and can make risk-neutral predictions about men and women (Siegrist et al., 2002). Hsee and Weber (1997) suggest that people tend to reflect their perceptions of visible third parties while considering hypothetical agents to be more risk-neutral. However, Chakravarty et al. (2011) suggest that experimental agents do not differentiate between hypothetical and actual agents. In other words, people are likely to take more risks when making decisions regarding other people's earnings. Similarly, people are likely to display lower risk aversion in hypothetical situations (e.g., Holt & Laury, 2002, 2005; Harrison, 2006). Overall, since hypothetical cases do not impact anyone's earnings, decisions in this context are likely to be free from social pressures.

Unlike in sociology and psychology, survey experiments are not a standard tool in economics. Current studies in economics show that survey experiments appear to provide reliable insights into experimental agents' decisions.

5 Background on human language

This section discusses the relevant qualities of human language, including cognitive and neural features, and social characteristics. There is also information regarding the data type (e.g., private, public, written, and spoken samples). The essential background information given in this section's next and some further paragraphs mainly refers to Pennebaker's (2011) comprehensive work, *The Secret Life of Pronouns*, which discusses distinct studies, including the LIWC development process.

In the early 1980s, social psychologist James Pennebaker found that people who had experienced trauma benefited from reflective writing, that is, keeping a journal by elaborating past events through feelings and emotions related to these events. Although researchers are aware that writing is helpful in confronting a traumatic past, there is no explanation of how. Since every personal experience is unique, it is challenging to group traumatic events and make general assumptions about shifting psychological states. Moreover, even if one is an experienced clinical psychologist, the human mind cannot analyze the details of journal entries to identify changes in people's writing as they start to feel better.

In the early 1990s, computational linguistics began to bloom. However, there were no programs available to analyze trauma writings. To fill this gap, Pennebaker et al. developed the LIWC software. Although the initial idea was to analyze trauma essays, the project soon produced a well-accommodated dictionary.

In the beginning, Pennebaker et al. postulated that better-off individuals would use different content words than non-reflective writers. Indeed, expressive writing increases positive emotion word usage, while still elaborating on negative thoughts in the form of a well-composed story. These qualities differ from the writing a traumatized mind produces. As a reflection of the subconscious, scrambled thoughts and disorganized structures are typical of non-reflective writers. However, initial analyses showed no differences among the essays. It soon became clear that function words (e.g., pronouns, prepositions, articles) constitute the distinguishing factor to reveal the genuine human mindset. For example, people recovering from trauma shift their focus from themselves and onto others by employing fewer first-person singular pronouns (e.g., I, me) and more references to others (e.g., they, us).

"The" and "that" do not have meaning on their own. Rather, these small words serve as glue to create meaningful content. Most importantly, function words indicate how people deliver their messages. Each individual has a unique way of expressing themselves. An individual's style stems from their social and psychological status, education level, age, gender, and personality. Hence, function words provide broad insights into various human characteristics. For example, people from higher social ranks use first-person plural pronouns more frequently than people with a lower social status. This tendency also appears in the linguistic styles of people in leadership positions as compared to their subordinates.

Although they account for slightly more than half of the total word count in typical speech, humans do not pay particular attention to these small words because they are cognitively taxing. Therefore, scientists may need to perform multiple rounds of editing on their essay drafts, for example. Furthermore, it is difficult to control and alter function words while speaking. In everyday conversation, a self-focused speaker must make a solid effort to replace "I" with "we" in their speech. Furthermore, nonnative speakers may struggle to use articles correctly. Speakers of foreign languages acquired at an adult age are likely to suffer from permanent article and preposition misuse.

Content words reveal people's attentional focus, which provides the context for language. These words indicate an action or describe an object or a feeling. For example, "a red chair" creates a corresponding mental picture. Furthermore, a happy person may choose to describe their emotional state by stating “I am happy today” instead of saying “I do not feel sad today.” These two sentences convey the same meanings on the surface. However, the second sentence comes from a person with a more pessimistic mindset. Table 1 displays examples of function and content words.

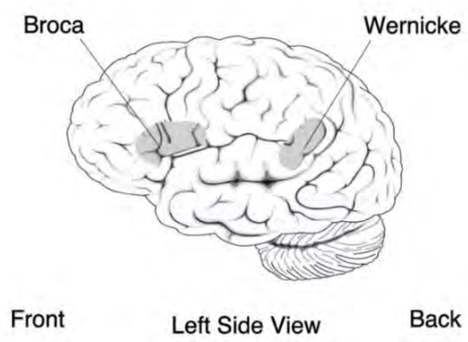
Function words		Content words	
Pronouns	they, it, us	Nouns	car, chair
Articles	a, an, the	Common verbs	to go, to write
Prepositions	over, up, on	Common adjectives	red, slow
Auxiliary verbs	is, are, have	Modifying adverbs	happily
Negations	never, no		
Conjunctions	because, but		
Quantifiers	most, less		
Common adverbs	very		

Source: The Secret Life of Pronouns (Pennebaker, 2011)

Table 1 Function words and content words with examples

Here, an interesting question is whether words give clues about the human psyche, or vice versa. Pennebaker and associates show that words are likely to reflect the content of the human mind. Intentionally using specific words is unlikely to create the desired effects. Experiments involving students have shown that incorporating specific words into their writing did not influence the students’ mental states. This manipulation of word groups changes the language leaders use in the public domain (Slatcher et al., 2007). For instance, changing pronouns to portray the image of a celebrated leader is likely to convey an aloof, disconnected leader when the person in question has no real bond with society.

Content and function words do not occur in a vacuum. Broca's area (which is located in the prefrontal cortex and plays a role in personality, social skills, and emotion regulation) is responsible for speech production. Using words accurately in written and spoken language and producing speech sounds are the main functions of Broca's area. Specifically, function words are primarily associated with this region. Wernicke's area (located in the temporal lobe) is responsible for speech comprehension and language processing. In other words, Wernicke's area produces content words. The association between these brain areas and word groups is a simplification. However, sustaining damage in one of those areas can prevent people from employing language properly. This connection between language and the brain indicates that humans cannot suppress their spontaneous speech or speak in contradiction to their reality.



Source: https://www.nidcd.nih.gov/sites/default/files/Documents/health/voice/Aphasia6-1-16.pdf

Figure 2 Broca's and Wernicke's areas

Boroditsky (2019) suggests that language makes humans unique. Language formation involves various neural processes and is part of brain activity. As an illustration, humans cannot understand the meaning of the word denoting the color red from the content of a conversation. In fact, it may be impossible to describe red to a visually impaired person. Since speech generation is a cognitive activity, people summon the mental picture of red when mentioning it. This process may also be expanded to attach different meanings to the same word. That is, people may not have the same shade of red in mind, or red may not evoke similar feelings universally. For some people, red can mean violence, while for others the color can be related to warm, positive emotions.

Similarly, not every language has the same word diversity for red. Some languages have a range of words available to describe the various shades of red, while in other languages, the spectrum of redness may be limited to a couple of words. These differences indicate that language carries cultural excerpts (Casaponsa & Athanasopoulos, 2018). The English language has different underpinnings in different native-speaking countries due to the evolution of separate societies with distinct histories and geographies. In addition to neural construction, all these exogenous features contribute to and shape human language.

This may be why some sociolinguistic studies tend to have a massive amount of data from various eras and English-speaking countries. The most particular study explores gender differences in 14,000 text samples from the United States, New Zealand, and England. Of those, 5,971 belong to men, and 8,353 belong to women. Two-thirds of the participants were undergraduate students (aged between 18 and 22 years). These data were collected 70 different international studies from the 17th century and the decades between 1980 and 2002 (Newman et al., 2008). The samples are from diverse contexts, including undergraduate psychology exams and assignments, talk show conversations, spontaneous descriptions of drawings (e.g., the Rorschach test), bestselling fiction novels, and written samples from two different experiments.

The findings of this study provide insights into gender differences in language use. Generally, female language is social, involved, and emotional. Women make more references to psychological processes, feelings, and the domestic sphere to describe human emotions and relationships. Unlike women, men pay attention to current concerns. Stereotypically, male language is informative and instrumental. Men employ words related to sports, finance, and professions. Hence, male language includes the frequent use of long words (i.e., words longer than six letters), prepositions, articles, numbers, and swear words. As part of male language's informative, technical style, Newman et al. (2008) suggest that long words indicate complex thinking. However, there is no concrete explanation for why men use long words more.

Men and women in a leadership context might employ long words to the same degree, given their higher education. Indeed, Yu (2014) suggests that long words appear at a comparable frequency in both male and female language in congressional speech. However, Slatcher et al. (2007) associate long words with male language, older age, and a candidate's presidential quality. Analysis of inaugural speech shows that long words mark male presidential language. In light of these findings, an explanation can depend on context. In formal environments (e.g., congress), people may use more complex language, whereas in casual settings (e.g., interviews), women might be more authentic and assume a storyteller stance (i.e., narrative style), while men may create distance by maintaining their formal approach. In addition, any analysis of US presidential samples produces more meaningful implications for male language, as there has not yet been a female president, and the country only swore in its first female vice president as recently as 2021.

Gender differences also extend to pronoun use. Compared to men, women use more first-person singular pronouns. This sort of pronoun use indicates that women have more self-focus while talking about events surrounding their lives from their perspective. Pennebaker (2011) suggests that the stereotypical characteristics of female language might match a depressive mindset, low social rank, and a younger person's linguistic style. In other words, suicidal and emotionally unstable people employ more first-person pronouns, and unhappy people focus more on themselves, as do younger people.

In contrast to the patterns observed for first-person singular pronoun use, Newman et al. (2008) found that men and women employ first-person plural pronouns (e.g., we, us) comparably. Social psychology and linguistics literature appears not to provide a clear-cut answer to whether women or men use "we" more frequently. Jones (2016) associates first-person plural pronouns with male language. However, the paper does not provide a well-documented motivation for male use of first-person plural pronouns. Nevertheless, this connection is likely related to the stereotypical association between men and high-ranking positions. Due to the social divide between men and women that has persisted across history, male language overlaps with the linguistic style of people with a higher social status, which is a consequence of the fact that leadership positions have traditionally been male-dominated (Coates, 2015).

Building on Newman et al.'s (2008) findings, Jones (2016) created two composite variables. First-person plural pronouns, articles, prepositions, words expressing anger, long words, and swear words are the ingredients of the masculine linguistic style. Similarly, the stereotypical characteristics of female language constitute the feminine linguistic style. Jones (2016) suggests that total pronouns (e.g., we, they), first-person singular pronouns, common verbs, auxiliary verbs (e.g., are, do, will), social references, emotion words, cognitive processes (e.g., think, know), and tentative words (e.g., guess, chance) are the most distinguishing characteristics of stereotypical female language. These word groups touch on women's social, involved linguistic style, pointing to women's fondness for telling stories about themselves and the people around them.

The interesting point here is why women employ first-person singular pronouns despite being sociable and other-oriented. In addition to the previously discussed reasons, Argamon et al. (2003) suggest that women's greater involvement in the social context results in higher usage of first-person singular pronouns and second-person pronouns. Women build social ties in their dyads, and in one-on-one relationships, women tend to share their personal views, while simultaneously focusing on what the other person thinks.

In contrast, Mulac et al. (2001) argue that first-person singular pronouns are a typical male language characteristic. This inclination appears to be explained by men's self-centered attitude compared to women's other-oriented attitude. However, Pennebaker et al. (2003) argue against these findings, explaining them as counter findings based on a significantly low number of samples.

Another interesting point is tentative words. Newman et al. (2008) show that women and men employ tentative words comparably. However, the motivation for this feminine classification, according to Jones (2016), can be associated with previous sociolinguistic findings. Lakoff (2004) distinguishes between powerful (i.e., male) and powerless (i.e., female) speech. Intensifiers (e.g., so, really), hedges (e.g., maybe), and tag questions (Aren't we?) are distinct characteristics of powerless speech. Argamon et al.'s (2003) computational study also supports that female language is more polite and employs more tag questions. These findings overlap with the tentativeness of female language.

Other similar word groups in both male and female language are sexual words, time references (e.g., until, season), conjunctions, word count (i.e., length of speech or writing), and anger words (e.g., hate). Among these gender-common categories, the most surprising finding relates to anger words. Stereotypically, anger is an emotion associated with men. In leadership positions, it is acceptable for women to display anger to a certain degree. Jones (2016) adds to this distinction, on the one hand, by including anger words as part of the masculine linguistic style. On the other hand, Jones (2016) suggests that women stereotypically employ more positive and negative emotion words. The latter includes anger words as a subcategory.

Keeping these findings from the literature in mind is crucial when evaluating arguments from linguistic studies with multi-sourced text data. Since Newman et al. (2008) gathered samples from different parts of the world, there might be inconsistencies regarding words' cultural meanings. For example, New Zealand English entails different cultural references than US English. Moreover, writing and speech samples come from different eras. A word may not have the same implication today as it did two centuries ago. Finally, the data come from a mix of adult and student participants. Although 14,000 text samples are sufficient to mitigate unbalanced data problems, it is still essential to evaluate data outcomes with a critical eye.

Human language includes many clues that refer to various human aspects. After analyzing the data, researchers may need to return to the text to explore the background. Not every word related to anxiety implies depression; rather, first-person singular pronouns may appear more frequently due to the data structure (e.g., interviews). Therefore, it may be misleading to immediately link these indicators with some implications.

While studies on LIWC have been increasing, Pennebaker and associates have released improved versions of the program over the years. Researchers have conducted studies to test LIWC software and run replications using new versions. First, in 1999, Pennebaker and King conducted an exploratory study to investigate personality using written text samples. This study attempted to test a newly created LIWC before its release in 2001.

The software is promising and capable of evaluating many data files concurrently. The findings suggest that written language provides broad insights into the five-factor personality framework. Mehl et al. (2006) extended these efforts by employing only spoken text samples. In contrast to a set of recordings from a laboratory environment, undergraduate students wore an electronically-activated recorder (EAR) for two consecutive days during their regular everyday life. EAR made recordings without the participants' active knowledge (e.g., 30 seconds on and 10 minutes off).

AI (Mairesse et al., 2007) has clarified the difference between written and spoken language using data from Pennebaker and King (1999) and Mehl et al. (2006). Multiple machine-learning experiments have revealed that spoken text samples produce the most reliable clues regarding personality. The disadvantage of AI studies is interpretability. For example, anger words appear to be positive indicators of emotional stability (Mairesse et al., 2007). Another problem is that of the experimental sample. Studies conducted with students may not translate into the adult world. Words related to studying may imply conscientiousness for students, while words related to work indicate conscientiousness for adults.

In a 2020 study, Tackman and associates eliminated these drawbacks by replicating Mehl et al. (2006). This study employs a recent version of the LIWC (2015). Adult participants from different backgrounds wore EAR for up to six days. The findings show that men and women do not appear to have significant personality differences.

In addition to the sample, Tackman et al. (2020) utilized linguistic summary variables in their analyses. The 2015 version of the LIWC has four composite scores. Analytical thinking, clout, authenticity, and emotional tone are a combination of function and content words (e.g., structural resemblance to the feminine and masculine linguistic style indices from Jones [2016]).

The analytical thinking score derived from a study conducted to explore whether function words can predict academic success. This study also tested whether a sole LIWC analysis would be sufficient to make a reliable conclusion about thinking patterns from essay samples. Pennebaker et al. (2014) explored answers to these questions using a sample of more than 50,000 college admissions essays belonging to approximately 25,000 applicants between 2004 and 2007. The results revealed that function words indeed predicted academic success during the four years of undergraduate studies. In addition, researchers recognized that LIWC alone serves as a reliable tool to explore human cognition through text data, without the need for another text analysis tool.

The findings on analytical thinking and academic success imply that highly-educated people use more abstract thinking and cognitive complexity. While researchers associate formal patterns with articles, higher usage of prepositions explains the latter pattern. Students who used more articles and prepositions in their admission essays employed a well-structured form, with formal and proper description of objects and events. Students who employed pronouns, auxiliary verbs, conjunctions, adverbs, and negations had lower analytical thinking scores. These students tended to tell personal stories, focusing on a specific time. In other words, narrative thinking or dynamic language (Pennebaker et al., 2014) is more social and less cognitively complex. Furthermore, students' academic success was not dependent on their academic field (e.g., engineering, humanities).

The clout score derived from a study to determine the relationship between pronouns and ranks in social hierarchies. Kacewicz et al. (2014) conducted four experiments and used written data analysis to explore pronoun usage between members of high and low ranks within social hierarchies. Undergraduate psychology students were the participants in the first three studies. The fourth study included graduate students and faculty members. The final study differs directionally with the analysis of letters written by lower and higher-ranking people in Saddam Hussein's regime.

Although the study is based on a non-representative sample, the researchers believe that the findings provide reliable insights. These findings have general implications for pronoun use across social ranks without strictly distinguishing leadership, power, and social status as unique concepts. Instead, Kacewicz et al. (2014) provide a brief discussion of their effects on these three notions. Leaders' features resemble those of persons from higher social ranks. The main difference is that leaders use power in a non-threatening or nonauthoritarian way. People who have a high social status and power bear a greater resemblance to leaders as they embrace members and support the common good.

People with a high social status use more first-person plural pronouns than those occupying lower ranks. First-person plural pronouns indicate a shared identity, shared goals, and an other-oriented perspective. Gardner and Avolio (1998) suggest that influential leaders stimulate a collective identity by employing "us," "we," and "our." To illustrate, flight teams led by pilots employing first-person plural pronouns experience better performance and higher team connectivity during their subsequent flights (Sexton & Helmreich, 2000).

Since higher social ranks are concomitant with rapport-oriented views, these people appear to use more second-person pronouns (e.g., you). This pronoun is the general outcome of focusing on others, supporting subordinates, resolving conflicts, and forging ties among group members. Higher-ranking people stay aligned with the necessities that arise from common interests (e.g., What is your idea on this matter?)

rather than focusing on themselves (e.g., What do I want?); they therefore use fewer first-person singular pronouns.

Indeed, first-person singular pronouns indicate a higher self-awareness and more attention to personal feelings and ideas. Here, low-ranking individuals' focus on others arises from what people holding higher ranks think of them as subordinates. Email conservations between managers and their assistants provide an illustrative example. In an email, the subordinate might need to explain their needs, actions, and motives in detail. This writing style requires a high concentration of first-person singular pronouns to grab the manager's attention.

To summarize, Kacewicz et al. (2014) suggest that people from higher social ranks use first-person plural pronouns and second-person pronouns more often. In contrast, people from lower social ranks employ first-person singular pronouns the most frequently. These findings provide insights into how people from different social ranks express themselves. Instead of evaluating these three pronoun groups, researchers may use the aggregate variable, namely the LIWC clout score (2015).

As a linguistic summary variable, the authenticity score derived from a study exploring deceptive language indicators (Newman et al., 2003). A higher authenticity score indicates that the speaker is honest and vulnerable. Authentic speakers are more abstract; they talk about general facts using other references (i.e., third-person pronouns). In contrast, a lower authenticity score implies distancing oneself from facts by creating a deceptive story. Hence, such people use fewer first-person singular pronouns. Creating a new story while talking about it is a cognitively exhausting process. Therefore, inauthentic language appears to be less sophisticated (i.e., fewer exclusive words; but, although). Deceptive talk features more motion verbs (e.g., go, talk) and fixation on a particular individual. Guilt about lying may also involve the use of more negative emotion words, specifically anxiety words.

Emotional tone is the net positive emotion that a person employs in their speech. Depending on the context, emotional tone serves as a tool to explore emotional well-being. Cohn et al. (2004) developed this linguistic summary variable to determine people's emotional positivity after the September 11 attacks. Researchers explored 1,084 samples from an online journal over four months (spanning two months before the attacks and two months after).

Compared to self-reports, personal journals reveal organic emotional fluctuations during a national crisis. In a general public sample, researchers' questions in the aftermath of a crisis are likely to influence personal responses. In addition, human memory may distort reality depending on the emotional disruption that a person undergoes. However, it may not be possible to peruse political leaders' personal journals.

In such cases, social psychologists generally rely on public speech samples. Comparable to private linguistic samples, this type of data may never fully reveal a leader's mindset. However, analyses of public samples still provide valuable insights into a leader's psyche amidst a national crisis. The findings show that the New York City mayor's language became more complex, emotional, personal, and future-focused in the aftermath of 9/11. These results indicate that a vulnerable person employs simple language to connect with the public. Complex language indicates a mind preoccupied with the future. Pennebaker and Lay (2002) obtained these findings by analyzing 35 press conferences given by a single leader.

Similarly, Jones (2016) explored 567 Hillary Clinton interview and debate transcripts from 1992 to 2013. Slatcher et al. (2007) focused on two presidents and two vice presidents in 2004, perusing 271 interview, debate, and press conference transcripts. Although studies exploring political leaders' characteristics have provided relevant information about linguistic styles, these studies utilized limited samples.

It might be wise to consider these findings with a grain of salt. In addition to small sample sizes, context is essential when exploring language. What suits political leaders may not apply to the corporate world. Politicians' public talks appeal to the masses and create a solid follower base. Societal norms tend to be stronger in this case. In contrast, corporate leaders may not be concerned about the general public. Indeed, not all CEOs want to be under the spotlight. Perhaps the language of CEOs and other high-ranking corporate executives is geared toward industrial appeal. However, some executives have become well-known public figures (e.g., Elon Musk, Jack Dorsey, Sheryl Sandberg). Future studies exploring such people's social media posts and interviews might offer insights into distinguishing the linguistic characteristics of the corporate world.

This section provided an overview of the characteristics of human language. In addition to discussing personality differences, the next section offers more detailed information about studies that have provided linguistic clues related to personality.

6 Conceptual framework

Societal beliefs are a part of people's lives. These expectations shape how men and women should be and how ideal men and women should behave. However, not everyone follows these beliefs. In other words, nonstereotypical people are part of every society. Management research and social sciences explore the economic outcomes, strategies, and characteristics of men and women who contravene descriptive and prescriptive beliefs.

In this context, the first essay explores the economic outcomes of nonstereotypical women by building on behavioral economics literature. Next, the second essay investigates nonstereotypical women's linguistic strategies, mainly through management literature using a sample of women who hold senior leadership positions. The upper echelons of corporate leadership constitute a stereotypically male domain. Therefore, women who pursue senior executive positions behave nonstereotypically. The third essay builds on the second essay as it employs the same data, with a focus on female leaders' personality traits to provide insights into nonstereotypical women's personalities. In addition to utilizing insights from management literature, the final essay also draws on an applied psychology research background.

The researcher explores nonstereotypical women through the lenses of three fields. The first field, behavioral economics, is a branch of economics that explains human behavior and decision-making processes through various factors, including cultural differences and cognitive misconceptions (Mullainathan & Thaler, 2000). Stereotypes may resemble cognitive misconceptions as a social concept (e.g., Kahneman, 2003). The human mind may implicitly operate according to descriptive and prescriptive norms. Consequently, people tend to misjudge men and women who defy societal expectations without questioning their behavioral motives and the context of these decisions. All else being equal, human cognition is highly accurate in sorting men from women based on significant differences in physical appearance (Bordalo et al., 2016). Expecting similar behavior from men and women who are associated with different environments or who have different skillsets and backgrounds is most likely to fail. For example, a woman who works in the finance industry is more likely to be a risk seeker than a woman working in a non-competitive environment (e.g., Adams & Ragunathan, 2017).

However, possessing this knowledge may not immediately change people's attitudes toward nonstereotypical women. Hence, the inspiration for the first essay came from a media article on discrimination against top female senior executives (Chira, 2017). The article discusses such discrimination as twofold. Firstly, some female senior executives almost become CEOs, only to end up in secondary positions. Second, discrimination can persist in the boardroom and work against even those women who do manage to become CEOs. That is, female CEOs are likely to lose their position whenever their strategies fail. The article suggests that in the boardroom, risky decisions from male versus female leaders may not be viewed with the same attitude. Women are prone to lose their position when their risky decisions do not pay off. In contrast, the executive boardroom appears to compensate for male leaders' losses.

Although the abovementioned article is anecdotal, studies also show that women operating outside their gender domain may be subject to stricter evaluations when they fail. Sarsons (2017 b) suggests that female surgeons might receive fewer referrals than male surgeons after an unsuccessful operation. Similarly, compared to male advisors, female financial advisors are more likely to be fired due to misconduct. Moreover, these severed female advisors may struggle to find new jobs (Egan et al., 2017). For women, operating in a nonstereotypical domain means defying societal expectations.

As is the case with certain occupational domains, people associate specific behaviors with men or women. For example, previous studies show that women (Aguiar et al., 2009) or both women and men (Brañas-Garza et al., 2018) are inclined to perceive females as altruistic, less competitive, and embodying the nurturer or caretaker role (e.g., characteristics suitable for stereotypically feminine positions in the labor market).

Similarly, people stereotypically attribute risk-seeking to men and risk-averse tendencies to women. In line with stereotypical expectations, studies in behavioral economics show that women tend to take fewer risks than men (e.g., Croson & Gneezy, 2009; Charness & Gneezy, 2012; Dohmen et al., 2011). Findings from applied psychology literature also support that women are likely to be more risk-averse in various activities, including smoking, sexual behavior, and driving (Byrnes et al., 1999).

Due to stereotypical expectations, people may fail to recognize actual behavior (e.g., Schubert et al., 1999). Women's inclination toward risk can be context-specific or may be related to individual women's characteristics. In short, women are likely to be comparable to men. Differences can also diminish with age (e.g., Byrnes et al., 1999; Friedl et al., 2020). However, stereotypical expectations can exacerbate the gap between men and women.

Specifically, people tend to perceive women as more risk-averse than is actually the case (e.g., Ball et al., 2010; Eckel & Grossman, 2002; Grossman & Lugovskyy, 2011). Eckel and Grossman (2008) also support this general conclusion regarding stereotypical expectations after reviewing experimental studies with different instruments (e.g., abstract gambling experiments, field studies).

Unlike the current literature, the findings presented in the first essay come from a survey experiment. Although the main result shows that stereotypes do not influence spectators' decisions regarding risk-taking women, further findings partially overlap with previous literature, in which different methodologies have been employed. Specifically, compared to unlucky female workers, men tend to leave unlucky male workers with no earnings. Implicit biases did not influence these decisions.

Experimental studies show that people operating outside their stereotypical domain face some bias. Hence, the main finding might be related to either the experimental design or the spectators' fairness concerns. Firstly, exerting no discernible impact on hypothetical workers' earnings is unlikely to evoke bias. For example, Booth and Nolen (2012) suggest that a hypothetical stake presented with survey questions and a real stake presented with lotteries may not produce similar outcomes. However, in such a case, there would be no difference between male and female spectators. Second, the short scenario describes a work situation that resembles an online task. An assumption could be that spectators might relate to the workers in the hypothetical case (e.g., a shared identity; Konow et al., 2020). Thus, spectators are likely to show genuine fairness concerns about even hypothetical workers.

Compared to male spectators, female spectators have egalitarian inclinations toward risk-taking male and female workers. This finding supports the hypothesis that female spectators may have more robust fairness concerns than male spectators. Moreover, findings from different experimental settings support that women may have more substantial fairness concerns than men (e.g., Heinz et al., 2012; Krysowski & Tremewan, 2021; Sharma, 2015).

Alternatively, women may hold nonstereotypical women to higher standards premised on the consideration that men may not have an advantage over their female counterparts. Evidence from committee studies (e.g., Bagues et al., 2017; Broder, 1993) suggests that women may assess other women more strictly in nonstereotypical contexts. Boards and selection committees may also subject women to stricter evaluations than their male counterparts in environments where there are relatively fewer women. This argument may translate into an alternative explanation of female spectators’ indifference toward female workers. Since the risk domain could be a nonstereotypical context, female spectators might not consider female workers to be at a significant disadvantage.

In contrast to the risk domain, female spectators impose inequality between male and female workers in a merit environment. This outcome appears to be related to seeing men as more advantageous in this particular setting. More specifically, female spectators tend to view male workers as expending less effort than comparably unsuccessful female workers. Consequently, female spectators tend to compensate for women when they fall behind. However, men do not receive similar compensation when they are unsuccessful (Cappelen et al., 2019).

In addition to the hypothetical stakes, the first essay applies an implicit association test (IAT) to further investigate stereotypes. This measure aims to understand whether spectators’ latent beliefs have an impact on their decisions. Self-reports may not be sufficient to determine implicit beliefs about stereotypes. As in the case of any self-report inventory, people may provide politically correct answers. Such information is likely to be insufficiently illustrative to understand genuine beliefs.

Furthermore, people may be unaware of their stereotypical beliefs. For this reason, there is a distinction between conscious and implicit beliefs. While the former operate within a person's awareness, the latter may exist in an individual’s subconscious as the outcome of past experiences and cultural norms. Studies in psychology show that IAT illuminates people’s underlying beliefs (e.g., Greenwald et al., 2009; Greenwald et al., 2015). However, there are mixed findings in the economics literature.

Bertrand et al. (2005) suggest that IAT reveals implicit attitudes, especially when people are distracted or facing time pressure or ambiguity. These situations arise when researchers instruct participants to engage in multiple tasks simultaneously. For example, Bertrand and Mullainathan’s (2004) field experiment shows that people tend to eliminate resumes with ethnic names due to ambiguity and time pressure. Specifically, because experimental agents have no clear guidelines for defining a successful profile (i.e., ambiguity), they must sort resumes as quickly as possible (i.e., time pressure). Under these conditions, it is easy to follow implicit associations and discriminate unintentionally.

In other words, when there is a cognitive load, people may reveal the deep-rooted prejudices and biases that exist in their minds. By virtue of its structure, IAT can easily capture these types of associations. The task requires that experimental agents press the left or right keys as quickly as possible when matching a target in the middle of the computer screen with concepts in the upper left or right corner.

Reuben et al. (2014) show that IAT is powerful enough to identify implicit gendered stereotypical beliefs that influence male and female behavior toward women in the context of STEM. However, Lee (2018) argues that IAT fails to explain racial biases in dictators’ decisions. Specifically, the study suggests that people might carry implicit beliefs without acting on them. In both laboratory experiments, the participants completed the IAT as the final step of the experiment, similar to the design in the first essay.

Based on the arguments from economics literature, there could be alternative explanations for why IAT fails to explain bias in the first essay. Spectators may be highly aware of the purpose of IAT and answer accordingly. As a feature of this test, prolonged answers do not serve as valid results. The general findings show that the test reveals implicit male stereotypes. Thus, spectators may not answer the questions at random. Alternatively, the design does not limit the duration of the IAT. Since there is no time pressure, spectators might have a lower cognitive burden, which can mask underlying beliefs. In contrast to these two possible explanations, implicit biases may not impact spectators' decisions automatically.

In part, this essay contributes to the economics literature by exploring evaluation bias. As already discussed, studies show that women who engage in nonstereotypical behavior are subject to strict evaluation. The first essay provides alternative evidence showing that when women fail, they may not face punishment. Due to stereotypical associations between setting and gender, male-dominated environments may consider women more likely to fail. This implies that men may associate failure with women in male-dominated environments and react by providing some sort of protection. Thus, this tendency can be understood as the outcome of benevolent sexism (Glick & Fiske, 1996).

Furthermore, the results may provide alternative implications for why women fail to advance in male-dominated occupational segments or risk-involving positions. Since women appear to be associated with failure, nonstereotypical environments may not give them the challenging tasks needed to demonstrate their skills or capability to ascend to higher ranks (e.g., King et al., 2012). These implications may shed light on the lack of female participation in STEM, AI, and various senior leadership positions.

In addition to this essay's contributions to the behavioral economics literature on stereotypes, the present research makes broad methodological contributions. The first essay shows that hypothetical designs can provide insights into human decisions. Survey experiments may serve as an alternative tool to explore the impact of stereotypes on economic outcomes.

This section discusses nonstereotypical women's attitudes and outcomes, which have been explored in the economics[3] and management literature. Research in this context also investigates methods to overcome bias in nonstereotypical domains. A significant example is Goldin and Rouse's (2000) famous symphony study. This paper illustrates how blind audition procedures implemented in the 1970s increased the number of women in the US symphony orchestras. Previously, hiring directors believed that female musicians might not be skillful enough to be part of a symphony orchestra. Blind auditions help 50% of female musicians advance past their initial audition. Due to impartial screening, women now make up 30% of hires.

While blind auditions appear to be effective in limiting the application of stereotypical beliefs, Bohnet et al. (2016) conducted an experimental study to explore whether joint evaluations would produce a comparable effect. The findings show that the joint evaluation of candidates in gender-stereotyped tasks can reduce discrimination. In contrast, separate evaluations tend to suffer from gender bias (e.g., associating men with quantitative tasks and women with qualitative tasks). Joint reviews are likely to prompt evaluators to focus on performance rather than candidates' gender.

These studies have focused on the limiting impact of stereotypical expectations. The findings demonstrate how alternative evaluation practices can open pathways for highly-skilled women in nonstereotypical domains. In addition to the relevance and effectiveness of institutionalized practices, women can apply individual strategies to ease the burden of stereotypes. For instance, senior executives' self-presentation strategies may provide insights into how women operate in

[3] Economics literature mostly explores sex-based discrimination through the lenses of taste-based discrimination (Becker, 1971) and statistical discrimination (Arrow, 1973). In broad terms, the former theory explains labor market outcomes through sexism. In contrast, the latter theory argues that inequality among different groups in society may arise from each having imperfect information about one another.

nonstereotypical settings. This topic constitutes the primary subject in management studies.

Broadly, management research explores stereotypes in an organizational context (e.g., Brenner et al., 1989). Theoretically, research on this axis borrows insights from psychology and sociology. Studies explore how women present themselves in different managerial positions (e.g., Bolino et al., 2016), including the strategies women utilize while engaging in nonstereotypical tasks (e.g., Bowles & Babcock, 2013). Furthermore, these studies focus on the outcomes of women who emphasize their nonstereotypical features (e.g., Amanatullah & Morris, 2010; Guadagno & Cialdini, 2007).

Rudman and Glick (1999) discuss the outcomes of agentic women in middle managerial positions. Communal and agentic traits are among the desired middle manager characteristics. When women promote themselves as agentic candidates to obtain such positions, they appear to have fewer social skills than agentic male prospects. In other words, women's agentic characteristics may help them present themselves as hirable, but they also represent women as socially insufficient. In a previous paper, Rudman (1998) defined this dislike of agentic women as a "backlash effect." Women are likely to mitigate backlash by promoting communal characteristics. In so doing, women appear to be likable but incompetent. Hence, the impression management dilemma refers to an alternating misfit between agentic and communal characteristics for women in a nonstereotypical context (e.g., Phelan & Rudman, 2010; Rudman et al., 2012).

Impression management originates from sociology as a notion of self-promotion in various environments. The desired image in a particular setting is a blend of societal expectations and personal characteristics. The critical point here is to draw a genuine picture and not impose something that is different from reality (e.g., Goffman, 1959). Altering images to meet contextual expectations may backfire. For example, portraying a relatively masculine image in their cover letters may not increase women's chances of being hired in stereotypically male fields (He & Kang, 2019).

Indeed, context plays an essential role in how women portray themselves. Stereotype defying women appear to be more acceptable in mixed or female dominated settings (Smith et al., 2013). Conversely, in male-dominated environments, being more agentic may damage women's image and engender dislike (e.g., Guadagno & Cialdini, 2007). It is essential to highlight that men's attitudes are not the sole trigger behind backlash. Rudman (1998) suggests that assertive women may also face dislike from fellow females.

Women can dodge backlash by taming their agentic impressions. In negotiations, women's assertiveness engenders dislike when they make demands in their own best interest. However, aggressively negotiating for others does not evoke backlash for women. In the former case, women defy societal expectations. The latter situation complies with the communal side of the gender stereotype by emphasizing benevolence for others. In light of this, when making personal demands, women may be less competitive and ask for less in order to avoid backlash (Amanatullah & Morris, 2010). Bowles and Babcock (2013) suggest that women can receive their full benefits and avoid backlash by promoting personal demands as beneficial for the common good.

Similarly, female managers can create an aggressive impression as long as they highlight some communal aspects within their strategy (Rudman & Phelan, 2008). Furthermore, to mitigate dislike, women may communicate their communal features explicitly in their nonstereotypical environment. The most obvious form is to provide information about additional roles related to family and motherhood (Heilman & Okimoto, 2007). When this cannot be done, women may conceal their agentic side by refraining from manifesting themselves as overtly dominating (Williams & Tiedens, 2016).

Overall, an assertive image may enhance male leaders' favorability, but for female leaders, it is likely to engender dislike. Specifically, people may express dislike toward female displays of power (Rudman & Kilianski, 2000). In the political leadership domain, this dislike might morph into severe backlash when people view women as desiring power, regardless of whether this is the leader's actual motive or merely spectators' perception. Furthermore, both men and women are likely to react with backlash against power-seeking female politicians (Okimoto & Brescoll, 2010).

Alternatively, the following stereotypes are likely to help women build a favorable image. Even if backlash can be avoided, this strategy can only help to a limited extent (Bolino & Turnley, 2003). A communal image may fall short of creating an adequate leader impression (Heilman, 2001; 2012). The only exception may arise in top leadership positions. Communal characteristics can strengthen senior leaders' image. In other words, proving oneself as an accomplished leader may eliminate question marks about female leaders' competence. In these positions, female leaders appear to be both more agentic and more communal than their male counterparts; possessing both qualities portrays these women as highly competent leaders. However, this favorable attitude does not extend to women in junior and midlevel positions (Rosette & Tost, 2010).

Apart from gender, ethnic background or race is another feature that leaders may utilize as part of their impression management strategy (Bolino et al., 2016). As an anecdotal example, in the political leadership arena, a leader, such as a presidential candidate, could claim membership to a particular ethnic group in order to reach a specific segment of voters. However, in cases where such a claim has no basis in reality (e.g., the leader did not grow up in the relevant society and did not experience similar struggles), these claims may only hurt the candidate's image (Nilsen, 2019).

There are few public examples of top leaders' self-promotion failures through genetic distribution. However, publicly available speech samples can reveal leaders' impression management strategies. The political science literature suggests that female leaders employ an agentic linguistic style as statespeople (e.g., Cameron, 2005; Jones, 2016). This strategy is so diffused in the political context that when describing female politicians with higher chances of election victory, media reports employ fewer female pronouns (von der Malsburg et al., 2020).

While these findings provide insights into how political leaders should portray themselves, it is unclear whether these strategies apply to senior corporate executives. Indeed, when it comes to gender and stereotypes, few studies focus on how women in top corporate positions strategize through language. Moreover, as already discussed, the psychology and political science literature investigate political leaders using a limited sample size.

Unlike the current literature (e.g., Choudhury et al., 2019; Jones, 2016; Pennebaker & Lay, 2002), the second essay provides insights from a large sample of 522 top leaders. The sample includes 1,082 text samples, of which 850 are in interview style, and 232 are either talks or essays. The total interview and essays/talks word counts are 1,333,929 and 446,133, respectively. Furthermore, the dataset includes questions from 850 interviews with a word count of 260,901.

The second essay employs a diverse range of publicly available text data from these leaders. The dataset includes interviews from newspapers, lifestyle magazines, popular podcasts, talks, and essays from various media outlets. The nonsingularity of the sources provides a wholesome public image of the executives rather than speaking to a particular environment (e.g., corporate, industry).

Overall, the findings suggest that senior female executives tone down their agentic features with communal characteristics. That is, these leaders balance the agentic features that are concomitant with their professional position with the communal features of being a woman. In terms of linguistic style, female leaders speak in a more agentic fashion than men in the general population, but their style is less agentic than male leaders'. Conversely, this inclination appears to shift in female-dominated industries. Based on the second essay's sample, female leaders from the consumer discretionary sector (Desilver, 2018) tend to deviate from societal norms and speak more agentically.

As discussed in Section 3, female-dominated environments are likely to allow women to circumvent societal norms. A similar welcoming attitude may also exist in gender-diverse environments. For example, boards with influential female executives are more likely to appoint female CEOs than boards with no women. Here, the differential point is not the number of females, but the power of female executives (Cook & Glass, 2015). This finding may also indicate that consumer discretionary sector leaders might be highly influential, despite mostly occupying non-CEO positions. Hence, women in the sector may have the flexibility to circumvent gender norms.

In addition, female leaders emphasize themselves while talking about their professional roles. These women's use of first-person singular pronouns may indicate that the upper echelons do not accept women as team members. Female executives' answers tend to highlight their individual contributions to the corporation. In contrast, male executives tend to use first-person plural pronouns to emphasize their efforts as part of a group and not as a single person. In other words, this finding may point to the extra individual effort female leaders invest in an attempt to become part of the team (e.g., Barbulescu & Bidwell, 2013).

Under such circumstances, female leaders may receive less support from their male counterparts, especially in positions involving high risk (e.g., the glass cliff; Glass & Cook, 2016). Although the second essay does not offer any implications regarding such positions, one way to explain the difference in pronoun use may arise from having sole responsibility for challenging tasks. Female executives may need to invest extra effort to compensate for being a woman in a nonstereotypical environment. Hence, women may be inclined to accept demanding jobs in order to prove themselves in their professional contexts. This choice may come at a price. Female leaders may not receive much support when undertaking these challenges; in other words, they may be on their own (e.g., Ryan et al., 2016).

Regarding content words, an intriguing finding is that female leaders appear to portray a maternal image. An implication could be that for female leaders, a "strong mother" image may not evoke as much bias as a plainly strong leader image. This finding overlaps with those of previous backlash studies and also supports recent anecdotes from the political arena, in which leading female public figures appear to legitimize their image in male-dominated environments by accentuating their maternal identity (e.g., Bennett, 2021; Miller & Haridasani Gupta, 2020).

The literature provides insights into the general characteristics of female corporate leaders. The overall findings show that these women still have stereotypical features (e.g., agreeableness). However, survey studies show that nonstereotypical women are comparable to their male counterparts in terms of nonstereotypical characteristics (e.g., competitiveness and assertiveness) (e.g., Adams & Funk, 2012; Wille et al., 2018).

While these findings are illustrative, the data in these studies come from self-administered inventories. Surveys may provide biased results because responders can manipulate their choices to meet the desired criteria instead of providing a genuine self-description. In contrast, linguistic samples provide a more realistic picture of personality. Still, social expectations can impact how people express and perceive the Big Five traits.

The Social Role Theory highlights that it is impossible to disentangle social roles from personality. Prescriptive norms impact what men and women display to the outside world (Eagly, 1987; Mehl et al., 2006). Studies in sociolinguistics have voiced these concerns as the tangle between image and personality. Hence, what people present to the outside world is their personality expression rather than their genuine character.

Personality expression occurs with *identity claims* and *behavioral residue* (Gosling et al., 2002). Identity claims define an individual's efforts on how they wish to be perceived. In other words, image is comparable to personality expression or *persona,* which is a collection of characteristics that a person aims to convey to their environment (Ibarra, 1999). Behavioral residue refers to physical traces of a person that were left unintentionally.

When it comes to linguistic data, the behavioral residue refers to the content of a speech (Schwartz et al., 2013). However, public speech samples (e.g., interviews, university talks, ted talks) may not reflect subconscious traces. Communication advisors can influence the personality expression of senior executives. Due to this factor, public speech samples mostly reflect conscious traces rather than subconscious traces.

Focusing on particular environments when collecting speech samples may help identify contextual factors' influence on personality expression. Indeed, management scholars have recently started to study CEOs' language-based personality features by analyzing conference calls (Harrison et al., 2019) and social media posts (Wang & Chen, 2020). Since these studies focus more on contextual elements, the text data come from a particular environment.

Studies on personality and social psychology have identified language-based personality differences between men and women. These findings come from general public participants, undergraduate students, and text samples taken from social media data. Although these studies' methodological approaches yield valuable insights, the sample may provide only a limited understanding of the senior executive context. That is, these studies may not offer relevant insights into nonstereotypical women's characteristics. Hence, the third essay in this dissertation advances the literature a step further by focusing solely on female executives' personalities.

The findings presented in the third essay overlap with the current literature: Female leaders are more agreeable than male leaders. Interestingly, personality psychology studies suggest that women are likely to be more agreeable than men in a general public sample. Similarly, women in the senior corporate context may be more compassionate toward and supportive of others. In turn, this characteristic is likely to appear in women's general leadership style.

Agreeableness is a stereotypically communal trait that builds relationship bonds. In the leadership context, women may lean toward this communal trait to help strengthen their position. Additionally, this trait may help build networks that can be tapped to climb the ranks.

However, a single trait is likely not sufficiently illustrative to support these arguments. In the rest of the dimensions, female leaders appear to be more extraverted but less conscientious, open, and emotionally stable. Among these characteristics, extraversion appears to be the most crucial personality feature in a leadership context. Text data can provide insights into leaders' warmth through the use of positive emotion words. Researchers have observed similar extraversion characteristics in leaders' linguistic samples (Bono & Judge, 2004). However, the extent to which extraversion arises from dominance is unclear. Perhaps language-based extraversion speaks more to female than male characteristics. Hardworking target-oriented people are likely to be naturally conscientious. That female leaders lack this trait appears to be an outcome drawn from content (e.g., talking about product lines and gender issues). Openness indicators are mostly negatively associated with the social style of the general female language. In addition, it might be challenging to observe openness to experience based on text samples (Mehl et al., 2006). Like conscientiousness, emotional stability appears to be a natural trait among leaders. The third essay indicates that female leaders are likely to talk about the emotional experience of climbing the ladder.

When exploring personality through publicly available data, it is essential to remember that people may not fully reveal their identity. This is why sociolinguistic research relies on daily life recordings. The third essay applies findings from a study (i.e., Tackman et al., 2020) with no resemblance to any nonstereotypical context (e.g., STEM workers, any particular leadership context). Therefore, linguistic indicators may not be sufficiently illustrative for a senior executive sample.

Interestingly, Tackman et al. (2020) report no personality differences between men and women. However, their results come from a regular flow of daily life recordings. Unlike in an interview, there is no disruption during the course of speech. Indeed, more than half of the samples in the third essay come from interview-style texts. Answers to questions inadvertently impact personality scores. Female leaders may receive more questions regarding gender issues. In this case, women may appear to be less conscientious. Questions regarding social contexts (e.g., various relationships) can also negatively impact openness indicators. Most neuroticism categories, which overlap with extraversion, also appear positively in female speech.

An alternative explanation could be that linguistic samples from interviews might reveal what people want to see in a public figure's personality rather than these people's true identity. It might become ambiguous as to what extent female leaders talk about topics that contribute to agreeableness as a strategy versus out of personal willingness. The latter motivation may indicate that female leaders are more open about their lives during interviews than male leaders.

Apart from personal willingness, the third essay illustrates public expectations about how male and female leaders should behave. For example, female leaders may receive more questions regarding the work–family balance. In contrast, interviewers may assume that male leaders have no demanding family obligations. Thus, male leaders may not receive questions about fatherhood. A similar scenario could involve corporate policies regarding a gender egalitarian approach and paid leave for women. Interviewers may consider these topics to be more relevant to female leaders. Consequently, text data of this nature indirectly reveal societal expectations of male and female leaders.

Overall, these findings may indicate that female leaders have characteristics that may overlap with the female stereotype, implying that nonstereotypical women may not be strictly different from their general fellow females. Alternatively, in parallel with Kark et al.'s (2012) findings, female leaders may have comparably high agentic and communal qualities, assuming that women in this context are effective leaders.

As social psychology and personality literature suggests, it may not be possible to divorce cultural implications and contextual elements from personality manifestation. In this case, the Big Five features may somewhat resemble the strategies utilized by women in upper echelons who balance communal and agentic qualities.

As already discussed in detail, all three essays take the nonstereotypical woman as the central theme. The first essay explores whether society sees nonstereotypical women as entitled to their gains. The modified dictator game serves as a tool for investigating this attitude. In the second essay, senior female executives serve as a sample to explore the strategies women utilize in a nonstereotypical domain. The final essay partially utilizes the second essay's data by focusing solely on nonstereotypical women's personality features. Similarly, the second and the third essays employ linguistic measures and quantitative text analyses.

Essay	Name	Main Literature	Context	Methodology
1	Gender, Inequality, and Risk-Taking	Behavioral Economics	Nonstereotypical women from the general public	Experiments
2	Verbal Impression Management Strategies by Top Female Executives	Management	Women from a nonstereotypical environment	Text analysis
3	Personality Expression by Language among Business Executives	Applied Psychology & Management	Personality expression of nonstereotypical women	Text analysis

Table 2 The elements of the three essays

7 Methodology

This dissertation combines two approaches. The first essay is an experimental study, and the second and third essays employ linguistic analyses. Having two main empirical frameworks opens up different possibilities for research prospects. For example, in experimental studies, linguistic data are more likely to prevail as a means of gaining insights into human characteristics. Word groups associated with emotional states in a spoken context can replace survey assessments (e.g., Proto et al., 2019).

Similarly, experiments may be necessary in studies that mainly employ text data. Although the experimental approach in economics does not fully overlap with psychology, both disciplines draw upon a general common background in terms of experimental design. This knowledge provides essential elements for conducting experiments. Specifically, applied psychology studies employ experiments to determine the validity and reliability of linguistic measures (Pennebaker & King, 1999). In turn, these linguistic measures serve as personality indicators in various studies.

Alternatively, as in this dissertation, it is possible to employ these two empirical approaches separately. Not every experiment uses textual measures to determine participant characteristics. Furthermore, studies can rely solely on linguistic measures to explore a sample's characteristics. The first subsection provides insights into conducting experiments in online labor markets. The second subsection provides essential details about the methodology employed in the second and third essays.

7.1 Online survey experiments

Amazon Mechanical Turk (mTurk) provides the opportunity to collect a large amount of data in a short period. Researchers can compile data from a large number of participants without having to establish a laboratory setting and create a participation schedule. Furthermore, many experimental agents can concurrently complete a published individual effort human intelligence task (HIT). Since there are no physical space limitations, researchers can obtain an appropriate amount of data within a single day.

There are several advantages to conducting online experiments. Firstly, studies employing inventories initially published in English can make HITs available to native English speakers only. Second, global participants are within reach. Researchers have the advantage of focusing on a particular country or specific regions without any physical outreach concerns. Third, the global participants are adults, which may serve as an advantage over using bachelor students. Buhrmester et al. (2016) suggest that the quality of collected data meets psychometric standards and provides a compelling alternative with a diverse worker population as opposed to undergraduate college students.

mTurk specifically serves as a suitable setting for survey experiments in which participants read and answer questions. However, the inability to observe participants raises some concerns among researchers. The main problem is not being able to see whether participants are concentrating on the task at hand. However, although it is possible to track experiment completion time, there is no guarantee that participants are entirely focused even in the lab setting. An unfamiliar artificial environment may distract people from the decision-making process. For example, a person might seem fully engaged when they are actually preoccupied with their thoughts. Another concern is whether the participants can complete the experiment thoroughly. To help assuage this concern, mTurk provides the experimental agents' HIT acceptance rate. This information details whether a person has a history of being committed to experiments.

7.2 Natural language processing (text analysis)

Natural language processing (NLP) integrates empirical approaches from computer science and AI as a linguistic subfield. Through NLP, it is possible to explore large amounts of data in a significantly short time. Indeed, researchers from various fields have gained a vast amount of information about the relationship between human language and human characteristics during the last three decades.

This in-depth knowledge is the outcome of computational approaches. Multiple applications are available for analyzing text data. The dictionary-based methodology, feature selection, statistical methods, and supervised learning algorithms are part of this empirical approach.

7.2.1 Data details

The dataset includes written and spoken speech samples and demographic and organizational details to measure the linguistic styles of 522 executives ($N_{male} = 318$, $N_{female} = 204$). The executive sample is from the *NYT* "Corner Office" column, which appears in the Sunday business section.[4] The interviewer was Adam Bryant, who ran the column from March 2009 to October 2017.

The main reason for choosing the *NYT* is its status as a globally well-established news outlet ("Prizes and Awards," 2021). We assume that executives recognized by the *NYT* are "successful" professionals. However, talking to a highly prestigious outlet may not reveal much about an executive's actual characteristics. Thus, the data contain additional text samples to the "Corner Office" interviews.

[4] https://www.nytimes.com/column/corner-office

This aim aligns with the motivation to expand the data with as many words as possible. Hence, for each executive, there is at least one additional text sample gathered from five different sources. There are six main media outlets and 1,082 linguistic samples, of which 48% are from the *NYT*, 22% are from alternative interviews, 4% are from podcasts, 16% are from YouTube captions, 5% are from essays, and 4% are from quotes. There is no particular balance between these additional sources. Table 3 lists the word counts. Although the *NYT* interviews account for half of the sources, podcasts have a higher word count due to the limited space and timeframe. A podcast may have a duration of up to 3 hours, which can translate into a 50-page transcript. In contrast, a regular interview has a limited number of questions and a standard transcript length of one or two pages. Consequently, it is the number of words in a source rather than the share of media outlets that shapes the composition of the data.

Source	Male			Female		
	N	M	SD	N	M	SD
Essay	40	1,108.22	1,283.50	18	927.50	584.81
Caption	107	2,694.68	2,375.64	69	2,965.03	3,660.82
Podcasts	28	6,074.29	4,724.36	15	5,805.33	3,506.37
Interviews	130	1,508.00	1,415.98	105	1,167.96	612.99
Other	31	690.06	543.80	17	673.71	769.83
The New York Times	318	1,189.97	361.96	204	1,171.61	357.50

Table 3 Descriptive statistics for the text based on media outlet and gender

There are two main reasons for the lack of balance. Firstly, this sample is comprised of executives who migrated to the country as young adults. When this is the case, it can become a challenge to find alternative interviews conducted in English. For this reason, there are supplementary essays and quotes, for example. Second, not every executive is popular or has attained celebrity status. While some executives have been interviewed and reported on at various media outlets, many executives have more reserved profiles.

Thus, collecting additional sources was a random process. Indeed, there is also variety within each additional source. For example, alternative interviews include samples from professional and corporate-related outlets and lifestyle magazines. Conversely, podcasts have a more consistent structure, although they are collected from various programs. Most of these talks are casual interviews that do not focus on one particular issue; rather, they are concerned with the interviewee's overall life. Podcast transcripts are publicly available and only lightly edited.

YouTube captions include raw text data downloaded directly from videos. The captions include video samples clipped from alternative interviews and podcasts, in addition to talks, presentations, and spontaneous interviews. Essays range between professional and personal texts, including blog posts, opinion pieces, and formal articles. Quotes refer to short texts. This group includes excerpts from news reports and quick responses to questionnaire-style interviews.

Given the multi-source structure of the data, slicing is possible through media outlets or interview style (i.e., whether the text comes from answers to a set of questions, or from a talk or essay). For further analysis, the data also include questions from interview-style texts.

Executives' demographic characteristics include an MBA degree, native speaker status, and age. Information about MBA degree and age comes from the text samples, Google, Wikipedia, LinkedIn, or a Bloomberg search. When age information was not available, the data were approximated based on the executive's graduation year. The native speaker criterion refers to whether an executive was born and raised in the United States. Organizational characteristics refer to the publicly traded status and the company's industrial sector. Sector information was gathered based on the Bloomberg industry classification system. Companies' publicly traded status was obtained via a Google search.

Women comprise 51% of consumer discretionary executives ($N_{CD} = 159$), 43% of communications executives ($N_C = 106$), 22% of technology executives ($N_T = 143$), and 39% of others (i.e., consumer staples, financials, government, industrials, health care, and utilities, $N_{other} = 114$). Table 4 presents descriptions of the remaining characteristics.

Feature	Male		Female	
	M	SD	M	SD
MBA Degree of the executive	0.31	0.47	0.24	0.43
Publicly traded status of the company	0.34	0.48	0.11	0.31
Native speaker status of the executive	0.73	0.44	0.87	0.33
Age of the executive	56.19	10.22	56.07	9.05

Table 4 Descriptive statistics for the individual and company characteristics

7.2.2 Dictionary approach: LIWC

Iliev et al. (2015) have suggested that the LIWC is the most popular software for dictionary-based methodologies. There are three different versions of this dictionary. Pennebaker et al. released the LIWC in 2001, followed by an improved version in 2007 (Pennebaker et al., 2007). The most recent release was in 2015 (Pennebaker et al., 2015). Each version expands on the previous version. All dictionary versions have both reliability and external validity.

Tausczik and Pennebaker (2010) provide a detailed review of the studies conducted using the early versions of the LIWC. Pennebaker (2011) also discusses the essential characteristics of the LIWC methodology. Insights related to employing and analyzing LIWC data have been accumulated from a wide range of applications in psychology studies. In addition, political science research utilizes the LIWC to investigate readily available speech samples from political leaders. The implications of these studies also employ prolific approaches to the LIWC.

The most recent LIWC release (Pennebaker et al., 2015) includes 90 categories and 6,400 words and word stems. In addition to word count (WC), there are four linguistic summary variables (i.e., analytic thinking, clout, emotional tone, and authenticity), three descriptive groups (i.e., words per sentence [WPS], percent of words matched with the LIWC dictionary [dic], percent of words longer than six letters), and 21 primary word groups (e.g., common adverbs, negations). There are 41 groups related to psychological foundations (e.g., social process, cognitive processes), 6 categories pertaining to personal concerns (e.g., money, religion), 5 informal linguistic categories (e.g., swear words), and 12 different groups of punctuation.

Other than WC, WPS, and linguistic summary variables, all categories are scored as a percentage. When analyzing a text document, the LIWC places each word in one of the abovementioned categories. The software then calculates each word group's percentage score by dividing the total specific WC into the document's overall WC. However, this does not mean that a single word only belongs to one specific category. Any word may belong to multiple groups. For example, happiness belongs to "positive emotions" and "affective processes." The latter group captures various emotions, whereas the former focuses solely on positive emotion words.

The LIWC aims to eliminate human bias as much as possible, and the software produces the same results for the same piece of data in every round of analysis. However, the LIWC does not understand context. Words with sarcastic and ironic connotations can have multiple meanings. The LIWC does not evaluate the content to accurately capture the meaning. For example, "mad" may refer to a strange person, anger, or positive affection, depending on the context.

Instead of focusing on single word groups (e.g., positive emotions), scores may be evaluated in broad categories (e.g., affective processes). Moreover, researchers can use linguistic summary variable scores (i.e., analytical thinking, clout, authenticity, and emotional tone), depending on the purpose. Pennebaker et al. (2015) constructed these four linguistic summary variables with broad and narrow LIWC categories. Each linguistic summary variable has a standardized score ranging from 0 to 100.

Pennebaker et al. (2014) published an algorithm to calculate the analytical thinking score (Eq. 1). However, other related studies only discuss the relevant word categories for clout (Kacewicz et al., 2014), authenticity (Newman et al., 2003), and emotional tone (Cohn et al., 2004). These studies do not provide explicit algorithms. Equations (2), (3), and (4) provide insights, but do not present the actual calculations.

$$Analytical\ thinking = 30 + article + preposition - personal\ pronoun - impersonal\ pronoun - auxiliary\ verb - conjunction - adverb - negation \quad (1)$$

$$Clout = first\ person\ plural\ pronouns + second\ person\ pronouns - first\ person\ singular\ pronouns \quad (2)$$

$$Authenticity = first\ person\ singular\ pronouns + third\ person\ pronouns + exclusive\ words - motion\ verbs - negative\ emotion\ words \quad (3)$$

$$Emotional\ tone = positive\ emotion\ words - negative\ emotion\ words \quad (4)$$

Over time, as the LIWC has reached wider scientific audiences, further applications have become possible. In a political science study, Jones (2016) develops linguistic style measures to identify Hillary Clinton's use of the feminine and masculine styles during her career. These composite indices have elements from the stereotypical characteristics of the male language (Eq. 5) and the female language (Eq. 6).

$$Masculine\ linguistic\ style = first\ person\ plural\ pronouns + articles + prepositions + anger\ words + words\ longer\ than\ six\ letters + swear\ words \quad (5)$$

$$\begin{aligned} &Feminine\ linguistic\ style \\ &= pronouns + first\ person\ singular\ pronouns + verbs \\ &+ auxiliary\ verbs + social\ references + emotion\ words \\ &+ cognitive\ processes \\ &+ tentative\ words \qquad (6) \end{aligned}$$

The second essay employs modified versions of these linguistic measures (Eqs. 5, 6). Unlike Jones (2016), the second essay does not employ pronouns as part of the communal linguistic style index. Since there are already first-person singular pronouns (a subcategory of pronouns), including pronouns would have caused double-counting. In addition, the second essay did not take the weighted average of each word category. The LIWC readily calculates scores for each category using the WC ratio.

7.2.2.1 T – test and effect size

As a standard procedure in sociolinguistic studies (e.g., Newman et al., 2008), the second and third essays compare the mean scores from the LIWC categories and calculate the effect sizes when there is a significant difference.

Both essays employ an independent sample t-test to compare the linguistic categories of male and female executives. Equation (7) provides the calculation. Here, the formula employs the mean scores and standard deviations that belong to the same word group (e.g., positive emotion words). $\bar{X}_{male}$ represents the male sample's mean score, and $\bar{X}_{female}$ represents the female sample's mean score. μ denotes the expected mean score. The denominator denotes the calculation of the standard error of the difference between the male and female mean scores.

$$t = \frac{(\bar{X}_{male} - \bar{X}_{female}) - (\mu_{male} - \mu_{female})}{\sqrt{SD_{pooled}(\frac{1}{N_{male}} + \frac{1}{N_{female}})}} \tag{7}$$

Equation (8) below illustrates Cohen's d. The effect size was calculated using a standard measure. SD_{pooled} is the weighted average of the standard deviation of the two groups. N denotes each group's sample size. SD^2 represents the variance. The denominator of SD_{pooled} is the adjusted population for degrees of freedom.

$$d = \frac{\bar{X}_{male} - \bar{X}_{female}}{SD_{pooled} = \sqrt{((N_{male} - 1)SD^2_{male} + (N_{femal} - 1)SD^2_{female})/(N_{male} + N_{female} - 2)}} \tag{8}$$

Cohen (1992) suggests that $d = 0.2$ corresponds to a small effect size, $d = 0.5$ means a moderate effect size, and $d = 0.8$ indicates a large effect size. After comparing the means via a t-test, it is essential to calculate the effect size. A 5% significance may not be enough to explain why researchers reject a hypothesis. Numerically, a P-value very close to .05 indicates a significant difference between men and women. In contrast, a P-value slightly above .05 indicates that there is no significant difference. In the case of a significant difference with a small P-value, the effect size shows whether the difference is trivial. A small effect size implies that men and women use a specific word group comparably. A large effect size implies that there is a considerable difference between men and women in terms of their usage of a particular word category (Field, 2017).

Subsection 9.2.3 below introduces feature selection. In the second essay, feature selection is employed to explore the content as a whole. The third essay utilizes feature selection for certain portions of the data. By adopting parts of Argamon et al.'s (2005) procedure, the essay ranks analytic thinking, authenticity, and clout in descending order. It takes the top and bottom thirds of the population. Then, the third essay explores the content with the highest and lowest scores for men and women separately. This step helps to distinguish between the low and high ends of a dimension.

9.2.3. Feature selection

This subsection and the following subsections discuss applications implemented using the open-source software Python. The information presented in these subsections comes from essential books concerning text analysis (Müller & Guido, 2016), data collection (Mitchell, 2018), and machine learning applications (VanderPlas, 2016), as well as statistical learning (Hastie et al., 2009; James et al., 2013).

The standard procedure for text data analysis starts with cleaning the data. This process includes eliminating raw data features (e.g., HTML tags, advertisements), removing punctuation, expanding contractions, and changing capital letters to lower case letters. There are multiple open-source software programs (e.g., Python, R) that can be used to implement these changes. The researcher employed Python 3.6.5, using the sci-kit learn library (Pedregosa et al., 2011).

The second step is to eliminate rare or commonplace features (i.e., words and word groups). For example, researchers may consider selecting features that occur in more than 20% and less than 80% of a given text. Removing stopwords (e.g., function words with three letters or fewer, e.g., the, a, I) is also a common step for researchers who are solely exploring content. However, function words may be necessary when investigating linguistic styles.

The third step is to use a stemmer or lemmatizer to avoid feature recounting. As an illustration, “played,” “playing,” and “plays” refer to the same word, “play.” Instead of counting three words, these algorithms only count one. Gentzkow et al. (2019) have suggested that Porter Stemmer (Porter, 1980) is the standard stemming tool for NLP studies. Stemmer trims common suffixes. An incorrect form or incorrect spelling may occur with stemming. For example, Stemmer replaces “was” with "wa" and "worse" with "wors." Unlike stemming, lemmatization (e.g., WordNetLemmatizer; Fellbaum, 1998) returns the dictionary version of a root form. Continuing the above example, "was" returns as "be" and worse returns as "bad." Lemmatization ensures that the word belongs to the English language. A lemmatizer distinguishes a noun from a verb, as in the example of "play."

Although both serve the same function, the stemming algorithm works faster. Lemmatization operates with a corpus (e.g., readily available text data including word stems), is more time-consuming, and may require part-of-speech tagging (i.e., marking words with grammatical labels).

Researchers can implement either one or choose to use both. When the meaning or dictionary form is more important, a lemmatizer is the more appropriate choice. It is also possible to use both processes consecutively, as in the second essay, which first employs a stemmer, followed by a lemmatizer. This procedure eliminates meaningless words from the stemmer and reduces the number of features.

After this in-depth preprocessing, researchers can implement the "bag of words" (BoW) procedure. Simply put, this system resembles throwing the words comprising a text in a bag and counting them one by one. This procedure does not consider word order, synonyms, or grammar structures. In computational linguistics, a computer algorithm performs the process, for example, CountVectorizer in the sci-kit-learn library in Python 3.6.5.

The steps below illustrate the overall feature extraction procedure to provide a general understanding. Sample sentences uttered by Speakers 1 and 2 represent the text data employed in the procedure.

> Speaker 1: *"The female corporate executives are more benevolent than male corporate executives."*
>
> Speaker 2: *"These findings come from management research, but more studies might be necessary."*

After the preprocessing algorithm removes all stop words and punctuation, all letters are changed to lower case, followed by stemming and lemmatization.[5] The text data include ten unique words: *female, corporate, executive, benevolent, male, finding, come, management, research,* and *study.* Table 3 presents the WCs for the overall text and each speaker. The vector presentation of the numbers in the table is as follows.

Speaker 1: [1, 2, 2, 1, 1, 0, 0, 0, 0, 0]

Speaker 2: [0, 0, 0, 0, 0, 1, 1, 1, 1, 1]

Overall text		Speaker 1	Speaker 2
Word	Frequency	Frequency	Frequency
female	1	1	0
corporate	2	2	0
executive	2	2	0
benevolent	1	1	0
male	1	1	0
finding	1	0	1
come	1	0	1
management	1	0	1
research	1	0	1
study	1	0	1

Table 5 Feature counts

CountVectorizer produces a 2 x 10 matrix.

$$\begin{bmatrix} 1 & 2 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

[5] This example is a straightforward approach. A computerized stemmer and lemmatizer produce more accurate results than represented by the researcher's knowledge about word stems and lemmas. For a list of stopwords, see https://www.ranks.nl/stopwords

Table 6 illustrates the computer output organized on an Excel spreadsheet augmented with metadata. The columns after the speaker refer to sociodemographic characteristics (e.g., gender and age). After these features, the token columns also resemble a matrix representation of all the text data.

Speaker	Gender	Age	....	female	corporate	executive	benevolent	male	finding	come	management	research	study
1	F	40	....	1	2	2	1	1	1	1	1	1	1
2	M	45	....	0	0	0	0	0	0	0	0	0	0

Table 6 Computer output of feature counts

After preprocessing and extracting the features, researchers can analyze the text data further using the applications discussed in the following two subsections.

7.3 Data modeling versus algorithmic modeling

Before delving further into the details of data analyses, it is essential to discuss the differences between statistical applications (e.g., data) and machine learning (e.g., algorithm) applications. Machine learning may seem like fancy statistics to the naked eye. However, this type of approach is an oversimplification. The distinguished statistician Breiman (2001a) discusses the differences between the two approaches in "Statistical Modeling: The Two Cultures." This section is mainly based on that paper; it provides the fundamental background for understanding the methodology of the second essay.

As subsets of data science, data modeling and algorithmic modeling differ in terms of how researchers approach data. In data modeling, statisticians decide which model to use to explore a dataset. This model then draws conclusions about the variables. As Breiman (2001a) illustrates, the black box produces the data. From one side, a researcher pours a vector of input variables x. The other side produces the outcome, that is, the response variable y. The black box is the nature of the various functions that researchers use to explore the relationship between dependent and independent variables. In this process, the first goal is to predict future input variables. The second goal is to gain information about the relationship between response and predictor variables.

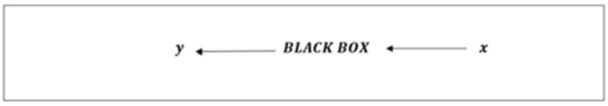


Figure 3 Black box

Data modeling is the mainstream approach used in statistics. In other words, a stochastic data model produces relevant information. This process results from random draws of different parameters, as stated in Equation (9). Researchers estimate the values of these variables from the data. Then, from nature, the model makes predictions about the relationship between x and y and produces information. A model validation technique includes goodness-of-fit measures (e.g., R^2).

$$y = f\ (predictor\ variables, parameters, random\ noise) \tag{9}$$

The primary purpose of statistics is to obtain information about the essential mechanisms in the data. Data modeling provides simple, easily interpreted information about the association between the input and response variables. The main disadvantage is that researchers assume that their choice of model fits the data. The conclusions drawn from this approach may not generate a picture of the data itself. Instead, the outcome may explain the model's operating mechanism on the data.

Another problem may arise from these data. Researchers may need to determine whether the data at hand can answer the research question and, similarly, whether a data model would help with understanding the data. Finally, the analysis may not provide a clear picture when the primary focus is on the model and not the data itself. In addition, testing hypotheses on the entire dataset may not produce reliable results.

Data models strive for accuracy by utilizing the entire sample. When researchers pour x into nature, they obtain y. In the same way, they also obtain $\hat{y}$ from x. Prediction accuracy is approximately equal to y and $\hat{y}$. This approach can be misleading because researchers do not test the accuracy of unseen data. First, the model parameters are calculated using the data. Then, using the model, researchers make predictions about the data. This loop, which recruits the entire dataset, indicates prediction quality.

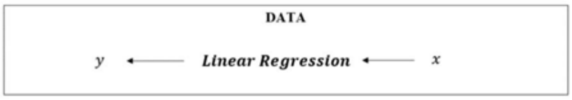


Figure 4 Data models

Algorithmic modeling has not evolved primarily within statistics because it is a computer-intensive process. Big data, as well as more compact datasets, can utilize algorithmic modeling. As an alternative to data modeling, algorithmic modeling appears to produce more accurate information. This argument does not mean that one methodology is superior to the other. Given that algorithmic modeling employs various accuracy measures, researchers in the field highlight that the results are likely to be more reliable.

Here, the data have an unknown mechanism. Moreover, the nature of the interior is entirely unknown. Researchers aim to find the best algorithm $f(x)$ for x in a test set such that $f(x)$ becomes a good predictor of y. In other words, machine learning algorithms learn data patterns to produce an accurate model. Subsequently, the researchers validate the model with predictive accuracy. Employing training, test sets, and cross-validation (CV) (Stone, 1974) are approaches to reduce bias in predictive accuracy (discussed in detail in Section 9.4).

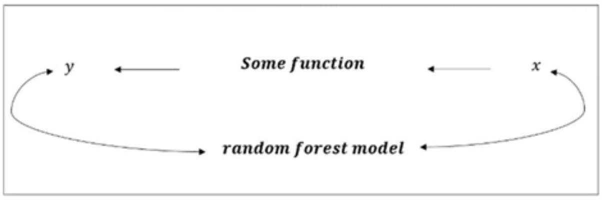


Figure 5 Algorithmic models

It is important to note that unlike statisticians and data scientists, social scientists are likely interested in the features of an algorithm with acceptable accuracy. The supplementary content analysis in the second essay operates with this motivation. We are specifically interested in the parameters the algorithm selects. This allows us to go beyond the surface and examine the details to which feature extraction may not attend.

Before introducing the details of the models in the second essay, Subsection 9.3.1 below discusses a fundamental topic in model fitting: the bias–variance tradeoff.

7.3.1 The bias – variance tradeoff

Equation (10) provides an alternative way to write Equation (9). In this setting, Y is a continuous variable, and there are p input variables, $X = (X_1,\ X_2, \dots, X_p)$. f is a fixed but unknown function that describes the relationship between Y and X. Here, the assumption is that there is some relationship between Y and X. $\in$ is the "random noise" or independent error term with a zero mean.

$$Y = f(X) + \in \qquad (10)$$

The main aim here is to calculate f. Hence, researchers can obtain Y from readily available X. The model functions at optimum capacity with low bias and low variance. Equation (11) shows the calculated $\hat{f}$ and predicted $\hat{Y}$. Given that the error term approximates to zero, $\in$ may not appear explicitly in the equation.

$$\hat{Y} = \hat{f}(X) \tag{11}$$

As discussed, $\hat{f}$ is a $black\ box$. In this setting, the main concern is to accurately predict Y, assuming that the black box can produce accurate output. Hence, the accuracy of Y depends on the accuracy of $\hat{Y}$. In turn, the prediction quality of $\hat{Y}$ is not independent of $\hat{f}$ and $\in$. Firstly, it may not be possible to find a perfect model $\hat{f}$. However, researchers can eliminate inaccuracies arising from the estimated model. Secondly, there most likely to be some independent error, $\in$, that researchers cannot control. The former case describes $reducible\ error$, while the latter refers to $irreducible\ error$.

Irreducible error refers to the unmeasured variables that may help predict Y. In the second essay, this might be the executives' personalities. Personal characteristics may still play a role in linguistic style, but we do not employ them. Some other aspects that we almost certainly cannot distinguish in the text data may also increase variation. For example, feelings on the specific day of an interview or interaction with the interviewer may impact linguistic style.

Equation (12) denotes the $expected\ value$ or the average of the squared difference between the predicted and actual values of Y; $E\left(Y - \hat{Y}\right)$ by assuming that both $\hat{f}$ and X are fixed. The $irreducible$ portion of the equation denotes the variance of the error term $\in$.

$$\begin{aligned} E\left(Y - \hat{Y}\right)^2 &= E\left[f(X) + \in - \hat{f}(X)\right]^2 \qquad (12) \\ &= \underbrace{\left[f(X) - \hat{f}(X)\right]^2}_{Reducible} + \underbrace{Var(\in)}_{Irreducible}, \end{aligned}$$

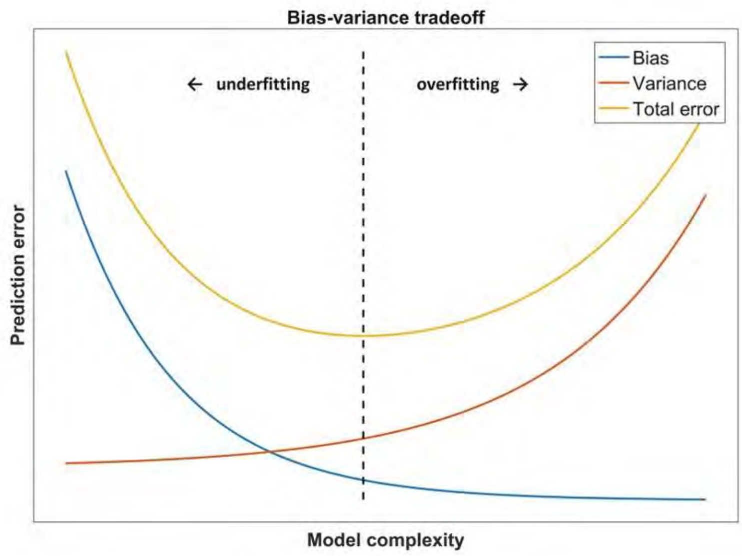
Variance and bias (Eq. 13) generate the $reducible$ part of Equation (12).[6] To minimize error, the model needs to have low bias and low variance.

$$E(Y - \hat{Y})^2 = Var\left(\hat{f}(X)\right) + \left[Bias(\hat{f}(X))\right]^2 + Var\,(\epsilon) \qquad (13)$$

When calculating $\hat{f}$, the main aim is to have a model that closely approximates the contents of the black box, f. In simple linear models (e.g., Eq. 14), a model may not be close to the unknown f. Flexible models (e.g., random forest) can mitigate this problem, especially when there are many variables. However, this approach is costly. Complex models can overfit the data by vigorously memorizing each detail. Hence, a slight change in the data leads to a significant change in $\hat{f}$. In other words, flexible models, also known as high-variance models, tend to follow random noise too closely.

The advantage of flexible models is that they have less bias than simple linear regressions. By virtue of its many features, a flexible model can make more accurate predictions. However, introducing more features can only improve the model's accuracy to a limited extent. Initially, as shown in the illustrative graph (6), this may reduce bias, but model flexibility can inflate variance after a certain point. Conversely, decreasing variance increases bias, and in this case, the model fails to capture relevant details (i.e., underfitting).

[6] Please see the Appendix for proof.



Source: Dankers et al. (2018)

Figure 6 The bias – variance tradeoff

The following subsection illustrates the data model. We calculated the linear model parameters from the data and made predictions about the data. This approach paints a general picture of linguistic styles and gender.

7.3.2 Simple linear regression

The second essay employs a simple linear regression to determine whether there is a meaningful relationship between linguistic style and an executive's gender. In Equation (14), Y_i represents either one of the linguistic style scores as dependent variables (i.e., the agentic or communal linguistic style). β is the difference between the agentic and communal linguistic styles, while controlling the analysis for various characteristics (e.g., the executive's age, the organization's publicly traded status). The vector $\mathbf{Z}_i$ represents various individual and organizational characteristics that serve as control variables.

$$Y_i = \alpha + \beta Female_i + \gamma' \mathbf{Z}_\mathrm{i} + \epsilon_i \qquad (14)$$

Here, we assume that the model fits the data because, in simple terms, the linguistic style (i.e., a continuous variable) may have a positive or negative relationship with being a woman (i.e., a binary variable). Consequently, we employ linguistic style scores rather than a set of linguistic features. This step helps to simplify the complex linguistic data by reducing it to a two-digit number from a set of words and word groups.

Although this does not form part of the analysis in the second essay, we made sure that the data structure did not alter our results. Table 7 shows that our results remained essentially unchanged.

Variable	Model 1 Agentic Linguistic Style	Model 2 Communal Linguistic Style
Female	−1.13**	1.77**
	(0.00)	(0.00)
MBA degree of an executive	0.59*	−0.92†
	(0.05)	(0.07)
Native speaker status of the executive	-0.57	1.04†
	(0.11)	(0.06)
Publicly traded status of the company	0.19	-0.406
	(0.57)	(0.47)
Age of an executive	-0.01	0.03
	(0.45)	(0.15)
Caption	−6.64**	7.98**
	(0.00)	(0.00)
Podcast	−9.30**	13.19**
	(0.00)	(0.00)
Interviews	−2.57**	4.80**
	(0.00)	(0.00)
Other	−4.05**	3.94†
	(0.00)	(0.08)
The New York Times	−10.61**	19.68**
	(0.00)	(0.00)
Constant	48.32**	50.92**
	(0.00)	(0.00)
Observations	1,082	1,082
R-squared	0.42	0.50

Note: Robust standard errors in parentheses.
†$p < 0.10$
*$p < 0.05$
**$p < 0.01$

Table 7 Linguistic style differences among the executives

Algorithmic modeling eliminates the need for simplification by producing the best possible model to explore the data—and by extension, a realistic picture of how male and female executives speak. As part of the second essay's supplementary analysis, the next section discusses nonlinear models. The second essay employs this approach to explore spoken content's contribution to linguistic style.

7.3.3 Supervised nonlinear classification

The second essay employs a supervised classification algorithm. In the dataset, for each input variable in $x_i, i = 1, \dots, n$, there is a related response variable, y_i. A supervised learning algorithm fits a model that accurately predicts the future output variables. In algorithmic modeling, which is different from a data model, researchers build a classifier from a training set $(x_1, y_1), \dots, (x_n, y_n)$. The main aim is to have a classifier that performs well on the training set and then on unseen data, that is, the test set. The split of the training and test sets can be 80% and 20% of the entire dataset, respectively.

Generally, the response variable y is continuous or quantitative (e.g., the linguistic style score). However, there are some problems associated with a categorical or qualitative y (e.g., gender). Classification deals with categorical response variables. An algorithm predicts, or in this case, classifies an observation under a label (e.g., male or female). Classification algorithms resemble regression models because the model first calculates the probability of a variable belonging to one category. Then, it predicts whether the speaker is male or female. The most well-known application of classification is the logistic regression.

However, the second essay utilizes a computer application. Random forest is an algorithm that operates using many decision trees. The first application of decision trees dates back to the mid-1980s. Initially, computer scientists, engineers, and physicists were the primary users of decision trees. Researchers employ these advanced models in image, speech, and handwriting recognition, as well as nonlinear time series, and financial market predictions (Breiman 2001a).

When using these algorithms, the main aim is to achieve high predictive accuracy. However, this goal is associated with several complications. Accordingly, Breiman (2001a) identified three main problems with algorithmic modeling. The first is picking the best model among many good models. As an illustration, researchers aim to identify the five best covariates among 30 variables to produce the best linear regression. Researchers choose a model with a minimal mean squared error (MSE) or the lowest test error when such a problem is the case. However, this criterion may not point directly to the best model. There are almost certainly many models with minimal errors. Each model with a different set of variables draws a different picture of the data, so it is unclear which one would be the best choice. Random forests are likely to eliminate this problem by employing many predictors and returning a majority vote.

The second problem is accuracy versus interpretability. Linear regression models have good interpretability. They are simple and easy to understand when explaining the relationship between x and y. However, accuracy can be problematic. Similarly, decision trees are easy to understand, but models may struggle to achieve high accuracy. As opposed to growing a single tree, growing a forest may mitigate these problems. For example, in a classification, when providing a new set of x for each tree, obtaining a vote from each tree regarding the class may yield a better prediction. Bagging (Breiman, 1996) is a famous example of growing a forest. The main idea here is to introduce each successive decision tree using a random element. For example, from among many predictor variables, a tree chooses variables at random to split a node. Alternatively, the tree can choose a random combination of variables (Ho, 1998).

As an ensemble method, Breiman (1996) built random forests as a bagging algorithm to eliminate accuracy problems as much as possible. Unlike decision trees, forests appear to play a significant role in making accurate predictions. However, the random forest algorithm is similar to a black box, as it may be challenging to understand the mechanism behind the simultaneously growing decision trees. Thus, interpretability is problematic. Hence, there is a tradeoff between the two: Better accuracy means complex algorithms, while simple models, though easy to understand, may not have the best performance. Firstly, it might be helpful to focus on predictive accuracy. Then, exploring why the algorithm picked some variables over others may provide a better road map.

Finally, high dimensionality can be a problem. For example, the text data from the second essay include 522 observations and more than 700 weakly or moderately correlated features (i.e., words and word groups)—the so-called $p \gg N$ problem. Unlike simple linear models, penalized linear models (e.g., Lasso, Ridge, and Elastic Net) are suitable for high dimensionality. Regularization (shrinkage) methods minimize variance by reducing the value of some coefficients to (or close to) zero. The drawback is that penalized models are a better choice when the size of the feature set is comparable to the sample size, $p \approx N$. Moreover, penalized models can shrink the features from smaller subsets when the data are not well balanced.

7.3.3.1 Decision trees

Unlike the line in a simple linear regression, tree-based algorithms draw boxes in the XY plane. In other words, the algorithm divides a plane into layers. These spaces comprise several prediction regions. The rules regarding how to divide each space resemble the yes/no questions that a decision tree asks when splitting a node. Researchers make predictions based on each region's mean or mode in regression models.

However, in classification models, a decision tree is interested in whether an observation belongs to the commonly predicted class of the training set within a region. When interpreting results from decision trees, the number of observations from each training class is important. In other words, researchers likely need to know whether a particular group dominates a region.

Researchers need to set the criteria for pruning a tree (i.e., eliminating non-critical sections of a decision tree) and splitting a node. The Gini index is preferable to the classification error rate when pruning a tree.

However, the classification error rate is a better indicator of a pruned tree's prediction accuracy.

When growing a classification tree, the classification error rate indicates the ratio of misclassified observations (i.e., an observation that does not belong to the commonly predicted class) to the entire training population in a region. Equation (15) represents the classification error rate, where $\hat{p}_{kj}$ is the ratio of the training set observation from the j^{th} class in the k^{th} region.

$$E = 1 - \max_{j}(\hat{p}_{kj}) \tag{15}$$

The classification error rate may not be sufficient to grow decision trees. The Gini index indicates the probability of incorrect classification based on a random sample of a given node. Thus, the Gini index indicates node purity. If a node predominantly contains one class, the Gini index has a smaller value.

In other words, Gini is a measure of variance over J classes, where $j \in \{1, 2, \dots, J\}$. Equation (16) provides the calculation:

$$G = 1 - \sum_{j=1}^{J} (\hat{p}_{kj})^2 \tag{16}$$

Figure 3[7] presents an illustrative decision tree. This example determines whether a person is male or female by examining LIWC features. The tree has a depth of two. In the absence of depth, the nodes would split until a node contained a single sample. To split the node, the algorithm asks whether the male text has a female words score less than or equal to .55. "True" leads to a function node in which the male text function words score is less than or equal to 60.64. "False" leads to insight words

[7] This output is one of the random prints from the researcher's experimentation with the model. The researcher obtained this output by using LIWC features only with the collected data for this dissertation.

in which females supposedly score less than or equal to 3.24. Following the "true" path further, function words are split into two nodes. The final node of the outer left split shows that according to the text data from 196 people (162 males and 34 females), function words appear more in male text samples, with 82% correct classification (a Gini of 1).

As an illustration, based on 439 samples, there is a 47% chance that the root node (top node) misclassifies men and women. Since this is a binary classification task, the algorithm takes the sum of the predictions for men ($\frac{272}{439}$) and women ($\frac{167}{439}$) in a node, and takes the squares of each ratio (.38, .14), then subtracts the sum (.52) from 1 (estimation with a calculator gives .48).

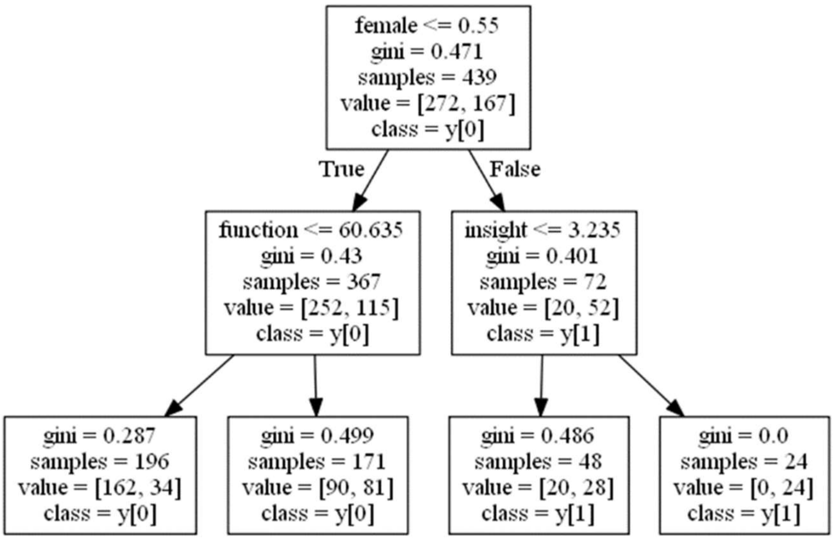


Figure 7 Decision tree classification

7.3.3.2 Random forest classification

Random forests are an ensemble of many decision trees. For a single decision tree n, the model generates a random vector of θ_n. This vector is independent of the previous vectors, $\theta_1, \theta_2, \dots\ \theta_{n-1}$, with an identical distribution. Each tree grows on a random vector θ_n and a training set, producing a classifier $h(\boldsymbol{x}, \theta_k)$, where $\boldsymbol{x}$ denotes an input vector. The algorithm repeats this process to produce N trees. Subsequently, the ensemble votes for the most popular class at the input vector $\boldsymbol{x}$. This procedure defines the random forest classification (Breiman, 2001b).

In a classification problem, the algorithm fits a model to the training data $Z = \{(x_1, y_1), (x_2, y_n), \dots, (x_N, y_N)\}$, making a prediction $\tilde{f}(x)$ at the input vector $\boldsymbol{x}$. The algorithm takes the average of predictions from many decision trees in a forest. For each bootstrap sample Z^{*b}, where $b = \{1, 2, \dots . B\}$, model fitting produces a tree, $\tilde{f}^{*b}(x)$.

The algorithm randomly selects m variables from the p variables to grow a random forest tree $\tilde{f}^{*b}(x)$. Among m variables, the algorithm decides on the best splitting point to further grow the tree. The algorithm splits the nodes into two branches until it reaches the terminal node, n_{min}. Each tree in a forest repeats this process, and the result is an ensemble of trees $\left\{\tilde{f}^{*b}(x)\right\}_1^B$.

Equation (17) demonstrates the random forest classification algorithm $\widehat{C_b}\ (a)$, making a new prediction of the b^{th} tree at a new input a by taking the majority vote from B number of trees.

$$\hat{C}_{rf}^{B}(a) = majority\ vote\ \left\{\widehat{C_b}\ (a)\right\}_1^B \qquad (17)$$

7.4 Accuracy in algorithmic models

In general terms, model performance or accuracy indicates whether an algorithm makes predictions that are close to the actual value. In classification models, researchers wish to estimate f by employing some training observations $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where $y_1, \dots, y_n$ represent classes and, hence, qualitative variables.

For classification problems, the mainstream approach to calculate the accuracy of $\hat{f}$ is to quantify the training error rate. This is the ratio of the misclassifications to the training data. For i^{th}observation, $\hat{y_i}$ denotes the predicted class label.

$$\frac{1}{n}\sum_{i=1}^{n} I\ (y_i\ \neq \hat{y}_i) \tag{18}$$

$I\ (y_i \neq\ \hat{y_i})$ represents a dummy variable, in which,

$$I\begin{cases} 1 & if\ \ y_i \neq \hat{y}_i \\ 0 & if\ y_i = \hat{y_i} \end{cases}$$

The algorithm correctly classifies the i^{th} observation if $I\ (y_i \neq \hat{y}_i) = 0$ and misclassifies otherwise.

The test error rate provides the ratio of misclassified observations in the test set. For observations (x_0, y_0), Equation (19) gives the test error rate after applying a classifier to the test set. $\hat{y}_0$ denotes the predicted class. An accurate classifier produces a small test error.

$$Ave\big(I(y_0 \neq \hat{y}_0)\big) \tag{19}$$

7.4.1 Ten – fold cross – validation

CV is a resampling method. The procedure's primary mechanism is to draw samples from a training set and fit the chosen model. In each repetition, refitting provides information about the fitted model by exploring the extent to which the results differ. For classification problems, the number of misclassified observations represents the test error.

Equation (15) denotes n-fold CV, where $n \in \{1, 2, \ldots. n\}$. Here, the performance metric could be the test set error, as in Equation (19), or the receiver operating characteristic under the curve (ROC AUC) score, as explained in Section 9.4.2.

$$CV_{(n)} = some\ performance\ metric \tag{20}$$

Bradley (1997) suggests that calculating the ROC AUC with tenfold CV is a reliable metric to assess accuracy. In the second essay's supplementary analysis, each round of tenfold CV operates on a randomly selected 90% training set, and the remaining 10% is employed as a test set. In every CV process, the ROC AUC determines model performance. The average of the ten rounds of testing produces the model's accuracy score.

7.4.2 Receiver operating characteristic under the curve

This section primarily discusses how researchers derive the ROC AUC and its purpose as a performance measure. Theoretically, the ROC AUC borrows basic features from the Neyman-Pearson lemma. The receiver operating characteristic (ROC) measurement originates from radar signal detection theory, which dates back to World War II. Regarding quality and forecast assessment, psychology, medicine, and meteorology are some fields that widely employ ROC as a performance metric (Mason & Graham, 2002).

A ROC curve shows the classification model's performance at different thresholds. However, this evaluation does not reveal much about the model's performance. The estimation of the area under the curve (AUC) measures model performance. The AUC is equivalent to the probability that a model correctly ranks a randomly selected positive instance higher than a randomly selected negative instance (i.e., the Mann-Whitney U test) (Bradley, 1997).

The AUC score takes a value between 0 and 1 because it is the fraction of the area of the unit square. In binary classification, as the AUC score approaches 1, the model becomes better at distinguishing between the two groups. When the AUC is equal to or less than .5, the model fails to differentiate and may misclassify. The ROC graph serves as a better performance measure than standard accuracy tests. This is mainly when data include unbalanced classes, as in the second essay where the sample is comprised of $N = 318$ men and $N = 204$ women (Fawcett, 2006).

The core ROC AUC elements come from metrics based on a model's classification performance. As an illustration (Fawcett, 2006), a classifier maps an instance K (e.g., a set of feature vectors) to a set of class labels $\{p, n\}$ (i.e., positive and negative). In this case, $\{Y, N\}$ denote the predicted class labels. Since this is a binary classification, there are four different possibilities. In Figure 8, a "true positive" shows that the input is positive, and the algorithm classifies it as positive. However, if the algorithm classifies this input as negative, it becomes a “false negative.” A “true negative” shows that the input is negative, and the algorithm classifies it as negative. A “false-positive” is when the algorithm classifies this input as positive.

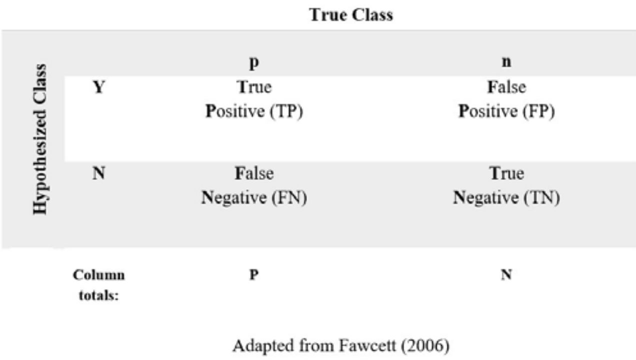


Figure 8 Confusion matrix

The confusion matrix (Figure 8) presents these outcomes. The following performance metrics were calculated based on the confusion matrix. Equation (21) shows the estimation of the true positive rate, which is also defined as recall and sensitivity:

$$True\ positive\ rate\ \approx \frac{Positives\ correctly\ classified\ (TP)}{Total\ positives\ (P)} \tag{21}$$

Equation (22) shows the calculation for the false positive rate:

$$False\ positive\ rate\ \approx \frac{Negatives\ incorrectly\ classified\ (FP)}{Total\ negatives\ (N)} \tag{22}$$

The F-score (23) is the harmonic mean of recall and precision. Compared to taking a simple average, the harmonic mean punishes extreme values. For example, a precision of .73 and a recall of .97 have an average of .85 but an F-score of .83. The F-score is a metric used to determine whether the model correctly identifies false positives and false negatives. A value of 1 is a perfect score.

$$F - score = \frac{2.Recall.Precision}{Recall + Precision} \quad (23)$$

Sensitivity (recall, 21), specificity (24), and precision (positive predictive value, 25) are also relevant for the ROC curve.

$$Specificity = \frac{True\ negatives}{False\ positives + True\ negatives} \quad (24)$$

$$Precision = \frac{True\ positives}{True\ positives + False\ Positives} \quad (25)$$

Researchers plot the ROC curve on the XY plane. The false-positive rate (FPR, 22) is on the X-axis, and the true positive rate (TPR, 21) is on the Y-axis. The ROC graph shows the tradeoff between costs (FPR) and benefits (TPR). Each classification outcome in the confusion matrix has a point in the ROC space. The perfect classification point is in the upper left corner of the ROC space with coordinates (0, 1) (Figure 7). The prediction has 100% sensitivity and 100% specificity. In other words, there were no false negatives or false positives.

The diagonal line dividing the ROC space, $y = x$, represents a random guess. The classification points above the line are better than a random guess. Conversely, a classifier below the line is likely to fail to apply the correct information for classification. The point (0.9, 0.9) denotes that the classifier can obtain true positives 90% of the time. However, the cost is high. At this point, the FPR is 90%.

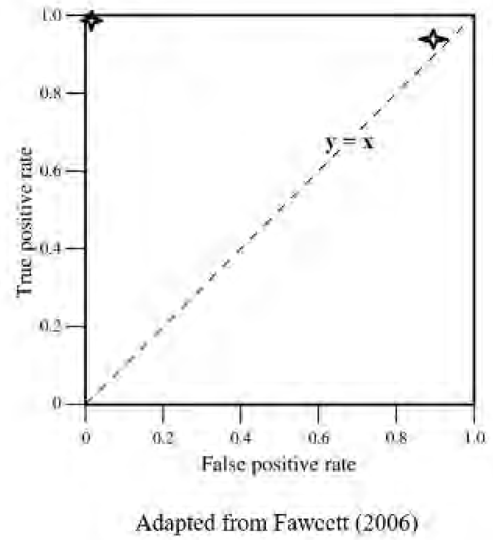


Figure 9 ROC space

The threshold is a limit for converting the probability to a binary value. As previously mentioned, a classifier first calculates the probability of an observation belonging to a particular class. For example, a model returning a .95 probability indicates that an email is probably spam. A probability of .002 indicates that the email is not spam. However, a value of .6 may not indicate a clear-cut decision. Thus, researchers may need to define decision thresholds. Curves in the ROC space help to identify an optimum threshold for a classification problem.

In a binary classification problem (e.g., Hernández-Orallo et al., 2013), $Y = \{1, 0\}$ denotes a $positve$ and a $negative$ class, respectively. x denotes an instance, and Y is the output space. A classifier is a function $m : X \rightarrow R$ that maps instances to scores. The classification model converts the scores on a decision threshold t to make predictions for the Y domain. For a predicted score $s = m(x)$, the instance x is in the positive class if $s > t$; otherwise, it is 0.

For an unspecified model and population, f_k denotes the score density for class k, and F_k represents the cumulative distribution function. Equation (26) shows the TPR or sensitivity and the ratio of correct classification for 1 at the decision threshold t.

$$F_1(t) = \int_t^{\infty} f_1(s)ds = P(s \leq t|1) \qquad (26)$$

Equation (27) represents the FPR at threshold t, that is, the ratio of incorrectly classified $class$ 0 as $class$ 1.

$$F_0(t) = \int_t^{\infty} f_0(s)ds = P(s \leq t|0) \qquad (27)$$

Bradley (1997) calculates the AUC by drawing insights from the Neyman-Pearson lemma. Trapezoidal integration is a viable approach when there are various thresholds and two known points $(\alpha, 1-\beta)$. α denotes a point on the $false\ positive\ rate$ $(F_0(t))$, and $1-\ \beta$ denotes a point on the $true\ positive\ rate$ $(F_1(t))$. Equation (28) approximates each interval for each ROC curve. Further, Equation (29) calculates the section between consecutive thresholds for true positives. Similarly, Equation (30) shows the section for false positives.

$$AUC = \sum_i \left((1-\beta_i \Delta_\alpha) + \frac{1}{2}(\Delta(1-\beta).\Delta\alpha) \right) \qquad (28)$$

Where,

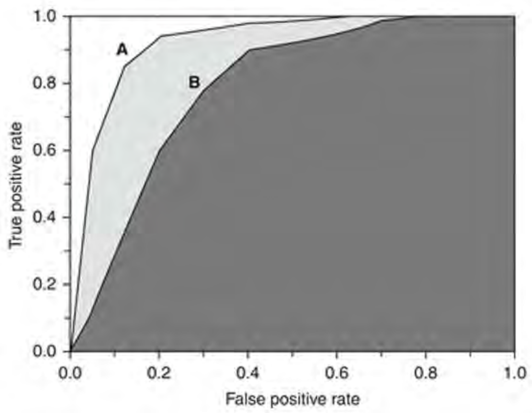
$$\Delta(1-\beta) = (1-\beta_i) - (1-\beta_{i-1}), \qquad (29)$$

$$\Delta\alpha = \alpha_i - \alpha_{i-1} \qquad (30)$$

Alternatively, the AUC of a classifier is equivalent to the Wilcoxon rank test (Bradley, 1997; Fawcett, 2006; Hanley & McNeil, 1982), which is also known as the Mann-Whitney *U* test (Mann & Whitney, 1947) or the Wilcoxon rank-sum test (Wilcoxon, 1945). These similar tests offer a nonparametric alternative to the independent t-test (Field, 2017). Equation (31) presents the formula for the Mann-Whitney *U* test for two groups with a sample size of n_1 and n_2, respectively. r_{1i} represents the rank of the i^{th} feature that belongs to the first group.

$$U = \sum_{i=1}^{n_1} r_{1i} - \frac{n_1(n_1 + 1)}{2} \tag{31}$$

In the example below (Figure 10), the AUC value of classifier A is larger than that of classifier B.



Source: Melo (2013)

Figure 10 Receiver operating characteristics under the curve

7.5 An example: Python output

This section presents the theoretical information introduced in the previous sections on a Python output. As discussed regarding the bias–variance tradeoff, a random forest model is a complex algorithm prone to overfitting. However, readily available packages in Python 3.6.5 (e.g., sci-kit learn; Pedregosa et al., 2011) calculate the optimum forest features.

In other words, hyperparameter tuning overcomes overfitting and improves model performance. Similar to a real-life tree, each decision tree has a root and numerous branches. Python packages allow researchers to determine the optimum limit for random forest features by calculating the number of branches on a tree and the number of trees in a forest.

The output below belongs to an algorithm with $1{,}200$ trees. Each tree has a maximum of 460 nodes. When splitting a node, a tree considers randomly selected features. The tree determines the number of random features to be considered by taking the square root of the total number of features in a node. Hyperparameter tuning produces all these characteristics to improve model accuracy.

The output below (Figure 11) belongs to a random forest algorithm with these algorithmic characteristics. The data have $522\ \left(N_{male} = 318, N_{female} = 204\right)$ observations and 747 features (i.e., words and word groups). The test set comprises 20% of the data ($N = 105$). In the confusion matrix, the left diagonal shows the number of correctly classified men and women.

In the classification report, support shows the number of samples in $class\ 0$ (i.e., $N_{male} = 63$) and $class\ 1$ (i.e., $N_{female} = 42$). The precision score shows that the model can correctly label men and women 73% and 90% of the time, respectively. Furthermore, recall shows a classifier's ability to identify positive instances. In this case, the classifier can identify men 97% of the time and women 45% of the time. The F1 score brings together precision and recall. Among all male-labeled people, 83% are actually men. Among all female-labeled people, 60% are women.

We determined the model's global accuracy using ROC AUC by running a tenfold CV. The output (Figure 11) shows the ROC AUC for each round. We are interested in the average of these ten scores. Therefore, the random forest model can sort men and women correctly with 86% accuracy. This score is acceptable because our primary interest is the model's input variable. As discussed in the second essay, the model employs features related to *women* and *family*. These details can be captured using the LIWC. However, the LIWC does not explicitly specify which words are part of the female- or family-related word category. Machine learning algorithms can fill this gap by providing seemingly miniscule but essential details.

```
=== Confusion Matrix ===
[[61  2]
 [23 19]]


=== Classification Report ===
              precision    recall  f1-score   support

           0       0.73      0.97      0.83        63
           1       0.90      0.45      0.60        42

=== All AUC Scores ===
[0.9203869  0.8608631  0.93005952 0.90550595 0.87109375 0.840625
 0.86171875 0.79375    0.825      0.83467742]


=== Mean AUC Score ===
Mean AUC Score - Random Forest:  0.8643680395545316
```

Figure 11 Python output

8 The US sample

This dissertation only employs samples from the United States, including experimental agents and the leader population. There are three main reasons for this. Firstly, the overall language is English, including tools and measures (e.g., personality inventories, LIWC). It was therefore reasonable to use native English speakers. Studies with an international sample tend to highlight language as a limitation. Murphy et al. (2021) report that nonnative speakers with a weak English-as-a-second-language background may struggle to comprehend the meaning of survey questions. When participants fail to understand the questions, the findings may not be genuine or reliable. In addition, the LIWC operates in English. Although there are various versions of the LIWC in different languages, the principal studies use US English samples. Therefore, it is easier to compare the results of these essays with those of other studies that also use US samples. Second, as Hambrick (2007) suggests, the US executive population has distinctly diverse demographics. A top executive sample from a daily US newspaper with an international readership (Wang, 2018) is representative of the global corporate world. Finally, focusing on the same country creates continuity among the three essays.

9 Summary

Essay 1: Gender, inequality, and risk taking

People might consider women to be more risk-averse than men in different settings (Croson & Gneezy, 2009; Charness & Gneezy, 2012). This view may be intact when people evaluate men's and women's risk-taking decisions (Daruvala, 2007; Eckel & Grossman, 2002; Siegrist et al., 2002). Indeed, in most cases, people accurately predict gender differences in risk attitudes. However, people fail to capture the actual differences between men and women.

This stereotypical view of gender differences in risk attitudes might lead to different evaluations based on a risk-taker's gender. The adverse outcomes of women's risky decisions may not be subject to a comparative evaluation as is the case with men's decisions. Findings from the literature show that female surgeons receive fewer referrals after a failed surgical procedure (Sarsons, 2017b). Furthermore, female financial advisors may have a higher chance of losing their jobs due to misconduct than their male counterparts (Egan et al., 2017).

In this essay, we specifically explore whether people evaluate the outcomes of men's and women's risky decisions differently using a novel survey experiment. We facilitated a hypothetical dictator game to explore the research question. This design involves hypothetical workers and spectators as the only experimental agents. Moreover, we explored spectators' motivations using an implicit association test on stereotypes.

In a short vignette, we described a decision-making situation involving a pair of workers. After completing a similar individual effort task, these workers can earn a specific sum in addition to their compensation by choosing a risky or safe earnings option. As the sole experimental agents, the spectators made redistribution decisions based on the hypothetical workers' earnings decisions. Between subjects, spectators evaluated either same-sex or mixed-sex pairs. In this setting, we randomly varied the gender of the better-off and worse-off workers in an otherwise identical pair. These variations resulted in 2 × 2 treatments. Specifically, this design gives us insights into whether spectators see better-off risk-taking men as more entitled to their gains than better-off risk-taking women or, similarly, whether the worse-off worker receives more compensation depending on their gender.

The results show that female spectators tend to make an equal distribution between a better-off and a worse-off worker, regardless of gender. However, male spectators appear to leave unlucky male workers with no earnings more frequently than they disenfranchise unlucky female workers. We further explore whether spectators' decisions are associated with implicit biases. Although spectators appear to favor stereotypical men, we find that these beliefs have no influence on redistribution decisions.

Manuscript: Fest, S., Yaldiz, N., & Kvaløy, O. (2021). Gender, inequality, and risk taking.

Essay 2: Verbal Impression Management Strategies by Top Female Executives

Creating a leader image involves specific challenges for women. Social role theory (Eagly, 1987; Eagly et al., 2000) highlights that these obstacles arise from the mismatch between female gender role and a stereotypical leader image. People expect women to be warm, cooperative, and supportive. In contrast to these communal features, the agentic features of the male gender role (e.g., assertive, aggressive, and competitive; Eagly & Carli, 2007) match a stereotypical leader image (Eagly, 1987; Eagly & Karau, 2002). Due to this incongruity between the communal features of being a woman and the agentic features that comprise a stereotypical leader image, women are likely to face an impression management dilemma (Rudman & Phelan, 2008). Studies show that female leaders experience backlash when they portray an agentic image. However, when women are more communal, they appear to be incompetent leaders.

However, the literature also suggests that displaying communion can benefit female executives. According to the role congruity theory, women can avoid backlash by confirming the gender prescription (Heilman & Okimoto, 2007; Rudman et al., 2012). Although this can risk female leaders appearing incompetent, the stereotype content model suggests that women at the top leadership levels can appear both competent and communal (Eckes, 2002; Fiske et al., 2002). Furthermore, communal characteristics can strengthen the image of female leaders by helping them appear as effective leaders and supporting women as a better match for transformational leadership style (e.g., Koenig et al., 2011).

The literature provides insights into how female leaders integrate agentic and communal characteristics in nonstereotypical domains (Amanatullah & Morris, 2010; Heilman & Okimoto, 2007; Williams & Tiedens, 2016). However, we do not know how they use verbal communication to

achieve this goal. This study focuses on top female executives and fills this knowledge gap. We analyze publicly available texts (e.g., interviews, speeches, podcasts, and editorial pieces) from 204 female and 318 male prominent executives with over one million words spoken or written by them.

This study mainly contributes to the interdisciplinary social psychology and management literature that focuses on women's strategies against backlash in non-stereotypical settings. We find that female executives use communal language (more personable and emotional styles) and communicate communal topics (e.g., family and motherhood). This finding is consistent with the current literature that women in non-stereotypical domains tend to soften their image with gender congruent features, and such communal display could benefit top female leaders. In addition, comparisons with the general public also support that male and female leaders have androgynous styles. Our findings also support the developing literature on the importance of leader androgyny and the shift in leader stereotypes.

Manuscript: Onozaka, Y., & Yaldiz, N. (2022). Verbal Impression Management Strategies by Top Female Executives

Essay 3: Personality Expression by Language among Business Executives

As the faces of the company, how business executives present themselves matters to both inside and outside of the organization (Schlenker, 2012). Leaders' public expression of images is shaped by what society expects of leaders, such as inspirational, charismatic, and reliable. These characteristics are linked to personality traits of extraversion and agreeableness. Accordingly, the images they choose to convey and how they communicate them are likely to reflect both the impression management strategies to fulfill the social expectations and their personalities.

As personalities are latent constructs, surveys are the commonly employed methods to assess one's personality (e.g., Goldberg, 1990). However, the recent advancement in computational linguistics provided new methodological outlets to explore personality from text data. Social science studies are already utilizing these new methodologies on student (e.g., Mehl et al., 2006; Sun & Vazire, 2019) and general public samples (e.g., Tackman et al., 2020).

Still, there is a further need to explore linguistic personality indicators through diverse sources and samples since language and personality traits are anchored to context (Stryker, 2007; Roberts, 2009). In this aspect, gender moderates how people express their personality and how these traits are perceived by their immediate environment. While social psychology and personality literature give insights into how the general public express themselves, the literature lacks insights into personality expression in the business leadership domain and how gender impacts these trait expressions. This study aims to fill this gap by exploring the personality expressions of business executives. Current findings show that the personality expressions of leaders align with social expectations.

Manuscript: Yaldiz, N. (2022). Personality Expression by Language among Business Executives

10 Discussion

The first essay explores whether people evaluate nonstereotypical women differently than their male counterparts. The findings partially overlap with previous literature that has employed registered data (e.g., Egan et al., 2017; Sarsons 2017b). In these studies, women are punished when they take risks and fail. However, the first essay finds that male spectators appear to leave unlucky male workers with no earnings more frequently than they disenfranchise unlucky female workers.

Furthermore, findings show that men are likely to compensate for unlucky women's losses. Stereotyping women as risk-averse and considering women to be more likely to fail at nonstereotypical tasks may explain men's behavior. Previous survey studies have reported comparable findings. People tend to overestimate men's risk-seeking behavior or consider women to be more risk-averse (e.g., Daruvala, 2007; Siegrist et al., 2002).

This essay differs from previous studies in that it employs a novel methodology. A further study that uses real stakes and workers within the same design may help legitimize this hypothetical design (i.e., assuming that results would overlap). Additionally, experimental studies may utilize linguistic samples more to collect clues about political views, personality, and other human characteristics.

The second essay explores the impression management strategies women utilize in a nonstereotypical context. Here, female senior executives may be considered nonstereotypical women because they occupy a position in a nonstereotypical domain. The findings mostly overlap with the previous psychology literature exploring women's strategies while engaging in a stereotypical male task. In this context, women ease backlash by emphasizing their communal side (e.g., Amanatullah & Morris, 2010; Heilman & Okimoto, 2007; Williams & Tiedens, 2016).

Unlike the substantial management literature, this essay employs big publicly available text data from diverse media outlets. It is virtually impossible to approach top US-based executives for interviews. Hence, publicly available text data serve as a unique alternative source for studying an actual executive sample. Furthermore, linguistic samples are likely to provide reliable insights into these women's strategies.

This essay utilizes a state-of-the-art methodology to investigate these strategies. The dictionary approach employing the LIWC is a widespread methodology in psychology studies and is expanding into management research. Moreover, supervised classification algorithms capture specific details to further interpret insights gained from linguistic data.

Future studies may focus on dissipating the ambiguity surrounding whether female leaders discuss feminine issues due to their personal willingness or due to the nature of the questions that are posed to them. Although the second essay aims to eliminate this issue via a supplementary analysis, additional research on women's maternal image may improve clarity. For example, Hillary Clinton's former Twitter biography started with "wife, mom, and grandma," and only introduced her professional qualifications thereafter. In contrast, her significant other did not mention any family affiliations (Grady, 2018). Future analysis can examine whether women mention their family on their professional webpage or social media account.

The final essay employs the second essay's data and explores female leaders' personality features. The final essay also provides insights into nonstereotypical women's characteristics. Findings partially overlap with the literature on female leaders' agreeableness (e.g., Adams & Funk, 2012). Nevertheless, further studies might be necessary to arrive at more substantial implications related to female leaders' characteristics. It is unclear whether female leaders start their career with these personality features or if they evolve with experience. Findings from applied psychology also show that the Big Five personality traits are dynamic; that is, they change with advancement in leadership positions (Li et al., 2020). As people ascend, they can enhance their (people) skills through formal training (e.g., MBA education). Life experiences along the way may also help people develop specific characteristics. All of these qualities, in turn, may impact a leader's personality, causing them to evolve into a different person (Avolio, 2005).

The third essay prepares the foundation for future studies, mainly focusing on the leadership literature. The findings from the third essay imply that female leaders may have characteristics that match the traits of a transformational leader. The only difference is that female leaders are not as open as male leaders. A future study may explore female leaders' transformational style through textual samples, primarily focusing on corporate policies. Furthermore, as machine learning applications have begun to find a place in management and leadership research (e.g., Doornenbal et al., 2021; George et al., 2016; Wenzel & Van Quaquebeke, 2018), future studies can utilize more extensive data and employ various machine learning models to explore transformational leadership at the intersection of gender and personality. Finally, the essay can integrate questions into the analysis to clarify the influence of societal expectations.

References

Adams, R. B. (2016). Women on boards: The superheroes of tomorrow? *The Leadership Quarterly, 27*(3), 371-386.

Adams, R. B., & Funk, P. (2012). Beyond the glass ceiling: Does gender matter?. *Management science*, *58*(2), 219-235.

Adams, R. B., & Ragunathan, V. (2017). *Lehman sisters.* SSRN. http://dx.doi.org/10.2139/ssrn.3046451

Ahern, K. R., & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *The quarterly journal of economics*, *127*(1), 137-197.

Ahmadian, S., Azarshahi, S., & Paulhus, D. L. (2017). Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences*, *107*, 49-53.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, *115*(3), 715-753.

Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of personality and social psychology*, *98*(2), 256.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, *58*(1), 78–93. https://doi.org/10.1287/mnsc.1110.1348

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005). Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America* (pp. 1–16).

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & talk*, *23*(3), 321-346.

Arrow, K. (1973). The Theory of Discrimination. In O. Ashenfelter, & A. Rees (Eds.), *Discrimination in labor markets* (pp. 3-33). Princeton University Press

Arvate, P. R., Galilea, G. W., & Todescat, I. (2018). The queen bee: A myth? The effect of top-level female leadership on subordinate females. *The Leadership Quarterly*, *29*(5), 533-548.

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, *82*(1), 179-210.

Avolio, B. J. (2005). *Leadership development in balance: Made/born*. Psychology Press.

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, *107*(3), 714-47.

Baert, S., & De Pauw, A. S. (2014). Is ethnic discrimination due to distaste or statistics? *Economics Letters*, *125*(2), 270-273.

Bagues, M. F., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, *77*(4), 1301-1328.

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, *107*(4), 1207-38.

Baird, J. (2016, February 26). Sarah Palin's Mustache. The *New York Times*, A29.

Ball, S., Eckel, C. C., & Heracleous, M. (2010). Risk aversion and physical prowess: Prediction, choice and bias. *Journal of Risk and Uncertainty*, *41*(3), 167-193.

Barbulescu, R., & Bidwell, M. (2013). Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers. *Organization Science*, *24*(3), 737-756.

Bayer, A., & Rouse, C. E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, *30*(4), 221-242.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, *124*(4), 1497-1540.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, *335*(6068), 582-586.

Becker, G. (1971). *The economics of discrimination* (2nd ed.). University of Chicago Press.

Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, *25*(5), 581-589.

Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *The Journal of Socio-Economics*, *37*(5), 1775-1784.

Ben-Ner, A., McCall, B. P., Stephane, M., & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior & Organization*, *72*(1), 153-170.

Bennett, J. (2021, January 17). Meet the Groundbreaking New Second family. *The New York Times,* A14.

Bertrand, M. (2011). New Perspectives on Gender. In D. Card & O. Ashenfelter (Eds.). *Handbook of Labor Economics* (pp. 1543-1590). Elsevier. https://doi.org/10.1016/S0169-7218(11)02415-4

Bertrand, M. (2013). Career, family, and the well-being of college-educated women. *American Economic Review*, *103*(3), 244-50

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2019). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies*, *86*(1), 191-239.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, *95*(2), 94-98.

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American economic journal: applied economics*, *2*(3), 228-55.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, *94*(4), 991-1013.

Bilén, D., Dreber, A., & Johannesson, M. (2020). *Are women more generous than men? A Meta-Analysis.* SSRN. http://dx.doi.org/10.2139/ssrn.3578038

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, *89*(1), 335-339.

Bohnet, I., Van Geen, A., & Bazerman, M. (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science*, *62*(5), 1225-1234.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, *109*(10), 3395-3436.

Bolino, M., Long, D., & Turnley, W. (2016). Impression management in organizations: Critical questions, answers, and areas for future research. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 377-406.

Bolino, M. C., & Turnley, W. H. (2003). More than one way to make an impression: Exploring profiles of impression management. *Journal of Management*, *29*(2), 141-160

Bolton, G. E., & Katok, E. (1995). An experimental test for gender differences in beneficent behavior. *Economics Letters*, *48*(3-4), 287-292.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: a meta-analysis. *Journal of applied psychology*, *89*(5), 901.

Booth, A. L., & Nolen, P. (2012). Gender differences in risk behaviour: does nurture matter? *The economic journal*, *122*(558), F56-F78.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753-1794.

Boroditsky, L. (2019). Language and the brain. *Science, 336*(6461), 13. https://doi.org/10.1126/science.aaz6490

Bowles, H. R., & Babcock, L. (2013). How can women escape the compensation negotiation dilemma? Relational accounts are one answer. *Psychology of Women Quarterly*, *37*(1), 80-96.

Boyd, K. S. (2008). Glass Ceiling. In Schaefer, R.T. (Ed), *Encyclopedia of Race, Ethnicity, and Society* (pp. 549-52). SAGE.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, *30*(7), 1145-1159

Brands, R. A., & Fernandez-Mateo, I. (2017). Leaning out: How negative recruitment experiences shape women's decisions to compete for executive roles. *Administrative Science Quarterly*, *62*(3), 405-442.

Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour. *Economics Letters*, *170*, 19-23.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Breiman, L. (2001a). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, *16*(3), 199-231.

Breiman, L. (2001b). Random forests. *Machine learning*, *45*(1), 5-32.

Brenner, O. C., Tomkiewicz, J., & Schein, V. E. (1989). The relationship between sex role stereotypes and requisite management characteristics revisited. *Academy of management journal*, *32*(3), 662-669.

Broder, I. E. (1993). Review of NSF economics proposals: Gender and institutional patterns. *The American Economic Review*, *83*(4), 964-970.

Bruttel, L., & Stolley, F. (2018). Gender Differences in the Response to Decision Power and Responsibility—Framing Effects in a Dictator Game. *Games*, *9*(2), 28.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). *Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?* In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (p. 133–139). American Psychological Association. https://doi.org/10.1037/14805-009

Buolamwini, J. (2019, February 7). *Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve it.* TIME. https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Bursztyn, L., Fujiwara, T., & Pallais, A. (2017). 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, *107*(11), 3288-3319.

Buss, D. (2015). *Evolutionary psychology: The new science of the mind* (5th ed.). Psychology Press.

Büchel, B. (2018, March 1). *Artificial intelligence could reinforce society's gender equality problems.* THE CONVERSATION. https://theconversation.com/artificial-intelligence-could-reinforce-societys-gender-equality-problems-92631

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: a meta-analysis. *Psychological bulletin*, *125*(3), 367.

Cadsby, C. B., Servátka, M., & Song, F. (2013). How competitive are female professionals? A tale of identity conflict. *Journal of Economic Behavior & Organization*, *92*, 284-303.

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. *Trends in cognitive sciences*, *7*(5), 225-231.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic perspectives*, *9*(2), 209-219.

Cameron, D. (2005). Language, gender, and sexuality: Current issues and new directions. *Applied linguistics*, *26*(4), 482-502.

Cappelen, A. W., Falch, R., & Tungodden, B. (2019). *The boy crisis: Experimental evidence on the acceptance of males falling behind.* NHH Dept. of Economics Discussion Paper, (06). http://hdl.handle.net/11250/2589254

Cappelen, A. W., Konow, J., Sørensen, E. Ø., & Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, *103*(4), 1398-1413.

Carpenter, J. (2017, August 7). *Women in the Fortune 500: 64 CEOs in half a century.* CNN Money. https://money.cnn.com/interactive/pf/female-ceos-timeline/

Casaponsa, A., & Athanasopoulos, P. (2018, April 16). *The way you see colour depends on what language you speak.* THE CONVERSATION. https://theconversation.com/the-way-you-see-colour-depends-on-what-language-you-speak-94833

Casselman, B. & Tankersley, J. (2019, March 18). Women Face High Levels Of Sex Abuse in Economics. *The New York Times*, B1.

Catalyst (2020, January 15). *Pyramid: Women in S&P 500 Companies.* CATALYST. https://www.catalyst.org/research/women-in-sp-500-companies/

Cha, Y. (2010). Reinforcing separate spheres: The effect of spousal overwork on men's and women's employment in dual-earner households. *American Sociological Review*, *75*(2), 303-329.

Chakraborty, P., & Serra, D. (2018). *Gender differences in top leadership roles: Does aversion to worker backlash matter?* Departmental Working Papers (No. 1807), Southern Methodist University. https://ideas.repec.org/p/smu/ecowpa/1807.html

Chakravarty, S., Harrison, G. W., Haruvy, E. E., & Rutström, E. E. (2011). Are you risk averse over other people's money? *Southern Economic Journal*, *77*(4), 901-913.

Chamorro-Premuzic, T. & Gallop, C. (2020). 7 Leadership Lessons Men Can Learn from Women. *Harvard Business Review,* https://hbr.org/2020/04/7-leadership-lessons-men-can-learn-from-women

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, *68*(1), 29-35.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50-58.

Chira, S. (2017, July 21). Why Women Aren't C.E.O.s. *The New York Times,* SR1.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, *40*(11), 1705-1732.

Chozick, A. (2017, December 9). A Feminist Hero, Not By Design. *The New York Times,* BU1.

Coates, J. (2015). *Women, men and language: A sociolinguistic account of gender differences in language* (3rd ed.). Routledge.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science*, *15*(10), 687-693.

Colquitt, J. A. (2008). From the editors publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal, 51*(4), 616 – 620.

Connor, R. A., & Fiske, S. T. (2018). Warmth and competence: A feminist look at power and negotiation. In C. B. Travis, J. W. White, A. Rutherford, W. S. Williams, S. L. Cook, & K. F. Wyche (Eds.), *APA handbook of the psychology of women: History, theory, and battlegrounds,* 321-342. American Psychological Association.

Cook, A., & Glass, C. (2015). Diversity begets diversity? The effects of board composition on the appointment and success of women CEOs. *Social Science Research*, *53*, 137-147.

Cook, A., & Glass, C. (2016). Do women advance equity? The effect of gender leadership composition on LGBT-friendly policies in American firms. *Human Relations*, *69*(7), 1431-1456.

Cox, T. (2018, May 21). *Voice control: why AI must resist our bad habit of stereotyping human speech.* THE CONVERSATION. https://theconversation.com/voice-control-why-ai-must-resist-our-bad-habit-of-stereotyping-human-speech-96310

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448-74.

Dankers F.J.W.M., Traverso A., Wee L., & van Kuijk S.M.J. (2019). Prediction Modeling Methodology In P. Kubben, M. Dumontier, & A. Dekker (Eds). *Fundamentals of Clinical Data Science,* chapter 8. Springer. https://www.ncbi.nlm.nih.gov/books/NBK543534/ doi: 10.1007/978-3-319-99713-1_8

Daruvala, D. (2007). Gender, risk and stereotypes. *Journal of Risk and Uncertainty*, *35*(3), 265-283.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of experimental social psychology*, *40*(5), 642-658.

De Dreu, C.K.W., Evers, A., Beersma, B., Kluwer, E.S., & Nauta, A. (2001), A theory-based measure of conflict management strategies in the workplace. *J. Organiz. Behav.,* 22: 645-668. https://doi.org/10.1002/job.107

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, *114*(23), 5964-5969.

Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly*, *27*(3), 456-469.

Desilver, D. (2018). *Few women in US companies are CEOs and in the jobs that lead there.* Pew Research Center Technical report. https://pewrsr.ch/2KlB5Dk

Dezsö, C. L., & Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic management journal*, *33*(9), 1072-1089.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, *9*(3), 522-550.

Doornenbal, B. M., Spisak, B. R., & van der Laken, P. A. (2021). Opening the black box: Uncovering the leader trait paradigm through machine learning. *The Leadership Quarterly*, 101515.

Drupp, M. A., Khadjavi, M., Riekhof, M. C., & Voss, R. (2020). Professional identity and the gender gap in risk-taking. Evidence from field experiments with scientists. *Journal of Economic Behavior & Organization*, *170*, 418-432.

Dublin, T. (1979). *Women at work: The transformation of work and community in Lowell, Massachusetts, 1826-1860*. Columbia University Press.

Dupas, P., Modestino, A. S., Niederle, M., & Wolfers, J. (2021). *Gender and the dynamics of economics seminars* (NBER Working Paper No. 28494). National Bureau of Economic Research. https://www.nber.org/papers/w28494

Eagly, A. H. (1987). *Sex differences in social behavior* (1st ed.). Psychology Press.

Eagly, A. H. (2007). Female leadership advantage and disadvantage: Resolving the contradictions. *Psychology of women quarterly*, *31*(1), 1-12.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Harvard Business Press.

Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological bulletin*, *129*(4), 569.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist, 75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eagly, A., Wood, W., & Diekman, A. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). New York: Psychology Press.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The economic journal*, *108*(448), 726-735.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic inquiry*, *39*(2), 171-188.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, *23*(4), 281-295.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. In C.R. Plott & V.L. Smith (Eds), *Handbook of experimental economics results*, (pp. 1061-1073). Elsevier. doi.org/10.1016/S1574-0722(07)00113-8

Egan, M. L., Matvos, G., & Seru, A. (2017). *When Harry fired Sally: The double standard in punishing misconduct* (NBER Working Paper No. 23242). National Bureau of Economic Research. https://www.nber.org/papers/w23242

Ely, R. J. (1994). The effects of organizational demographics and social identity on relationships among professional women. *Administrative science quarterly*, 203-238.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, *14*(4), 583-610.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817-868.

Fellbaum, C. (1998). *WordNet: An electronic lexical database.* The MIT Press

Fernandez-Mateo, I., & Fernandez, R. M. (2016). Bending the pipeline? Executive search and gender inequality in hiring for top management jobs. *Management Science*, *62*(12), 3636-3655.

Field, A. P. (2017). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll* (5th ed.). SAGE.

Finseraas, H., Johnsen, Å. A., Kotsadam, A., & Torsvik, G. (2016). Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *European Economic Review*, *90*, 363-374.

Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, *81*(1), 74-81.

Friedl, A., Pondorfer, A., & Schmidt, U. (2020). Gender differences in social risk taking. *Journal of Economic Psychology*, *77*, 102182.

Fu, L., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Tie-breaker: Using language models to quantify gender bias in sports journalism. *arXiv preprint arXiv:1607.03895*.

Fung, P. (2019, June 30). *This is why AI has a gender problem.* The Word Economic Forum. https://www.weforum.org/agenda/2019/06/this-is-why-ai-has-a-gender-problem/

Gardner, W. L., & Avolio, B. J. (1998). The charismatic relationship: A dramaturgical perspective. *Academy of management review*, *23*(1), 32-58.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535-74.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal, 59* (5), 1493-1507

Glass, C., & Cook, A. (2016). Leading at the top: Understanding women's challenges above the glass ceiling. *The Leadership Quarterly*, *27*(1), 51-63.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of personality and social psychology*, *70*(3), 491.

Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.

Goldberg, P. (1968). Are women prejudiced against women?. *Society, 5* (5), 28-30.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American economic review*, *96*(2), 1-21.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091-1119.

Goldin, C., & Katz, L. F. (2000). Career and Marriage in the Age of the Pill. *American Economic Review*, *90*(2), 461-465.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American economic review*, *90*(4), 715-741.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, *82*(3), 379.

Grady, C. (2018, April 24). *Hillary Clinton and Chimamanda Ngozi Adichie talk censorship, feminism, and Pizzagate.* Vox. https://www.vox.com/culture/2018/4/24/17271638/hillary-clinton-chimamanda-ngozi-adichie-arthur-miller-freedom-write-pen-america-world-voices

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically smalleffects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561. http://dx.doi.org/10.1037/pspa0000016

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, *97*(1), 17.

Grossman, P. J., & Lugovskyy, O. (2011). An Experimental Test of the Persistence of Gender-based Stereotypes. *Economic inquiry*, *49*(2), 598-611.

Guadagno, R. E., & Cialdini, R. B. (2007). Gender differences in impression management in organizations: a qualitative review. *Sex Roles*, *56*(7-8), 483-494.

Güth, W., & Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization*, *108*, 396-409.

Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review, 32,* (2), 334–343.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Harrison, G. W. (2006). Hypothetical bias over uncertain outcomes. In J. A. List (Ed.), *Using Experimental Methods in Environmental and Resource Economics* (pp. 41-69). Edward Elgar Publishing.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, *40*(8), 1316–1330. https://doi.org/10.1002/smj.3023

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

He, J., & Kang, S. (2019). Covering in cover letters: Gender and self-presentation in job applications. *Academy of Management Proceedings*. https://doi.org/10.5465/AMBPP.2019.275

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of social issues*, *57*(4), 657-674.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, *32*, 113-135.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: the implied communality deficit. *Journal of applied psychology*, *92*(1), 81.

Heinz, M., Juranek, S., & Rau, H. A. (2012). Do women behave more reciprocally than men? Gender differences in real effort dictator games. *Journal of Economic Behavior & Organization*, *83*(1), 105-110.

Herbert, B. (2006, May 18). Hillary Can Run, but Can She Win? *The New York Times.* https://www.nytimes.com/2006/05/18/opinion/18herbert.html

Hernández-Orallo, J., Flach, P., & Ferri, C. (2013). ROC curves in cost space. *Machine learning*, *93*(1), 71-91.

Hinchliffe, E. (2020, May 18). *The number of female CEOs in the Fortune 500 hits an all-time record.* FORTUNE. https://fortune.com/2020/05/18/women-ceos-fortune-500-2020/

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Analysis and Machine Intelligence, 20*(8) 832-844.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644-1655.

Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, *95*(3), 902-912.

Hsee, C. K., & Weber, E. U. (1997). A fundamental prediction error: Self-others discrepancies in risk preference. *Journal of Experimental Psychology: General*, 126, 45–53.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, *60*(6), 581.

Hymowitz, C., & Schellhardt, T. D. (1986, March 24). The Glass Ceiling: Why Women Can't Seem to Break the Invisible Barrier That Blocks Them from the Top Jobs. *The Wall Street Journal,* 1.

Ibarra, H. (1999). Provisional selves: Experimenting with image and identity in professional adaptation. *Administrative science quarterly*, *44*(4), 764-791.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and cognition*, *7*(2), 265-290.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* New York: Springer.

Jones, J. J. (2016). Talk "like a man": The linguistic styles of Hillary Clinton, 1992–2013. *Perspectives on Politics*, *14*(3), 625-642.

Jordan, K. N., & Pennebaker, J. W. (2017). The exception or the rule: Using words to assess analytic thinking, Donald Trump, and the American presidency. *Translational Issues in Psychological Science, 3*(3), 312–316. https://doi.org/10.1037/tps0000125

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, *93*(5), 1449-1475.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business, 59*(4), 285-300.

Kark, R., Waismel-Manor, R., & Shamir, B. (2012). Does valuing androgyny and femininity lead to a female advantage? The relationship between gender-role, transformational leadership and identification. *The Leadership Quarterly*, *23*(3), 620-640.

Kent, R. L., & Moss, S. E. (1994). Effects of sex and gender role on leader emergence. *Academy of Management Journal*, *37*(5), 1335-1346.

King, E. B., Botsford, W., Hebl, M. R., Kazama, S., Dawson, J. F., & Perkins, A. (2012). Benevolent sexism at work: Gender differences in the distribution of challenging developmental experiences. *Journal of Management*, *38*(6), 1835-1866.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181-209.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological bulletin*, *137*(4), 616.

Konow, J., Saijo, T., & Akai, K. (2020). Equity versus equality: Spectators, stakeholders and groups. *Journal of Economic Psychology*, *77*, 102171.

Krysowski, E., & Tremewan, J. (2021). Why Does Anonymity Make Us Misbehave: Different Norms Or Less Compliance? *Economic Inquiry*, *59*(2), 776-789.

Lakoff, R. T. (2004). *Language and woman's place: Text and commentaries*. Oxford University Press.

Larsen, R. J., & Seidman, E. (1986). Gender schema theory and sex role inventories: Some conceptual and psychometric considerations. *Journal of Personality and Social Psychology*, *50*(1), 205.

Lee, D. J. (2018). Does implicit bias predict dictator giving? *Games*, *9*(4), 73.

Lemoine, G. J., & Blum, T. C. (2021). Servant leadership, leader gender, and team gender role: Testing a female advantage in a cascading model of performance. *Personnel Psychology, 74(1), 3-28.*

Lerer, L. & Ember, S. (2020, November 7). Historic Milestone for Woman Steeped in Civil Rights Battles. *The New York Times,* A1.

Li, W.-D., Li, S., Feng, J. J., Wang, M., Zhang, H., Frese, M., & Wu, C.-H. (2020). Can becoming a leader change your personality? An investigation with two longitudinal studies from a role-based perspective. *Journal of Applied Psychology*, *106*(6), 882-901. https://doi.org/10.1037/apl0000808

Madgavkar, A., Manyika, J., Krishnan, M., Elingrud, K., Yee, L., Woetzel, J., Chui, M., Hunt, V., & Balakrishnan, S. (2019, June 4). *The future of women at work: Transitions in the age of automation.* McKinsey Global Institute. https://www.mckinsey.com/featured-insights/gender-equality/the-future-of-women-at-work-transitions-in-the-age-of-automation.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60

Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *128*(584), 2145-2166.

Matsa, D. A., & Miller, A. R. (2013). A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics*, *5*(3), 136-69.

Mead, R. (2006). *How the vote was won: woman suffrage in the western United States, 1868-1914*. NYU Press.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862.

Melo F. (2013) Area under the ROC Curve. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (Eds.) *Encyclopedia of Systems Biology*. Springer. https://doi.org/10.1007/978-1-4419-9863-7_209

Miller, C.C. (2016, March 14). Measuring Trump's Language: Bluster but Also Words That Appeal to Women. *The New York Times.* https://www.nytimes.com/2016/03/15/upshot/donald-trump-is-among-the-most-feminine-sounding-candidates.html

Miller, C.C., & Haridasani Gupta, A. (2020, October 15). Mothers in Public Office Still Walk a 'Tightrope.' *The New York Times,* A1.

Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media, Inc.

Mokyr, J. (1992). *The lever of riches: Technological creativity and economic progress*. Oxford University Press.

Mook, D. G. (1983). In defense of external invalidity. *American psychologist*, *38*(4), 379.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the national academy of sciences*, *109*(41), 16474-16479.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research*, *27*(1), 121-152.

Mullainathan, S. (2017, September 9). Sexism and Shopping: Female Players Get Most of the Odd Questions at the U.S. Open. *The New York Times.* https://www.nytimes.com/2017/09/08/upshot/sexism-and-shopping-female-players-get-most-of-the-odd-questions-at-the-us-open.html

Mullainathan, S., & Thaler, R. H. (2000). *Behavioral economics* (NBER Working Paper No. 7948). National Bureau of Economic Research. http://www.nber.org/papers/w7948

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. O'Reilly Media, Inc.

Murphy, H. (2018, March 16). Draw a Leader. What's She a Like? Trick Question! *The New York Times,* A1.

Murphy, S. A., Fisher, P. A., & Robie, C. (2021). International comparison of gender differences in the five-factor model of personality: An investigation across 105 countries. *Journal of Research in Personality*, *90*, 104047.

National Institute on Deafness and Other Communication Disorders. (n.d.). *Areas of the brain affected by Broca's and Wernicke's aphasia* [Fact Sheet]. https://www.nidcd.nih.gov/sites/default/files/Documents/health/voice/Aphasia6-1-16.pdf

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1-16.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The quarterly journal of economics*, *122*(3), 1067-1101.

Nilsen, E. (2019, February 5). *New evidence has emerged Elizabeth Warren claimed American Indian heritage in 1986.* Vox. https://www.vox.com/2018/10/16/17983250/elizabeth-warren-bar-application-american-indian-dna

Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, *36*(7), 923-936.

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, *65*(3-4), 703-713.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85*)*, 2825–2830.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us.* Bloomsbury Press.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS one*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007* [LIWC manual]. Austin, TX: LIWC.net

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*(3), 271-282.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, *54*(1), 547-577.

Phelan, J. E., & Rudman, L. A. (2010). Prejudice toward female leaders: Backlash effects and women's impression management dilemma. *Social and Personality Psychology Compass*, *4*(10), 807–820. https://doi.org/10.1111/j.1751-9004.2010.00306.x

Phillips, K.W. (2014, October 1). *How Diversity Makes Us Smarter.* SCIENTIFIC AMERICAN. https://www.scientificamerican.com/article/how-diversity-makes-us-smarter/

Porter, M. F. (1980). An algorithm for suffix stripping. *Program: Electronic library and information systems, 14*(3), 130–137. https://doi.org/10.1108/eb046814

Porter, C., & Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, *12*(3), 226-54.

Prizes and Awards. 100 Years of Award-Winning Work (2021, June 4). Retrieved from https://www.nytco.com/company/prizes-awards/100-years-of-award-winning-work/

Proto, E., Sgroi, D., & Nazneen, M. (2019). Happiness, cooperation and language. *Journal of Economic Behavior & Organization*, *168*, 209-228.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403-4408.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Rosalsky, G. (2020, January 7). *Economics Still Has A Diversity Problem.* PLANET MONEY: NPR. https://www.npr.org/sections/money/2020/01/07/793855832/economics-still-has-a-diversity-problem

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, *95*(2), 221.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, *74*(3), 629.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*(5), 1004.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin*, *26*(11), 1315-1328.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165-179.

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in organizational behavior*, *28*, 61-79.

Ryan, M. K., & Haslam, S. A. (2007). The glass cliff: Exploring the dynamics surrounding the appointment of women to precarious leadership positions. *Academy of management review*, *32*(2), 549-572.

Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. (2011). Think crisis–think female: The glass cliff and contextual variation in the think manager–think male stereotype. *Journal of Applied Psychology*, *96*(3), 470.

Ryan, M. K., Haslam, S. A., Hersby, M. D., Kulich, C., & Atkins, C. (2007). Opting out or pushed off the edge? The glass cliff and the precariousness of women's leadership positions. *Social and Personality Psychology Compass*, *1*(1), 266-279.

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly*, *27*(3), 446-455.

Sarsons, H. (2017 a). Recognition for group work: Gender differences in academia. *American Economic Review*, *107*(5), 141-45.

Sarsons, H. (2017 b). Interpreting signals in the labor market: evidence from medical referrals. *https://scholar.harvard.edu/files/sarsons/files/sarsons_jmp.pdf*

Schlenker, B. R. (2012). *Self-presentation.* In M. Leary & J. Tangney (Eds.), Handbook of self and identity (2nd ed.), 542–570. New York, NY: The Guildford Press.

Schubert, R., Brown, M., Gysler, M., & Brachinger, H. W. (1999). Financial decision-making: are women really more risk-averse? *American economic review*, *89*(2), 381-385.

Schwab, K. (2016, January 14). *The Fourth Industrial Revolution: what it means, how to respond.* The World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Sedivy, J. (2016, October 25). *Donald Trump Talks Like a Woman.* POLITICO MAGAZINE. https://www.politico.com/magazine/story/2016/10/trump-feminine-speaking-style-214391

Sergent, K., & Stajkovic, A. D. (2020). Women's leadership is associated with fewer deaths during the COVID-19 crisis: Quantitative and qualitative analyses of United States governors. *Journal of Applied Psychology, 105*(8), 771–783. https://doi.org/10.1037/apl0000577

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, *5*(1), 63-68.

Sharma, S. (2015). Gender and distributional preferences: Experimental evidence from India. *Journal of Economic Psychology*, *50*, 113-123.

Siegrist, M., Cvetkovich, G., & Gutscher, H. (2002). Risk preference predictions and gender stereotypes. *Organizational Behavior and Human Decision Processes*, *87*(1), 91-102.

Simonite, T. (2018, August 17). *AI is the Future—But Where Are the Women?* WIRED. https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/

Slatcher, R. B., Chung, C. K., Pennebaker, J. W., & Stone, L. D. (2007). Winning words: Individual differences in linguistic style among US presidential and vice presidential candidates. *Journal of Research in Personality*, *41*(1), 63-75.

Smith, A. N., Watkins, M. B., Burke, M. J., Christian, M. S., Smith, C. E., Hall, A., & Simms, S. (2013). Gendered influence: A gender role perspective on the use and effectiveness of influence tactics. *Journal of Management*, *39*(5), 1156-1183.

Solnick, S. J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*(2), 189-200.

Stephan, G., Dütsch, M., Gückelhorn, C., & Struck, O. (2014). When are bonus payments for managers perceived as fair? Results from a quasi-experiment. *Economics Letters*, *125*(1), 130-133.

Stevenson, J. E., & Orr, E. (2017). We interviewed 57 female CEOs to find out how more women can get to the top. *Harvard Business Review.*

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111-133.

Stryker, S. (2007). Identity theory and personality theory: Mutual relevance. *Journal of Personality*, *75*(6), 1083–1102. https://doi.org/10.1111/j.1467-6494.2007.00468.x

Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment?. *Psychological science*, *30*(3), 405-414.

Taalbi, J. (2019). Origins and pathways of innovation in the third industrial revolution. *Industrial and corporate change, 28*(5), 1125-1148.

Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley, S. A., Polsinelli, A. J., & Mehl, M. R. (2020). 'Personality in its Natural Habitat' Revisited: A Pooled, Multi–sample Examination of the Relationships between the Big Five Personality Traits and Daily Behaviour and Language Use. *European Journal of Personality*, *34*(5), 753–776. https://doi.org/10.1002/per.2283

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Ullmann, S. & Saunders, D. (2021, March 30). *Online translators are sexist—here's how we gave them a little gender sensitivity training.* THE CONVERSATION. https://theconversation.com/online-translators-are-sexist-heres-how-we-gave-them-a-little-gender-sensitivity-training-157846

United Nations Educational, Scientific and Cultural Organization. (2019). *Women in Science* [Fact sheet]. http://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc.

Vinkenburg, C. J., Van Engen, M. L., Eagly, A. H., & Johannesen-Schmidt, M. C. (2011). An exploration of stereotypical beliefs about leadership styles: Is transformational leadership a route to women's promotion?. *The Leadership Quarterly*, *22*(1), 10-21.

von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The Cases of the 2016 United States and 2017 United Kingdom elections. *Psychological science*, *31*(2), 115-128.

Wang, S. (2018). *As The New York Times extends its reach across countries (and languages and cultures), it looks to locals for guidance.* NiemanLab. https://www.niemanlab.org/2018/07/as-the-new-york-times-extends-its-reach-across-countries-and-languages-and-cultures-it-looks-to-locals-for-guidance/

Wang, S., & Chen, X. (2020). Recognizing CEO personality and its impact on business performance: Mining linguistic cues from social media. *Information & Management*, *57*(5), 103173.

Wenzel, R., & Van Quaquebeke, N. (2018). The double-edged sword of big data in organizational and management research: A review of opportunities and risks. *Organizational Research Methods*, *21*(3), 548-591.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1*, 80–83.

Wille, B., Wiernik, B. M., Vergauwe, J., Vrijdags, A., & Trbovic, N. (2018). Personality characteristics of male and female executives: Distinct pathways to success?. *Journal of Vocational Behavior*, *106*, 220-235.

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, *142*(2), 165.

Wilton, L. S., Sanchez, D. T., Unzueta, M. M., Kaiser, C., & Caluori, N. (2019). In good company: When gender diversity boosts a company's reputation. *Psychology of Women Quarterly*, *43*(1), 59-72.

Wiseman, D. B., & Levin, I. P. (1996). Comparing risky decision making under conditions of real and hypothetical consequences. *Organizational behavior and human decision processes*, *66*(3), 241-250.

World Economic Forum. (2018). *The global gender gap report 2018.* https://www.weforum.org/reports/the-global-gender-gap-report-2018

Wu, A. H. (2018). Gendered language on the economics job market rumors forum. *AEA Papers and Proceedings*, 108, 175-79.

Yu, B. (2014). Language and gender in Congressional speech. *Literary and Linguistic Computing*, *29*(1), 118-132.

Zhou, L. (2018, November 7). *It's official: a record-breaking number of women have won seats in Congress.* Vox. https://www.vox.com/policy-and-politics/2018/11/7/18024742/midterm-results-record-women-win

Appendix - Proof: Bias – variance decomposition

$Y = f(X) + \in$ (10) and $E(\in) = 0$

$\hat{Y} = \hat{f}(X)$ (11)

$$E[MSE] = E[\frac{1}{N}\sum_{i=1}^{N}(Y-\hat{Y})^2] = \frac{1}{N}\sum_{i=1}^{N}E[(Y-\hat{Y})^2]$$

$$E[(Y-\hat{Y})]^2 = E\left[\left(Y - f(X) + f(X) - \hat{Y}\right)^2\right]$$

$$= E[(Y-f(X))^2] + E\left[\left(f(X)-\hat{Y}\right)^2\right] + 2E\left[\left(f(X)-\hat{Y}\right)(Y-f(X))\right]$$

$$= E[\in^2] + E\left[\left(f(X)-\hat{Y}\right)^2\right] + 2\left(E(f(X)Y) - E(f(X)^2) - E(\hat{Y}Y) + E\left(\hat{Y}f(X)\right)\right)$$

$$E(f(X)Y) = Y^2 \; since \; f(X) \; is \; deterministic \; and \; E(Y) = f(X)$$

$$E(f(X)^2) = f(X)^2 \; since \; f(X) \; is \; deterministic$$

$$E(Y\hat{Y}) = E[\hat{Y}((f(X)+\in)] = E\left[\hat{Y}f(X) + \hat{Y}\in\right] = E\left[\hat{Y}f(X)\right] + 0$$

$$E[(Y - \hat{Y})]^2 = E[\epsilon^2] + E\left[(f(X) - \hat{Y})^2\right]$$

$$E\left[(f(X) - \hat{Y})^2\right] = E\left[(f(X) - E[\hat{Y}] + E[\hat{Y}] - \hat{Y})^2\right]$$

$$= E\left[(f(X) - E[\hat{Y}])^2\right] + E\left[(E(\hat{Y}) - \hat{Y})^2\right] + 2E\left[(E(\hat{Y}) - \hat{Y})\left(f(X) - E(\hat{Y})\right)\right]$$

$$= bias(\hat{Y})^2 + Var(\hat{Y}) + 2\left[E\left[f(X)E[\hat{Y}]\right] - E[E[\hat{Y}]^2] - E[\hat{Y}f(X)] + E\left[\hat{Y}E[\hat{Y}]\right]\right]$$

$$E\left[f(X)E[\hat{Y}]\right] = f(X)E[\hat{Y}] \; since\ f\ is\ deterministic\ and\ E\left[E[a]\right] = a$$

$$E[\hat{Y}f(X)] = f(X)E[\hat{Y}]$$

$$E\left[\hat{Y}E[\hat{Y}]\right] = E[\hat{Y}]^2$$

$$E\left[(f(X) - \hat{Y})^2\right] = bias^2 + Var(\hat{Y}) \text{ where } f(X) = Y \text{ and } \hat{Y} = \hat{f}(X)$$

$$E\left[(Y - \hat{Y})^2\right] = bias(\hat{Y})^2 + Var(\hat{Y}) + Var\,(\epsilon)$$

Source:
http://www.inf.ed.ac.uk/teaching/courses/mlsc/Notes/Lecture4/BiasVariance.pdf

Essays I – II – III

Gender, inequality and risk taking

Sebastian Fest Nur Yaldiz Ola Kvaløy *

January 29, 2022

Abstract

We study experimentally whether people find economic inequalities that result from risk taking equally acceptable when males rather than females are taking risks. In the experiment, participants that act as third-party spectators make re-distributive decisions involving a pair of workers who chose to be compensated for their work effort through a risky prospect. We randomly vary the gender composition of the worker pair while keeping opportunities, choices and the resulting inequalities between workers constant. We find no evidence that the gender of unlucky risk-takers affect how they are compensated for their loss. However, we find that relative to female participants, male participants are less inclined to redistribute earnings from lucky workers to unlucky workers if the lucky worker is male rather than female. Specifically, males are fourteen percent more likely to leave unlucky workers with no earnings if they are male rather than female. We further investigate whether this behavior can be attributed to differences in explicit and implicit attitudes towards gender roles in our sample. We find no indication that males' discriminatory behavior towards male winners is driven by either measure.

JEL Classification: C91, D63, J16
Keywords: Fairness, Discrimination, Gender-bias

---

*Fest: Norwegian School of Economics. Kvaløy and Yaldiz, University of Stavanger, UiS Business School, 4036 Stavanger, Norway (e-mail: sebastian.fest@gmail.com; nur.yaldiz@uis.no; ola.kvaloy@uis.no); We thank the participants of the Workshop on Behavioral Economics at FAIR in Bergen on Recognition and Feedback. Financial support from the Norwegian Research Council is gratefully acknowledged.

1 Introduction

In a number of contexts there seems to be a tendency towards woman being less risk tolerant than men (Eckel and Grossman, 2008a; Croson and Gneezy, 2009; Charness and Gneezy, 2012).[1] In addition, people take account of this tendency when asked about how likely they believe others will engage in a risky activity (Ball, Eckel, and Heracleous, 2010; Daruvala, 2007; Eckel and Grossman, 2002, 2008b; Grossman, 2013; Hsee and Weber, 1997; Siegrist, Cvetkovich, and Gutscher, 2002). In particular, people oftentimes correctly predict the direction of the gender gap in risk attitudes while incorrectly estimating the actual size of that gap.

While there is mixed evidence on gender differences in risk-taking behavior, the presence of beliefs of such differences, i.e. gender stereotyping, may still indicate that risk taking behavior might be evaluated differently depending on the gender of the risk taker. Recent evidence from labor market points to the existence of such discrimination. For example, female financial advisers have been found to be more likely to be fired for misconduct than their male counterparts (Egan, Matvos, and Seru, 2017). Likewise, female surgeons, who are rewarded less after a bad outcome of risky surgical procedures by receiving fewer future referrals (Sarsons, 2017).

In this paper, we ask whether outcomes that follow from risky decision are indeed evaluated differently based on the gender of the risk taker.

This experiment is based on a hypothetical dictator game. In the experiment, spectators make a distribution decision between two hypothetical workers. These individuals are same-sex or opposite-sex pairs engaging in similar tasks. After completing the given task and having their compensation, workers are given a chance to make additional earnings. There are two earning options, one is risky, and the other one is safe. Based on the worker's choice in this hypothetical situation, the spectator decides how to redistribute the earned amount. This gives us a chance to see how men and women hold risk takers responsible for unlucky outcomes. In particular, we investigate whether and who people evaluate and reward risk-taking decisions from men and women differently.

Our main results are as follow: We find no evidence that the gender of unlucky risk-takers affect how they are compensated for their loss. However, we find that relative to female participants, male participants are less inclined to redistribute earnings from lucky workers to unlucky workers if the lucky worker is male rather than female. Specifically, males are fourteen percent more likely to leave unlucky workers with no earnings if they are male rather than female. We

[1]Although recent evidence concerning a gender gap in risk taking is mixed, it has been used in an attempt to explain differences in life outcomes between men and woman including occupational choice (Sapienza, Zingales, and Maestripieri, 2009; Buser, Niederle, and Oosterbeek, 2014), asset allocation (Almenberg and Dreber, 2015; Marinelli, Mazzoli, and Palmucci, 2017), alcohol and drug use (Spigner, 1993), recreational choice (Boverie, Scheuffele, and Raymond, 1995) and sexual behavior (Schroth, 1996).

also find that the gender of spectators mattters. Female spectators compensate worse off workers more than male spectators, regardless of gender. In contrast, male spectators less willing to redistribute from lucky male workers rather than lucky female workers. We also find no evidence that implicit associations towards women or men can predict discriminatory behavior.

Previous studies provide mixed findings for redistribution decisions in dictator games. In merit settings, male dictators share less with male receivers (Sharma, 2015). In explaining dictators' giving behavior based on gender differences, Ben – Ner, Kong and Putterman (2004) find that females share less with female recipients than male and anonymous recipients. Kamas and Preston (2015) argue that female dictators tend to provide more egalitarian distribution than their male counterparts. A meta-analysis in dictators giving behavior also shows that female dictators share more than men (Engel, 2011).

Several studies employ spectator design to explore social preferences in various settings. This particular design provides a relevant foundation to understand the social preferences of third party individuals (e.g., Konow, 2009). Cappelen et al. (2019) study whether spectators hold low-performing males or females more responsible in a merit environment. Konow, Saijo, and Akai (2020) investigate distribution decisions based on shared group identity and performance in a merit context. Cappelen et al. (2013) study social preferences in the face of risky behaviors. Spectators tend to provide equal redistribution between two risk-takers rather than between a risk-taker and a non-risk taker.

Social preferences (e.g., Levitt & List, 2007) and risk-taking behavior (e.g., Filippin & Crosetto, 2016) have been a long subject to the behavioral economics literature. Specifically, studies employ hypothetical stakes to compare the outcomes of social preferences with real incentives (e.g., Ben-Ner,Kramer, & Levy, 2008; Bühren, & Kundt, 2015).

Recently, crowdsourcing markets as Mechanical Turk (mTurk) provide a new outlet to employ an international subject pool (Chandler & Kapelner, 2013) for survey experiments (Horton, Rand, & Zeckhauser, 2011) involving solely hypothetical stakes (e.g., Bechler, Green, & Myerson, 2015).

In this study, we employ a modified dictator game in an online survey experiment. The spectators make hypothetical redistribution decisions involving opposite or same-sex pairs, and luck serves as a source of inequality.

The structure of this paper as follows. Section 2 provides an experimental design and the details of explicit and implicit measures, samples and procedures, and hypotheses. Section 3 present the results, followe Section 4 conclusion. The appendix provides information about the instructions.

2 The experiment

2.1 Design

Subjects participating in the experiment were asked to make a redistribute decision for a hypothetical scenario involving two workers. In the scenario, two workers had been hired to work equally long on the same type of task for four hours before each of them was given the opportunity to decide on how to be compensated for their work. In particular, each worker had the same opportunity of choosing between a safe amount of 20 USD and a risky prospect for which there was an equal chance to end up with 60 USD or nothing. Importantly, we revealed no specific information about the type of task or differences in task performance to participants.

We furthermore described to participants - denoted spectators- that both workers in the pair made the same decision of choosing the risky prospect. However, we pair workers such that, as outcomes of the risky prospects are realized, one worker ends up with 60 USD in earnings while the other ends up with no earnings at all. Thus, in the hypothetical scenario described to participants, we presented an economic inequality between two otherwise identical workers who provided the same, unspecified level of effort, faced identical opportunities and made similar choices. We thus control for ability, effort and choice as factor that might motivate participants in our experiment to evaluate outcomes differently for different workers.

After having read the scenario, spectators were asked to state how much of the winning worker's earnings they wanted to transfer to the worker who earned nothing. Specifically, they could indicate how much they would be willing to transfer to the worse-off worker in increments of 5 USD up to the whole amount of the total earnings for the better-off worker. All participants were requested to make a single redistributive decision involving a hypothetical pair of workers.

In order to see if risk taking is considered equally acceptable when males rather than females are taking risks, we randomly vary the gender composition of the worker pair in the description presented to the spectators. In particular, between subjects, we randomly vary the sex of the better-off and worse-off worker in the pair by explicitly revealing the gender of the worker in the description. Spectators therefore either met with a pair of single-sex workers or a pair of mixed-sex workers in which the better- and worse-off worker was either male or female, respectively.[2] The resulting 2x2 treatment is summarized in Table 1. See also appendix for instructions to participants.

[2]These kinds of gender-based designs are prone to experimenter demand effect (e.g., De Quidt et al., 2018), but we believe that our neutral design limits this problem. To reduce the bias as much as possible, we chose these names from the list of most popular names for births between 1921 – 2020 (https://www.ssa.gov/oact/babynames/decades/century.html). Male names are John and William, and female names are Ann and Emma. We are aware of the bias that name could cause from previous studies (e.g., Bertrand & Mullainathan, 2004; Anseel & Duyck, 2008).

Since we only vary the gender compositions of the pair of workers, our treatments allow us to investigate whether better-off risk-taking men are considered to be more entitled to their gains than better-off risk taking woman. Furthermore, our design allows us to identify whether worse-off risk takers are compensated for their work effort more depending on their gender. In addition, treatments with different gender compositions help us to investigate whether risk taking women held more responsible for unlucky outcomes compared to unlucky men. We also observe whether risk-taking women held responsible more, in same sex context.

[Table 1 about here]

2.2 Explicit Scales

After making distributive decisions, the spectators completed two explicit scales of gender. The first scale is the Social Roles Questionnaire, measure of gender role attitudes. This scale aims to capture the different views on the social roles of men and women without pusing the idea of stereotyping. There are 13 items of which first five questions focuses on gender egalitarian view (Gender Transcendent Scale). The rest of the questions are centered around more traditional view. Internal and test reliability of SRQ has a Cronbach α of .86 for the General Scale and .71 for the Gender Transcendent Scale. High test-retest reliability is found for both the General scale r = .92, p < .001, and the Gender Transcendent scale, r = .81, p < .001 (Baber & Tucker, 2006).

The second expilicit test is Bem Sex Role Inventory, aims to determine how individuals describe themselves in terms of socially desirable female and male traits. This measure is the extension of defining gender as a cultural concept determined with social characteristics, regardless of one's sex. In total, there are 60 items. 20 items measure femininity and 20 items measure masculinity. There are 20 other filler items, which are gender neutral but focuses on social desirability. Coefficient alpha is calculated separately for three categories. The findings indicate reliable results for internal consistency (Masculinity α = .86; Femininity α =. 80; Social Desirability α = .75). Androgyny (a person having both strong masculine and strong feminine features) score is found as α = .85. Product – moment correlations show that BSRI has high reliability (Masculinity r = .90; Femininity r = .90; Androgyny r = .93; Social Desirability r = .89).

2.3 Implicit Measure

Implicit association test helps to discover automatic mental connections that individuals build towards the social constructs of their environment. Implicit measures most likely reveal the hidden association between different concepts (Greenwald, McGhee, & Schwartz, 1998). For this study we partially adapted a gender potency IAT design from Rudman, Greenwald and McGhee (2001).

The IAT generated with 60 stimulus items, which consists of 15 female names, 15 male names, 15 potent meaning words and 15 weak meaning words. Male names, and power attributes, appear on the top left side of the computer screen and female names, and weak attributes, appear on the top right side of the computer screen. The respondents press a left and a right key to match attributes and targets. The difference in response latencies between generally associated attributes compared to uncommon attributes determine the strength of implicit stereotyping (Rudman et al., 2001). Iatgen software is used to embed the whole test into Qualtrics (Carpenter et al., 2018).

2.4 Sample and procedures

We recruited 400 participants through Amazon Mechanical Turk from the United States. All participants followed a link to our survey experiment which was hosted externally on Qualtrics. Average completion time amounted to approximately 14 minutes. All completed submissions where approved within a day and awarded with 6 USD.

After being randomly assigned to one of the four treatments, each participant was taken to the distributive stage before continuing responding to items associated with implicit and explicit scales of discrimination and a general set of background questions.[3]

Table 2 shows background characteristics of subjects participating in the experiment. On average, the subjects have a 2-year college degree with a yearly gross income (before taxes) of 50 000 UDS and are 37 years old. The general political view of spectators mediates between being neutral to liberal. Importantly, our sample is highly gender-balanced, with 46% being females.

2.5 Hypotheses

We build the following hypotheses to explore whether spectators treat gender differently. We assume that stereotypes may drive redistribution decisions. Since our design cannot pinpoint the role of stereotypical beliefs in these decisions, we do not formulate hypotheses on whether or not gendered expectations motivate the results.

Hypothesis 1 (H.1) *Worse-off workers are compensated independent of their gender.*

Stereotypical beliefs about how men and women should behave when confronted with taking risks may lead to smaller compensation for worse-off females since they act contrary to these beliefs but still are considered to be entitled to

[3]We also ask participants to associate names used in the distributive stage with the correct gender. From the 400 participants, 17 did not correctly associate the correct gender. We discard these observations from the final analysis.

some amount of fair compensation (e.g., Egan, Matvos, and Seru, 2017; Sarsons, 2017). Alternatively, spectators behavior might solely be governed by their fairness concerns and compensation of the unlucky worker happens regardless of their gender (Charness and Rabin, 2002; Andreoni and Bernheim, 2009; Almas, Cappelen, Sorensen and Tungodden, 2010).

Hypothesis 2 (H.2) *Better-off workers retain the same amount from their winnings independent of their gender.*

By the same reasoning, stereotypical beliefs about how men and women ought to deal with risks may also lead better-off females to be considered less entitled to their gains from taking risks relative to better-off males. Alternatively, spectators might not take a better-off worker's gender into account when redistributing income to the worse-off worker.

Hypothesis 3 (H.3) *Spectators' implicit associations towards woman/men cannot predict their level of transfer.*

Personality and psychology literature suggests that implicit association tests have meaningful indications when investigating biases (Greenwald, Banaji, & Nosek, 2015). On the other hand, more recent findings from experimental economics literature suggest that implicit association tests may not explain biased redistribution decisions (Lee, 2018).

We explore the relationship between the gender of the risk-taker and spectators' redistribution decisions. Although we argue that the redistribution decisions can stem from stereotypical beliefs, our aim is not to determine which mechanisms drive these results. Consequently, our design cannot identify the mechanisms. Stereotypes are the potential mechanisms that can lead to these results. In this context, an implicit association test can tell us whether an experimental agent holds stereotypical beliefs. However, IAT does not help us to clarify whether stereotypes motivate redistribution decisions.

3 Results

We test hypothesis 1 by estimating the average treatment effect of changing the gender of the worse-off worker in the pair of workers on peoples' willingness to accept inequalities by measuring spectators decisions in two different ways. First, we measure the amount spectators wish to transfer to the worse-off worker and, second, we measure the level of inequality that would result from their transfer decision in the following way:[4]

$$\text{Inequality} = \frac{|\text{Income better-off } - \text{Income worse-off }|}{\text{Total Income}} \in [0, 1].$$

For our estimation we use ordinary least squares (OLS) regressions with robust standard errors and use the following specification:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 Female\ worse\text{-}off_i + \beta_2 Female_i \\ &+ \beta_3 Female\ worse\text{-}off \times Female_i + \gamma X_i + \varepsilon_i, \end{aligned} \tag{1}$$

where Y captures either the level of transfer or the level of inequality the spectator implements, $Female\ worse\text{-}off$ is an indicator variable taking the value of one if the worse-off worker in the pair is female, $Female$ is an indicator variable for when the spectator's gender is female and $Female\ worse\text{-}off \times Female$ is an interaction indicator variable for when the spectator is female and the worse-off worker in the pair is female as well, X_i is a vector of background characteristics for each worker i and ε_i is an error term.

Model I and IV in Table 3 report average treatment effect estimates for changing the gender of the worse-off worker in the pair from male to female. The estimate results reveal that average transfers to the worse-off worker as well as the level of inequality implemented is independent of the worse-off worker's gender ($p = 0.96$ and $p = 0.93$, respectively). Furthermore, model II and V show that while female spectators transfer more on average to the worse-off worker than male spectators, neither male or females condition their level of redistribution on the gender of the worse-off worker. This can be seen from the coefficient estimate for the $Female\ worse\text{-}off \times Female$ indicator variable that is not statistically different from zero ($p = 0.99$)

Model III and XI add a set of worker background variables and show the robustness of the results discussed so far. Background variables include age, education, household income and number of children in household. From the set of background variables we find that more educated workers transfer, on average, less to the worse-off worker and therefore implement a larger level of inequality ($p = 0.96$ and $p = 0.93$, respectively). The remaining background variables are not associated with the level of transfer.

In order to address whether the income the better-off workers retains is independent of their gender as stated in hypothesis 2, we estimate the average treatment effect of changing the gender of the better-off worker on both the amount spectators transfer from the better-off worker to the worse-off worker and the

[4]This inequality measure is equivalent to the Gini coefficient in the present set of distributive situations and takes a value of one if the spectator decides not to transfer anything and a value of zero if the spectator equalizes incomes. The inequality measure is unaffected by whether spectators implement inequality by transferring more or less than half of the total amount.

level of inequality. For estimating these effects, we again use ordinary least squares (OLS) regressions with robust standard errors using the following specification:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 Female\ better\text{-}off_i + \beta_2 Female_i \\ + \beta_3 Female\ better\text{-}off \times Female_i + \gamma X_i + \varepsilon_i, \end{aligned} \tag{2}$$

where Y captures either the level of transfer or the level of inequality the spectator implements, $Female\ better\text{-}off$ is an indicator variable taking the value of one if the better-off worker in the pair is female, $Female$ is an indicator variable for when the spectator is female and $Female\ better\text{-}off \times Female$ is an interaction indicator variable for when the spectator is female and the better-off worker in the pair is female as well, X_i is a vector of background characteristics for each worker i and ε_i is an error term.

Model I in Table 4 reports average treatment effect estimates for changing the gender of the better-off worker in the pair from male to female on the level of transfer. The estimate reveals that there is a seventeen percent larger albeit statistical insignificant transfer to the worse-off worker if the better-off worker in the pair is female rather than male ($p = 0.21$). Model IV shows that this change in the level of transfer corresponds to a change in the level of inequality of about eight percent. Similarly, this change also cannot be considered statistically significant ($p = 0.27$).

Model II and V add coefficients for the gender of the spectator as well as for the interaction effect of the gender of the better-off worker with the gender of the spectator. Estimation results from these models show that male spectators transfer more from the better-off worker to the worse-off worker if the better-off worker is female rather than male. Specifically, male spectators take 47% more from a better-off female worker than from a better-off male worker and transfer it to the worse-off worker ($p = 0.045$).[5] Females, while transferring a statistically larger amount from the better-off worker to the worse-off worker overall ($p = 0.018$), transfer an equal amount from the better-off worker to the worse-off worker independent of the better-off worker's gender. This can be seen from the linear combination of the $Female\ better\text{-}off + Female\ better\text{-}off \times Female$ coefficient estimate that is not statistically different from zero ($p = 0.648$).[6]

For hypothesis 3 we test whether implicit stereotypes towards male of females are associated with spectators transfer decisions in the experiment. To begin with, Figure 1 plots the distribution of D-scores from the implicit association test for male and female spectators. Using the standard cutoffs for classifying d-scores as "slight" ($|D| <= .15$), "moderate" ($.15 < |D| <= .35$) and "strong"

[5]There is no statistically significant change in the level of inequality associated with this effect.

[6]In Table S1 in the Appendix, we add an indicator variable to account for mixed gender groups. We find no qualitative difference in estimate results.

($.35 < |D| <= .65$), we find that 78.8% of male spectators and 67.48% show moderate or strong association towards either male or female stereotypes. Furthermore, we find that, on average, males exhibit a significantly larger implicit association towards male stereotypes than females ($diff = .269, p < 0.001$).

In order to estimate the correlation between a spectator's D-score obtained from the implicit association test and the transfer decision the spectator makes we use ordinary least squares (OLS) regressions with robust standard errors and use the following two specifications:

$$Y_i = \beta_0 + \beta_1 Pro\ better\text{-}off_i + \beta_2 Contra\ better\text{-}off_i \tag{3}$$

In the first specification above, Y captures either the level of transfer or the level of inequality the spectator implements, $Pro\ better\text{-}off$ is an indicator variable taking the value of one if the spectator exhibits a positive implicit association towards the better-off worker in the pair, $Contra\ better\text{-}off$ is an indicator variable taking the value of one if the spectator exhibits a negative implicit association against the better-off worker in the pair. To add to this analysis, we expand this specification further using the following variable setup:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 Pro\text{-}male_i \times Male\ better\text{-}off_i \\ &+ \beta_2 Pro\text{-}female_i \times Female\ better\text{-}off_i \\ &+ \beta_3 Pro\text{-}male_i \times Female\ better\text{-}off_i \\ &+ \beta_4 Pro\text{-}female_i \times Male\ better\text{-}off_i \end{aligned} \tag{4}$$

where Y captures either the level of transfer or the level of inequality the spectator implements, $Pro\text{-}male \times Male\ better\text{-}off$ is an indicator variable for spectators who exhibit a positive implicit association towards males and the better-off worker in the pair is male, $Pro\text{-}female \times Female\ better\text{-}off$ is an indicator variable for spectators who exhibit a positive implicit association towards females and the better-off worker in the pair is female, $Pro\text{-}male \times Female\ better\text{-}off$ is an indicator variable for spectators who exhibit a positive implicit association towards males and the better-off worker in the pair is female and $Pro\text{-}female \times Male\ better\text{-}off$ is an indicator variable for when the spectator exhibits a positive implicit association towards females and the better-off worker in the pair is male.

Table 5 shows estimation results for specification 3 and 4. Overall, none of the coefficient estimates are statistically significant suggesting that the variation in IAT scores is not statistically related to variation in spectator decisions.

To sum up, we fail to reject any null hypotheses. For Hypothesis 1 we show that average transfers from the worse-off worker to the better-off worker are not dependent on the worse-off worker's gender. This finding also supports that the level of inequality created between the two workers is gender independent.

There is also no statistically significant evidence regarding the compensation and level of inequality between a better-off and worse-off worker for the second hypothesis. However, estimations with spectator's gender show that male spectators redistribute the earnings of a better-off female worker more than a better-off male worker. In contrast, female spectators provide an equal distribution between a better-off and a worse-off worker, regardless of workers' gender.

For the third hypothesis we find no evidence to explain spectators' distribution behavior with the implicit biases. However, we gain insights into the implicit associations of the spectators. Male spectators have larger implicit associations to male stereotypes compared to female spectators.

4 Conclusion

In different environments, women might appear to be more risk-averse than men. Substantially, the general public has this view when evaluating risk-taking decisions from men and women. People mostly have accurate predictions about the gender gap in risk tolerance, while people fail to predict the actual difference.

Consequently, risk-taking behaviors from men and women might be subject to different evaluations due to this stereotypical view on risk tolerance. As findings from the literature show, the adverse outcomes of women's risky decisions may not be subject to a comparative evaluation as in males' decisions.

We explore whether people hold women more responsible for risk-taking decisions than men with a novel survey experiment. Including participants from the United States and utilizing a short vignette, we described a decision-making situation of two workers who engage in the same individual effort task. The only experimental agents, the spectators, are recruited from an online crowdsourcing platform of Amazon Mechanical Turk. These agents are asked to make distribution decisions about the earnings of two hypothetical workers. We find no evidence that the gender of unlucky risk-takers affect how they are compensated for their loss. However, our findings show that male spectators are less willing to redistribute earnings from lucky workers to unlucky workers if they are male rather than female. However, male spectators tend to leave unlucky male workers with no earnings compared to female workers. We further explore this discriminatory behavior through the lens of implicit and explicit biases. The results show that male spectators' discriminatory behavior may not be explained with either of those.

This paper builds the ground for future research. Firstly, it would be interesting to study a real effort task where the risk-taking decisions impact workers' actual earnings. This alternative study would provide an outlet to understand whether gender bias changes direction in the face of real consequences. Secondly, exploring the decision making in the leadership domain could provide an opportunity to compare the biases towards women from the general public and female leaders. Thirdly, the same design could be applied to a framework where men were benevolent (i.e., nonstereotypical). Fourthly, although not a common practice in the behavioral economics literature, as an alternative to personality inventories and implicit scales, written tasks would be used to explore participants' beliefs (e.g., Banker, Bhanot, & Deshpande, 2020).

References

Almås, I., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science, 328* (5982), 1176-1178.

Almenberg, J., & Dreber, A. (2015). "Gender, stock market participation and financial literacy," *Economics Letters, 137*: 140 – 142.

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77*(5), 1607-1636.

Anseel, F., & Duyck, W. (2008). Unconscious applicants: A systematic test of the name-letter effect. *Psychological Science*, *19*(10), 1059-1061.

Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles, 54*(7-8), 459-467.

Ball, S., Eckel, C. C., & Heracleous, M. (2010). Risk aversion and physical prowess: Prediction, choice and bias. *Journal of Risk and Uncertainty*, *41*(3), 167-193.

Banker, S., Bhanot, S. P., & Deshpande, A. (2020). Poverty identity and preference for challenge: Evidence from the US and India. *Journal of Economic Psychology, 76*, 102214.

Bechler, C., Green, L., & Myerson, J. (2015). Proportion offered in the Dictator and Ultimatum Games decreases with amount and social distance. *Behavioural processes, 115*, 149-155.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of consulting and clinical psychology, 42*(2), 155.

Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology, 25*(5), 581-589.

Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *The Journal of Socio-Economics, 37*(5), 1775-1784.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, *94*(4), 991-1013.

Boverie, P. E., Scheuffele, D. J., & Raymond, E. L. (1994). Multimethodological approach to examining risk-taking. *Current Psychology, 13*(4), 289-302.

Bühren, C., & Kundt, T. C. (2015). Imagine being a nice guy: A note on hypothetical vs. incentivized social preferences. *Judgment & Decision Making, 10*(2).

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409-1447.

Cappelen, A. W., Falch, R., & Tungodden, B. (2019). *The boy crisis: Experimental evidence on the acceptance of males falling behind.* NHH Dept. of Economics Discussion Paper, (06). http://hdl.handle.net/11250/2589254

Cappelen, A. W., Konow, J., Sørensen, E. Ø., & Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review, 103*(4), 1398-1413.

Carpenter, T. P., Pullig, C. P., Pogacar, R., Kouril, M., LaBouff, J., & Isenberg, N. (2017). Measuring implicit cognition in Qualtrics with iatgen: A free, user-friendly tool for building survey-based IATs. In *Annual Meeting of the Society for Personality and Social Psychology.*Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization, 90*, 123-133.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50-58.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics, 117*(3), 817-869.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448-74.

Daruvala, D. (2007). Gender, risk and stereotypes. *Journal of Risk and Uncertainty*, *35*(3), 265-283.

De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266-3302.

Eckel, C. C., & Grossman, P. J. (2002). "Sex differences and statistical stereotyping in attitudes toward financial risk," *Evolution and Human Behavior, 23*(4): 281 – 295.

Eckel, C. C., & Grossman, P. J. (2008a). Differences in the economic decisions of men and women: Experimental evidence. In C, R. Plott & V. L. Smith (Eds), *Handbook of experimental economics results* (1st ed., pp. 509– 519). Elsevier.

Eckel, C. C., & Grossman, P. J. (2008b). "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices," *Journal of Economic Behavior & Organization, 68*(1): 1–17.

Egan, M. L., Matvos, G., & Seru, A. (2017). *When Harry fired Sally: The double standard in punishing misconduct* (No. w23242). National Bureau of Economic Research.

Engel, C., (2011). Dictator Games: A Meta Study. *Experimental Economics 14*(4), 583-610.

Filippin, A., & Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science, 62*(11), 3138-3160.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically smalleffects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561. http://dx.doi.org/10.1037/pspa0000016

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology, 74*(6), 1464.

Grossman, P. J. (2013). Holding fast: the persistence and dominance of gender stereotypes. *Economic Inquiry*, *51*(1), 747-763.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, *14*(3), 399-425.

Hsee, C. K., & Weber, E. U. (1997). A fundamental prediction error: self–others discrepancies in risk preference. *Journal of experimental psychology: general*, *126*(1), 45.

Lee, D. J. (2018). Does implicit bias predict dictator giving? *Games, 9*(4), 73.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives, 21*(2), 153-174.Kamas, L., & Preston, A. (2015). Can social preferences explain gender differences in economic behavior? *Journal of Economic Behavior & Organization, 116*, 525-539.

Konow, J. (2009). Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare, 33*(1), 101-127.

Konow, J., Saijo, T., & Akai, K. (2020). Equity versus equality: Spectators, stakeholders and groups. *Journal of Economic Psychology, 77,* 102171.

Marinelli, N., Mazzoli, C., & Palmucci, F. (2017). How does gender really affect investment behavior?. *Economics Letters*, *151*, 58-61.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27*(9), 1164-1178.

Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, *106*(36), 15268-15273.

Sarsons, H. (2017). Interpreting signals in the labor market: evidence from medical referrals. https://scholar.harvard.edu/files/sarsons/files/sarsons_jmp.pdf

Schroth, M. L. (1996). Scores on sensation seeking as a predictor of sexual activities among homosexuals. *Perceptual and motor skills, 82*(2), 657-658.

Sharma, S. (2015). Gender and distributional preferences: Experimental evidence from India. *Journal of Economic Psychology, 50*, 113-123.

Siegrist, M., Cvetkovich, G., & Gutscher, H. (2002). Risk preference predictions and gender stereotypes. *Organizational Behavior and Human Decision Processes*, *87*(1), 91-102.

Social Security Administration (2021). *Top Names Over the Last 100 Years.* Accessed online https://www.ssa.gov/oact/babynames/decades/century.html

Spigner, C., Hawkins, W. E., & Loren, W. (1993). Gender differences in perception of risk associated with alcohol and drug use among college students. *Women & health, 20*(1), 87-97.

Table 1: Treatment table

	Better-off	
Worse-off	Male	Female
Male	Male vs. Male	Female vs. Male
Female	Male vs. Female	Female vs. Female

Note: The table gives an overview of the experimental design and shows the gender composition of the worker pair that participants were presented with in the re distributional situation.

Table 2: Background characteristics of subjects

	Age	**Female**	**Education**	**Income**	**Political**	
Better-off vs. worse-off	**Mean (se)**	**Mean (se)**	**Mean (se)**	**Mean (se)**	**Mean (se)**	**N**
Male vs. male	37.04 (1.07)	0.46 (0.05)	4.27 (0.13)	4.96 (0.22)	-0.31 (0.12)	94
Male vs. female	37.70 (1.10)	0.41 (0.05)	4.39 (0.16)	4.81 (0.23)	-0.31 (0.14)	96
Female vs. male	37.65 (1.12)	0.46 (0.05)	4.25 (0.16)	4.78 (0.26)	-0.43 (0.12)	93
Female vs. female	38.09 (1.01)	0.50 (0.05)	4.39 (0.14)	4.89 (0.25)	-0.20 (0.12)	100
All	37.61 (0.54)	0.46 (0.03)	4.32 (0.07)	4.86 (0.12)	-0.31 (0.01)	383

Note: The table reports background characteristics of subjects participating in the experiment. Subjects were recruited through the Amazon Mechanical Turk crowd-sourcing platform. “Age” is a continuous variable measuring participants’ age in years;“Female” captures the proportion of females;“Education” is an ordinally scaled variable: 1 = High School’, 2 = Some College, 3 = 2 year College Degree, 4 = 4 year College Degree, 5 = Masters Degree, 6 = Doctoral Degree;.

Table 3: Transfers conditional on the gender of the worse-off worker in the pair.

Dependent variable:	Transfer			Inequality		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Female worse-off	-0.059	0.209	0.086	-0.004	-0.015	-0.011
	(1.455)	(1.994)	(1.997)	(0.046)	(0.060)	(0.060)
Female		2.670	2.102		-0.123*	-0.105
		(2.009)	(2.109)		(0.063)	(0.065)
Female worse-off × Female		-0.562	-0.046		0.023	0.009
		(2.913)	(2.965)		(0.092)	(0.094)
Constant	11.556***	10.330***	10.097***	0.637***	0.693***	0.672***
	(1.006)	(1.386)	(1.652)	(0.031)	(0.041)	(0.051)
Controls	No	No	Yes	No	No	Yes
N	383	383	383	383	383	383
R^2	0.000	0.007	0.018	0.000	0.016	0.026

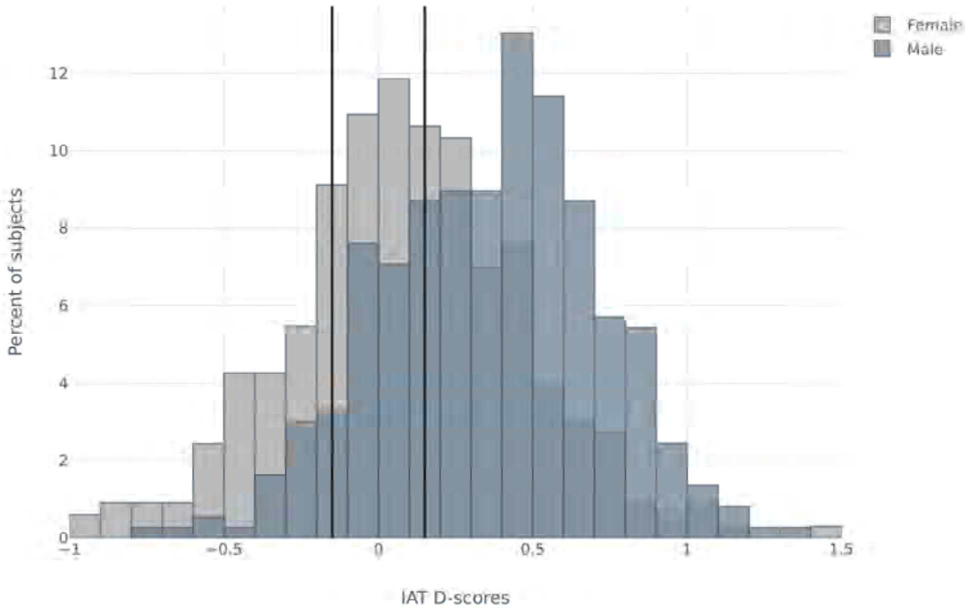
Note: The table reports linear regressions of the variable "Transfer" (columns (I)–(III)), capturing the amount the spectators assigns to the worse-off worker and of the variable "Inequality" (columns (IV)–(VI)), measuring the inequality implemented by the spectator, "Female worse-off": indicator variable for the worse-off worker being female, "Female": indicator variable for the spectator being female, Controls include "High age": indicator variable for the spectator's age being at or above the median in the sample, "High education": indicator variable for the spectator's education being at or above the median in the sample, "High income": indicator variable for the spectator having above median income. Robust standard errors in parentheses. ($*$: $p < 0.1$, $**$: $p < 0.05$, $***$: $p < 0.01$)

Table 4: Transfers conditional on the gender of the better-off worker in the pair.

Dependent variable:	Transfer			Inequality		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Female better-off	1.826	3.993**	4.077**	-0.050	-0.107*	-0.107*
	(1.451)	(1.989)	(2.003)	(0.046)	(0.059)	(0.060)
Female		4.801**	4.652**		-0.177***	-0.173***
		(2.027)	(2.088)		(0.065)	(0.067)
Female better-off		-4.964*	-5.414*		0.136	0.145
× Female		(2.912)	(2.967)		(0.092)	(0.095)
Constant	10.622***	8.532***	9.048***	0.660***	0.737***	0.713***
	(1.006)	(1.270)	(1.523)	(0.032)	(0.039)	(0.049)
Female +		-0.971	-1.337		0.028	0.037
Female better-off × Female		(2.127)	(2.193)		(0.071)	(0.072)
Controls	No	No	Yes	No	No	Yes
N	383	383	383	383	383	383
R^2	0.004	0.018	0.030	0.003	0.024	0.034

Note: The table reports linear regressions of the variable "Transfer" (columns (I)–(III)), capturing the amount the spectators assigns to the worse-off worker and of the variable "Inequality" (columns (IV)–(VI)), measuring the inequality implemented by the spectator, "Female better-off": indicator variable for the better-off worker being female, "Female": indicator variable for the spectator being female, Controls include "High age": indicator variable for the spectator's age being at or above the median in the sample, "High education": indicator variable for the spectator's education being at or above the median in the sample, "High income": indicator variable for the spectator having above median income. Robust standard errors in parentheses. ($*$: $p < 0.1$, $**$: $p < 0.05$, $***$: $p < 0.01$)

Figure 1: Histograms of transfers to the unlucky worker in the experiment



Note: The figure shows the distribution of standardized test score from the Implicit Association Test (IAT) separately for male and female spectators. The standard cutoffs for classifying D-scores as "slight" ($|D| <= .15$) are indicated as well.

Table 5: Transfers to the better-off worker in the pair depending of IAT classification.

Dependent variable:	Transfer		Inequality	
	(I)	(II)	(III)	(IV)
Pro better-off	0.114		-0.004	
	(2.534)		(0.084)	
Pro-male & male better-off		-0.000		0.000
		(2.696)		(0.090)
Pro-female & female better-off		0.642		-0.021
		(4.293)		(0.143)
Contra better-off	1.240		-0.019	
	(2.603)		(0.084)	
Pro male & female better-off		1.477		-0.022
		(2.776)		(0.089)
Pro female & male better-off		0.187		-0.006
		(4.525)		(0.151)
Constant	11.176***	11.176***	0.627***	0.627***
	(1.837)	(1.848)	(0.061)	(0.062)
N	173	173	173	173
R^2	0.002	0.002	0.000	0.001

Note: The table reports linear regressions of the variable "Transfer" (columns (I)–(II)), capturing the amount the spectators assigns to the worse-off worker and of the variable "Inequality" (columns (III)–(IV)), measuring the inequality implemented by the spectator, "Pro better-off": indicator variable if the spectator has an implicit bias towards the better-off. "Pro-male & male better-off": indicator variable if the spectator has an implicit bias towards males and the better-off worker is a male. "Pro-female & female better-off": indicator variable if the spectator has an implicit bias towards females and the better-off worker is a female. "Contra better-off": indicator variable if the spectator has an implicit bias towards the better-off. "Pro-male & female better-off": indicator variable if the spectator has an implicit bias towards males and the better-off worker is a female. "Pro-female & male better-off": indicator variable if the spectator has an implicit bias towards females and the better-off worker is a male. Robust standard errors in parentheses. ($*$: $p < 0.1$, $**$: $p < 0.05$, $***$: $p < 0.01$)

Appendix

Instructions

We now ask you to make a decision based on a hypothetical situation.

John and Ann were recruited to work on a four hours long task. Before they started, John and Ann were told that they would be paid a compensation of 10 USD.

After completing the task, John and Ann were told that in addition to the compensation, they were asked to choose a bonus earnings plan. There were two plans.

The first plan was to pay them a bonus of 20 USD with certainty. The second plan was to pay them a bonus based on a lottery draw.

In the lottery, it was possible to earn either 60 USD or 0 USD with equal chance.

Ann chose the lottery, and received an additional bonus earnings of 60 USD. John chose the lottery as well, but John was unlucky and earned 0 USD.

Ann and John were not informed about their final payment. However, they were told that a third person would be informed about the bonus earnings and decisions of each worker.

Ann and John were also told that this third person was given the opportunity to redistribute the bonus earnings between Ann and John, and determine the final payment of each worker.

This third person is you and we want you to decide whether to redistribute the earnings of Ann and John. Your decision is anonymous.

Ann chose the lottery and was lucky; she earned 60 USD. John chose the lottery and was unlucky; he earned 0 USD.

Do you prefer to transfer any share of 60 USD from Ann to John?

- I prefer not to transfer
- I prefer to transfer
  - Ann is paid 55 USD and John is paid 5 USD.

- Ann is paid 50 USD and John is paid 10 USD.
- Ann is paid 45 USD and John is paid 15 USD.
- Ann is paid 40 USD and John is paid 20 USD.
- Ann is paid 35 USD and John is paid 25 USD.
- Ann is paid 30 USD and John is paid 30 USD.
- Ann is paid 25 USD and John is paid 35 USD.
- Ann is paid 20 USD and John is paid 40 USD.
- Ann is paid 15 USD and John is paid 45 USD.
- Ann is paid 10 USD and John is paid 50 USD.
- Ann is paid 5 USD and John is paid 55 USD.
- Ann is paid 0 USD and John is paid 60 USD.

Additional tables and figures

Table S1: Transfers conditional on the gender of the better-off worker in the pair. Controlling for mixed groups.

Dependent variable:	Transfer			Inequality		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Female better-off	1.835	3.999**	4.042**	-0.050	-0.107*	-0.108*
	(1.460)	(2.006)	(2.022)	(0.046)	(0.060)	(0.061)
Female		4.787**	4.669**		-0.178***	-0.173**
		(2.043)	(2.097)		(0.065)	(0.067)
Female better-off		-4.961*	-5.379*		0.136	0.148
× Female		(2.935)	(3.006)		(0.093)	(0.095)
Mixed group	-0.395	-0.274	-0.251	-0.007	-0.013	-0.014
	(1.458)	(1.457)	(1.463)	(0.046)	(0.046)	(0.046)
Constant	10.812***	8.671***	9.238***	0.663***	0.744***	0.725***
	(1.258)	(1.479)	(1.688)	(0.039)	(0.043)	(0.051)
Female +		-0.962	-1.337		0.029	0.039
Female better-off × Female		(2.145)	(2.209)		(0.071)	(0.073)
Controls	No	No	Yes	No	No	Yes
N	383	383	383	383	383	383
R^2	0.004	0.018	0.030	0.003	0.024	0.034

Note: The table reports linear regressions of the variable "Transfer" (columns (I)–(III), capturing the amount the spectators assigns to the better-off worker) and of the variable "Inequality" (columns (IV)–(VI), measuring the inequality implemented by the spectator), "Female better-off": indicator variable for the better-off worker being female. "Female": indicator variable for the spectator being female. "Mixed group": indicator variable for mixed gender groups. Controls include "High age": indicator variable for the spectator's age being at or above the median in the sample, "High education": indicator variable for the spectator's education being at or above the median in the sample, "High income": indicator variable for the spectator having above median income. Robust standard errors in parentheses. ($*$: $p < 0.1$, $**$: $p < 0.05$, $***$: $p < 0.01$)

Verbal Impression Management Strategies by Top Female Executives

Nur Yaldiz and Yuko Onozaka

ABSTRACT

Creating and managing their public image is essential for leaders. From the role congruity perspective, female executives face a particular challenge; that they are disliked when following the societal image of agentic leaders (confident and assertive) but judged as ineffective when following female communal norms (benevolent and accommodating). However, there are reasons to believe that displaying communal qualities could be beneficial by enabling female leaders to avoid backlash, heightening the perception of competence (succeeded *despite* the gender obstacles), and establishing congruence to the new and contemporary androgynous leader image. This study empirically examines the spoken and written texts of 204 female and 318 male executives collected from various media outlets to reveal the images top female executives convey in public through communal and agentic language styles. Our findings show that female executives employ more communal language than male executives, but this effect is moderated by the industry sector such that executives in female-dominated industrial sectors tend to have a more agentic style. Female executives have a more agentic style than the general female language, while male executives exhibit more communal language than the general male language. We also find that female executives speak more about family and motherhood, irrespective of interview and non-interview occasions. This provides an additional contextual insight that family obligations generally being seen as obstacles for upward ascension for women in the labor market does not apply to the top executive levels. The overall findings are consistent with the notion that displaying communion benefits top executives.

Keywords: gender; linguistic styles; impression management; leadership; text analysis

INTRODUCTION

Creating and managing public image is essential for leaders, both to convey an image of the company to the outside world (Rosenfeld, Giacalone, & Riordan, 1995) and as an effective organizational management tool (e.g., Brotheridge, Lee, Riggio, & Reichard, 2008; Fombrun & Shanley, 1990; Sosik, Avolio, & Jung, 2002; Tucker, Turner, Barling, Reid, & Elving, 2006). According to the role congruency (Eagly & Karau, 2002), society expects its leaders to fit the descriptive norms of leaders (Eagly, 1987; Eagly & Karau, 2002), such as being assertive, competitive, and risk-taking (Eagly & Carli, 2007). These agentic characteristics are strongly associated with male descriptive norms, making it relatively straightforward for men to adopt such leadership qualities. Women, too, are expected to demonstrate their agentic qualities to appear competent, but being agentic can also invoke backlash (Rudman, Moss-Racusin, Phelan, & Nauts, 2012). The incongruity between communal female gender norms and agentic leaders poses a primary challenge for women to convey an appropriate image (Connor & Fiske, 2018; Eagly & Karau, 2002; Rudman & Phelan, 2008).

However, the literature also suggests that displaying communion can benefit female executives. From the role congruity standpoint, women can avoid backlash by confirming the gender prescription and appearing more communal (Heilman & Okimoto, 2007; Rudman et al., 2012). Although this can risk female leaders appearing incompetent, the stereotype content model suggests that women at the top leadership levels can appear both competent and communal (Eckes, 2002; Fiske, Cuddy, Glick, & Xu, 2002). The perceived competence of top female leaders can be high because of the very environment that makes it more challenging for women to ascend on the organizational hierarchy—the women who actually made it to the top *despite* these hindrances are seen as exceptionally competent (Rosette & Tost, 2010). Taken together, female executives can appear communal without being seen as weak or receiving backlash (Eckes, 2002; Fiske et al., 2002).

Furthermore, the image of "ideal" leaders is shifting from traditional agentic leaders to transformational leaders (Koenig, Eagly, Mitchell, & Ristikari, 2011). In particular, transformational leadership comprises four main components; idealized influence, individualized consideration, inspirational motivation, and intellectual stimulation (Burns, 1978). Since these components are generally associated with communal features, female leaders are seen as a better fit for transformational leadership than male leaders (Bass, 1990). This leadership style may also fit better with today's organizations (Ng, 2017). Therefore, this modern transformational leadership image is congruent with female communal qualities.

The display of communal qualities harms or benefits female leaders can depend on specific contexts (e.g., Cabrera, Sauer, & Thomas-Hunt, 2009). The hierarchical level is one of such contexts. Appearing communal can benefit female executives (Pillemer, Graham, & Burke, 2014; Rosette & Tost, 2010) but is linked to less adequate leadership for women in other levels of hierarchy (e.g., Cuddy, Fiske, & Glick, 2004; Smith, Rosenstein, Nikolov, & Chaney, 2019). Organizational contexts (Eagly & Carli, 2003; Paustian-Underdahl, Walker, & Woehr, 2014), industry-specific culture, and a heterogeneous degree of female representation (Dezsö & Ross, 2012; Ko, Kotrba, & Roebuck, 2015; Lemoine & Blum, 2021) are also likely to influence female executives' agentic/communal positioning strategies. For instance, more stereotypically male-dominated organizations tend to value a masculine leadership style (Kark, Waismel-Manor, & Shamir, 2012).

Effectively conveying the targeted communal/agentic images is important for top executives. The literature provides insights how they do this behaviorally, such as being assertive for the benefit of others during negotiations (Amanatullah & Morris, 2010), showing benevolence, and communicating information on family roles while engaging in stereotypically male-dominated tasks (Heilman & Okimoto, 2007), and displaying implicit dominance rather than being direct and demanding (Williams & Tiedens, 2016). However, we do not know how they use verbal communication to achieve this goal. This study focuses on top female executives and fills this knowledge gap. We analyze publicly available texts (e.g., interviews, speeches, podcasts, and editorial pieces) from 204 female and 318 male prominent executives with over one million words spoken or written by them. Investigating impression management strategies is difficult, as some view them too person- or situation-specific. Many existing studies may focus on an individual or a relatively small number of leaders (e.g., Choudhury, Wang, Carlson, & Khanna, 2019; Jones, 2016; Pennebaker & Lay, 2002) due to such challenge. This study aims to provide new and general insights by studying over 500 executives. Empirically, we quantify and compare the degree of communal and agentic languages of executives using the socio-linguistic approach. As the literature suggests the importance of contextual factors, our investigation includes individual and industry characteristics. We provide further anchoring in female executives' positioning of linguistic styles through the comparisons to those of general male and female language reported by Newman, Groom, Handelman, and Pennebaker (2008). Finally, we investigate both the most frequently used and the most gender-differentiating set of words to shed light on specific ways female executives verbally display communion and offer theoretical discussion.

This study contributes multiple ways to the interdisciplinary social psychology and management literature that focuses on women's strategies against backlash in non-stereotypical settings. First, we find that female executives use communal language (more personable and emotional styles) and communicate communal topics (e.g., family and motherhood). This finding is consistent with the current literature that women in non-stereotypical domains tend to soften their image with gender congruent features, and such communal display could benefit top female leaders. In addition, comparisons with the general public also support that male and female leaders have androgynous styles. Our findings also support the developing literature on the importance of leader androgyny and the shift in leader stereotypes.

We also add to the literature exploring the impact of context on leaders' self-presentation. Our findings complement the substantial literature by showing that female-dominant environments allow women to display non-stereotypical characteristics. We provide an additional contextual insight that family obligations, generally being seen as obstacles for upward ascension for women in the labor market, does not apply to the top executive levels. Finally, we methodologically contribute to the emerging interdisciplinary literature on social science, quantitative text analysis, and big data.

The remainder of the paper is organized as follows. The theoretical background section discusses the relevant findings from the literature of sociology, linguistics, management, and psychology. After that, we introduce hypotheses. The methodology section describes the main elements of quantitative text and exploratory content analyses, followed by the results section. Finally, we provide the main implications and insights of this study for management and leadership literature in the discussion and conclusions section.

THEORETICAL BACKGROUND

Role and Status Congruity

Gender is a social and cognitive construct considered the accumulation of socially significant characteristics (Larsen & Seidman, 1986). Society shares collective expectations, or norms, on how men and women are (*descriptive* norms) and what they should do (*prescriptive* or *injunctive* norms). A critical observation in this context is that typical attributes that men and women are associated with converge into two fundamental dimensions in social cognition, *agency,* and *communion* where the former constitutes goal-achievement and task functioning, while the latter relates to the maintenance of relationships and social functioning (Abele & Wojciszke, 2014; Bakan, 1966). Agentic qualities relate to assertiveness and confidence and are usually strongly linked to men, whereas communal qualities comprise caring and other-regarding, typically associated with women. The Social Role Theory (Eagly, 1987; Eagly, Wood, & Diekman, 2000) postulates that gender role expectations are derived from both descriptive and prescriptive norms that are tied to an agency (for men) and communion (for women), exerting a powerful influence on how people behave. Society either approve the appropriate behavior or sanction the deviations to reinforce the gender roles. For instance, deviations from descriptive norms invoke emotional responses such as surprise, and deviations from prescriptive norms are met with moral disapproval (Cialdini & Trost, 1998).

The agency and communion and associated gender norms influence the social standing of men and women, as an agency is linked with higher status, wherein communion with lower status (Abele & Wojciszke, 2014; Bakan, 1966). This leads to the view that women, who are prescribed to be communal and lower status, are unfit to lead. When women are in leadership roles, they invoke a mix of positive and negative reactions (or ambivalent reactions), which tend to lower female leaders' performance evaluation (Toneva, Heilman, & Pierre, 2020). These effects caused by the mismatch of female and leader traits are theorized in Eagly and Karau (2002) as the role congruity theory, an extension to the Social Role Theory. The Role Congruity Theory explains what is referred to as a double-bind. Women face two challenges when they wish to ascend to leadership roles. First, they need to fill the gap between the perceptions for leaders and those for women and demonstrate their competency. Second, even after they succeed in filling the perception gap, they face the backlash as they are regarded as less likable, hirable, promotable (Rudman & Glick, 2001), and influential (Carli, 2001) than men with similar credentials. The backlash towards agentic women is further underpinned as a penalty given to women who engage in high-status behavior. More precisely, when women exhibit dominant traits (e.g., controlling, arrogant and ambitious), they are not only behaving as women should *not* (proscriptive norms) but also acting as high-status, thereby challenging the existing gender hierarchy. This creates status incongruency and invokes dominance penalty by perceivers motivated to maintain the current social hierarchy (Rudman et al., 2012; Williams & Tiedens, 2016).

However, the existing studies also point out that women avoid backlash when they show communal qualities (Rosette & Tost, 2010; Vroman & Danko, 2020). From the role and status congruence perspectives, when women appear communal, which is also lower social status, this incurs less incongruency and, therefore, less penalty (Rudman et al., 2012). It is also argued that people (particularly men) are more willing to accept influence from women when they appear communal and likable (Carli, 2001). Thus, with the incongruence between leader and female norms, one strategy for women to exert leadership is to appear communal.

Female Leadership Advantage

The role congruity research shows that women are not considered fitting for leadership roles due to the mismatch between the expected leader behavior and the prescribed female roles. Women tend to be evaluated with lower performance ratings if they are communal and not considered for senior levels (Cuddy et al., 2004) and meet backlash when they ascend to leadership positions (Rudman & Glick, 1999; 2001). However, the very existence of the obstacles that women face as they climb up the organizational hierarchy can strengthen the perception of competence of these women, as they achieve the higher position *despite* the hindrances (Rosette & Tost, 2010). In addition, for female executives at the top levels only a selected few can achieve, these women are not in direct competition with most people in the organization. This can reduce the perception of competitiveness and increase the perception of warmth and communality (Eckes, 2002; Fiske et al., 2002). Thus, top female leaders can be considered competent *and* communal simultaneously. Accordingly, it is argued that top female leaders can simultaneously be evaluated as agentic *and* communal (Rosette & Tost, 2010).

The rise of transformational leadership also supports female leadership advantage (Koenig et al., 2011). Organizations realize that heavily agentic leadership traits may not match the needs of today's modern workplace or aid organizational performance in the long run (Ng, 2017). As such, today's leadership skills increasingly emphasize androgynous characteristics – a mixture of agency and communal qualities (Koenig et al., 2011; Pillemer et al., 2014). For instance, idealized influence, individualized consideration, inspirational, motivational, and intellectual stimulation are essential leadership skills (Bass, 1990) that require people-oriented communal qualities (Kark et al., 2012; Koenig et al., 2011). Due to the match with communal characteristics, female leaders generally perform transformational leadership better (Bass 1990; Eagly, Johannesen-Schmidt, & Van Engen, 2003; Powell, Bitterfield, & Bartol 2008). Still, this does not mean that agency is no longer relevant for leadership. For instance, inspiring and motivating followers for a common goal requires assertiveness (Vinkenburg, Van Engen, Eagly, & Johannesen-Schmidt, 2011), a typical agentic quality. What seems critical in today's leaders is the appropriate balance between agency and communion.

Moderating Factors

There seem to be specific contexts where communal display works for women, as discussed above. The hierarchical level (top leadership vs. mid/low level) is one of them, supported by the stereotype content model. Women in senior executive positions, who have already proved their competence, can benefit from displaying communal characteristics, i.e., double standards work in favor of female senior executives. By achieving top leadership positions, these women demonstrate strong competence. After establishing this capability, communal characteristics can boost women's image as exceptional leaders. However, similar advantages may not be possible in junior and mid-level management. Motherhood and display of communal characteristics may hamper women's success and chances for promotion (e.g., Connor & Fiske, 2018; Cuddy et al., 2004).

Another possible factor is the gender composition of the group. In female-dominated and mixed-sex settings, gendered expectations are less likely to restrict women's ascend. Female leaders in these contexts are seen as more competent than their female counterparts in stereotypically male-dominated environments (Ko et al., 2015). Furthermore, followers may expect female leaders to display a certain degree of agency with communion (Kark et al., 2012). Conversely, communal characteristics can be more beneficial when leaders share similar traits with their followers and when communion matches the organization's purpose (Lemoine & Blum, 2021).

Impression Management of Leaders

The role and status (in)congruence indicate that women are severely disadvantaged to be considered and promoted to top leadership roles. However, there are compelling arguments why women should display their communal characteristics, especially at the top level of leadership, to avoid backlash, appear competent, and be congruent to the new and contemporary leader image.

This poses a question: what image should top female executives convey as their impression management strategy?

Self-presentation is a process of influencing the impression of a person, a social entity, or an opinion (Goffman, 1959). The critical factor in this process is mediating the relevant information to influence the image created in the minds of others. The concept of impression management through self-presentation has since become a core practice of corporate leaders at the personal and organizational levels (Schlenker, 2012). For corporate leaders, self-image also serves as a tool to convey the targeted image of the company to the outside world (Rosenfeld et al., 1995). Indeed, creating suitable leadership images is a highly effective organizational management tool (e.g., Brotheridge et al., 2008; Fombrun & Shanley, 1990; Sosik et al., 2002; Tucker et al., 2006). The image of corporate leaders is not only self-serving or limited to inter-organization environments, as it extends beyond intra-organizational boundaries. For instance, the background, competencies, and lifestyle of a CEO feed into the organization's image (Love, Lim, & Bednar, 2017; Pollach & Kerbler, 2011).

The focus of the current study is the careful and deliberate use of language to manipulate or reinforce the image of leaders (Pan, McNamara, Lee, Haleblian, & Devers, 2018). One prominent example is how Hillary Clinton changed her language style to fit into the formal role she was portraying at different career stages. According to Jones (2016), Clinton employed a more masculine linguistic style to conform to stereotypical expectations of a political career. Conversely, when Clinton tried to create a more approachable image with a feminine linguistic style, the 2008 campaign strategy fluctuated with the image crisis. In this case, the shift from masculine and feminine linguistic styles failed to create an authentic and reliable leader image.

The socio-linguistic literature provides insights into how societal gender norms, with agency-men communion-women cognition, shape the language of men and women. For instance, agency relates to goal-achievement and task functioning, consistent with the male language that focuses on facts, objects, and orientation. Women's language tends to focus more on narratives and social relations, reflecting how communion relates to maintaining relationships and social functioning. As men historically occupy leadership positions, it is not surprising that male language overlaps with leadership language (Jones, 2016; Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014). Similarly, reflecting on women's historically lower social status, female language styles are associated with subordinates or people from lower social ranks (Coates, 2015). Employing male language can draw an agentic image for women, but the incongruence can result in appearing cold, arrogant, not authentic, and not likable (Jones, 2016).

HYPOTHESES

The incongruity between stereotypical leader image and female image (Eagly & Karau, 2002) undermines a woman's success as an effective leader (Smith et al., 2019; Cuddy et al., 2004; Rudman & Phelan, 2008; Spencer, Logel, & Davies, 2016). However, there are reasons to believe that displaying communal qualities benefit top female leaders. More specifically, appearing more communal enables her to avoid backlash and dislike (Rudman et al., 2012), improve the perception of competence and legitimacy (Rosette & Tost, 2010), and strengthen the congruence with transformational leadership style (Eagly et al., 2003). At the same time, the general acceptance and desire for the transformational leadership style can also motivate top male leaders to appear more communal, implying that male and female executive language styles could be similar. To investigate this, we hypothesize that top female executives use more of a communal language and less of an agentic language than male executives:

Hypothesis 1a. Female executives employ a more communal linguistic style than male executives.

Hypothesis 1b. Female executives employ a less agentic linguistic style than male executives.

A female leader's language style may also vary depending on the sector to which she belongs. Pennebaker (2011) argues that the environment influences the linguistic style of an individual. For example, the technology industry is known for its dominant male culture, so female leaders in this sector may require more effort to fit into such a culture. Furthermore, fewer female senior executives indicate stricter adherence to the traditional gender norms. Hence, female executives in this sector see their communal qualities as a disadvantage, but at the same time, the display of agency may evoke strong dislike. In male-dominated industrial sectors, polarity can explain disklike toward nonstereotypical women. Agentic traits are specifically assigned to men, while communal traits are strictly associated with women. There is also no fluidity between agency and communion for both sexes. In other words, female leaders may not have the freedom to display agentic traits when necessary (Ely, 1995). Consequently, not being direct and assertive may lead to subordinates considering female leaders incapable of fundamental managerial functions such as solving external and internal conflicts (Ko et al., 2015). Despite the downsides, women may still refrain from displaying agentic tendencies to avert dislike in stereotypically male-dominated settings (e.g., Dasgupta & Asgari, 2004).

Hypothesis 2. Female executives in male-dominant industry sectors employ more communal language than female executives in other industry sectors.

Suppose female and male executives differ in communal and agentic language. In that case, we are further interested in examining to what extent female and male leaders differ in communal and agentic language compared to general male and female language. In particular, if female executives appear more communal or agentic, how are their communication styles compared to the general female language? Similarly, if male executives are motivated to appear more communal, how communal are they compared to the general male language? We hypothesize that both male and female executives communicate with higher communal qualities than ordinary men to fit the androgynous leadership style of today. However, due to leadership descriptive and prescriptive norms linked to the agency, we hypothesize that executives communicate less communal than ordinary women.

It is important to note that since we do not have detailed knowledge about each component of LIWC 2015 composite scores, we limit our hypotheses to feminine and masculine linguistic styles instead of communal and agentic styles.

Hypothesis 3a. *Female executives employ a more feminine linguistic style than general male language.*

Hypothesis 3b. *Male executives employ a more feminine linguistic style than general male language.*

In addition, we conduct a complimentary analysis by comparing the speech content between female and male executives to explore the topics they choose to communicate. The content comparisons are exploratory by nature, and thus we do not state any formal hypotheses. Here, our focus is on identifying the key differences in what they say and offering theoretical and practical implications.

METHODS

Sample and Data

We collected our dataset from 522 executives featured in *The New York Times*' Corner Office column in the Sunday Business Section between March 2009 and October 2017. This dataset thus includes a broad spectrum of executives from various socio-demographic backgrounds and organizations (Bryant, 2017). Setting our sample in this way offers at least two advantages. First, as these executives are featured in a well-established and internationally recognized media outlet (Golan, 2006), we can reasonably assume that they are "successful." Second, there is only one interviewer during our sample period, meaning that our data contain a set of interview texts that are comparable across samples.

However, relying only on The New York Times interviews would severely limit the context in which executives exercise their impression management strategies. First, *The New York Times* is a prestigious outlet, and featured executives are likely to follow certain etiquette when talking about themselves and their organizations. Second, executives' responses are possibly molded by the interviewer's questions and the interest of the readership of *The New York Times*. To reduce the impact of the context-specific environment, we collect additional text data from different outlets, including YouTube captions, podcast transcripts, and other interviews, as described next.

The YouTube captions are unprocessed text extracted from video subtitles between December 2004 and May 2019. TED talks, commencement addresses, keynote speeches, and conference talks are formal-environment samples of this category. In this context, speakers mostly read or talk from prepared speeches with some on-the-spot questions. In contrast to this formal and stylized context, podcast and interview videos have an informal structure where people speak spontaneously and casually. In addition to YouTube captions, we collect publicly available podcast transcripts and lightly edited text of the interviews from November 2009 until November 2019.

The interviews from different sources belong to (i) business, management, and marketing- and advertising-oriented outlets (e.g., Forbes, Leaders Magazine, and The Drum) and (ii) lifestyle magazines (e.g., *Marie Claire*, *Cosmopolitan*, and *Harper's Bazaar*). Because some executives only choose to appear in corporate-oriented sources, we use management-related outlets. In general, we collect text edited by the interviewer.[1] These samples are from January 2003 to June 2019. Lastly, when interviews of international executives are unavailable in English, we search for news reports with the full name of the executive followed by "say," "said," "told," or "tells" to find direct quotes. This relatively small group primarily includes quotes from international executives or short responses from U.S. executives. We collect the samples from this category between February 2002 and June 2019.

After stripping out advertisements and headings, we organize the cleansed text data into a database where each executive-outlet pair is a single observation. Then, we remove all the words spoken by the interviewer (marked as bold text, italics, or beginning with Q) and markup language. Next, we use Python 3.7 and Linguistic Inquiry and Word Count (LIWC) software (Pennebaker, Boyd, Jordan, & Blackburn, 2015) to count the term frequencies. As a tested and validated dictionary, LIWC eliminates human bias as much as possible while exploring individuals' psychological and social states (Tausczik & Pennebaker, 2010).

[1] Some executives are rarely in the public eye and give infrequent interviews. For these executives, we look for personal blogs or any writing by them.

Finally, we augment the text data with additional information about the individuals and organizations as appropriate controls. Specifically, we collect information about individuals' age, final education, immigration background, and company information. Studies find that people use different word groups depending on their age (Pennebaker, 2011) and educational background (Pennebaker, 2011; Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014). Likewise, being in a competitive environment can affect the organizational culture (Barber & Odean, 2001; Eckel & Grossman, 2008; Niederle & Vesterlund, 2007). For instance, regardless of gender, leaders from publicly traded companies might have a stronger tendency toward an agentic linguistic style (Jones, 2016).

Dependent Variables

Communal linguistic style measures

Being expressive, people-oriented, social, kind, and soft are traits that are linked to communion (Abele & Wojciszke, 2014). Thus, we consider language associated with social and emotional qualities as communal. The 2015 version of LIWC has tested and validated composite scores related to these qualities, *authenticity* and *emotional tone*. *Authenticity* refers to an open and personal style in which people are vulnerable and honest (Newman, Pennebaker, Berry, & Richards, 2003); thus, it reveals itself as warm and sincere in spoken or written text data. For example, female leaders can be more open to talking about their stories in interviews. In this way, authentic language is more sociable and personal while less formal and analytical. *Emotional tone* represents the net positive emotion words (positive minus negative tone words) in the text (Cohn, Mehl, & Pennebaker, 2004). Similar to authenticity, showing one's emotion indicates vulnerability and desire to be socially connected, which is also part of female gender role, hence, matches communion (Abele & Wojciszke, 2014).

With the socially developed tie between communion and the female gender, it is not surprising that communal language overlaps the general female language. The female language is characterized by high usage of social words, emotion words, verbs, and cognitive process words. Women use these categories to describe people and events (Newman et al., 2008; Pennebaker, 2011; Tausczik & Pennebaker, 2010). Female language is also indirect (Mulac, Bradac, & Gibbons, 2001; Newman et al., 2008), as represented by the frequent use of auxiliary verbs and tentative words. First-person singular pronouns are another dominant category in female speech. While Newman et al. (2008) argue that first-person singular pronouns are the outcome of self-focus and avoiding assertiveness, Mulac et al. (2001) state that women pay attention to other people more than men and build dyadic ties first. Thus, female language style is inevitably linked to communion, and we use the feminine linguistic style index developed by Jones (2016) as our final measure of communal language:

$$Feminine\ Linguistic\ Style_i = First\ Person\ Singular\ Pronouns_i + Verbs_i + Auxiliary\ Verbs_i + Social\ Process\ Words_i + Positive\ Emotion\ Words_i + Negative\ Emotion\ Words_i + Cognitive\ Process\ Words_i + Tentative\ Words_i \qquad (1)$$

Agentic linguistic style measures

Agency is linked to instrumental, assertive, confident, and aggressive (Abele & Wojciszke, 2014). We employ linguistic styles representing these agentic qualities: *clout* and *analytic thinking*. Like *authenticity* and *emotional tone*, *clout* and *analytical thinking* are tested and validated composite measures from LIWC 2015. *Clout* represents self-confidence, higher social status, and leadership positions (Kacewicz et al., 2014), thus matching agency. *Analytic thinking* refers to a formal, logical, and distant linguistic style (Pennebaker et al., 2014). This language style is consistent with agentic characteristics of being logical and function-oriented.

Similar to how communal linguistic styles are strongly linked to female speech, agentic linguistic style is related to male speech patterns (Argamon, Koppel, Fine, & Shimoni, 2003; Jones, 2016; Koppel, Argamon, & Shimoni, 2002; Newman et al., 2008). Informative and instrumental style of male language (Mulac et al., 2001; Newman et al., 2008; Tausczik & Pennebaker, 2010) leads to more frequent use of long words (more than six letters), articles, and prepositions, anger words, swear words (Newman et al., 2008). Following Jones (2016), we also include the masculine linguistic style index as a final measure of the agency:

$$Masculine\ Linguistic\ Style_i = Long\ Words_i + First\ Person\ Plural\ Pronouns_i + Prepositions_i + Anger\ Words_i + Swear\ Words_i \quad (2)$$

Independent Variable

Female is an indicator variable for female executives. We record the gender of the person based on first names, photos, pronoun references within the texts, and common knowledge (e.g., well-known figures).

Control Variables

Age is a continuous variable that we collect from multiple sources (e.g., Wikipedia pages, Bloomberg executive data, LinkedIn pages). When information on age is unavailable, we approximate the age data based on graduation year (23 years for Bachelor's and 27 years for Master's/MBA)[2] (National Center for Education Statistics, 2011). *Publicly traded* is a binary indicator that takes one for a publicly-traded company and zero otherwise. *MBA* is an indicator of having an MBA degree. This information comes from the texts, Wikipedia, Bloomberg, LinkedIn, and Google Search. *Native speaker* is an indicator of a native English speaker, which takes one, if

[2] In a few cases in which no graduation data are available on any of the sources, we use the most recent photo of an executive as well as his/her work experience details from LinkedIn to make an educated guess.

the person was an English speaker born and raised in the United States, and zero, including speakers from other English-speaking countries (e.g., the United Kingdom, Canada, New Zealand). The primary motivation behind this categorization is the cultural implications of the same language in different countries (Pennebaker, 2011). As language shapes one's cognition early in life, we count people who migrated to the United States at an early age as non-native speakers (Pennebaker, 2011; Perani et al., 2003).

Sector categorizes the industry into consumer discretionary, communication, technology, and other business sectors based on the Bloomberg Industry Classification system. Consumer discretionary includes a range of sub-industries, including automotive, apparel and textile, home and office products, transportation, entertainment facilities, and non-profit organizations. The communication sector includes media companies focusing on publishing and broadcasting, advertising and marketing, and Internet-based services. Hardware and software companies belong to the technology sector. The other sector category is consumer staples, financials, government, industrials, health care, and utilities in the other sector category. Finally, *interview* is a binary variable coded one, when the text is from interview-style data, and zero, otherwise. Here, interview-style means structured interviews, podcasts, YouTube captions of podcasts, and interviews.

Empirical Specifications

To test hypotheses 1 and 2 empirically, we specify the following general linear regression model:

$$Y_i = \alpha + \beta Female_i + \gamma' \mathbf{Z}_i + \epsilon_i \qquad (3)$$

where the dependent variable Y_i represent the communal or agentic linguistic style measures discussed earlier, and vector $\mathbf{Z}_i$ represents the control variables, including both individual and firm characteristics. The key parameter is β, representing the gender difference in the communal or agentic linguistic style while controlling for the individual and firm characteristics.

Comparison to the general public

Hypotheses 3 compare the languages of executive women and men to that of general women and men. This investigation is conducted using the gender language comparison reported in Newman et al. (2008) based on the corpus of 14,000 written and spoken text samples (e.g., talk show conversations, bestselling fiction novels, psychology essay assignments) 67% belongs to college-age participants. These samples are from different periods (17th century and the decades between 1980 and 2002) and three English-speaking countries (England, New Zealand, and the United States), of which 60% of the data belongs to women. Newman et al. (2008) only report sample statistics and some of the measures we specified above, so our comparison is limited to what is available. We investigate the effect sizes (*d*) of available linguistic categories. Following the suggestion of Cohen (1992), we set the small effect size (d = 0.2), medium effect size (d = 0.5), and large effect size (d = 0.8).

Content Analysis

We compare the content in the speech by examining both frequent and gender-distinguishing terms using a standard "bag of words" approach. N-grams are tokens consisting of N words. For example, *leader* is a uni-gram, *strong leader* is a bi-gram (two words), and *strong female leader* is a tri-gram (three words). Counting frequent N-grams and comparing male and female executives allow us to highlight which terms they emphasize in their speech.

We identify gender-distinguishing terms by employing a classification model. Using relevant tokens as covariates, we estimate a probabilistic model that classifies speech into that spoken by men or women. Because the number of text-based covariates is higher than the sample size (the so-called p >> n problem), we cannot use linear models in this context (Varian, 2014). Instead, we employ a random forest classification model widely used in the data science and machine learning literature (Breiman, 2001). The random forest model, an ensemble learning algorithm, combines multiple decision trees that provide the highest prediction accuracy (Ho, 1995).[3] This analysis produces a list of most salient terms in classifying a given speech to either produced by female or male executives, thereby identifying the most distinguishing terms between them.

RESULTS

Table 1 summarizes the descriptive statistics of the text data, and Table 2 shows the characteristics of the executives in the sample. Despite the podcast corpora having a relatively low executive population (28 men and 15 women), the average word count for each sector is the highest across the whole corpus. Most podcasts last approximately three hours, providing more spoken words than other text sub-samples. More male than female executives have an MBA, and more men than women are heads of publicly traded companies. By contrast, more female executives than male executives are native speakers. On average, both male and female executives are 56 years.

[3] We do not dwell on the details of the random forest model here, as it is not our primary estimation. Please see Breiman (2001) for more details. We use the utils package from Python's Gensim library (Rehurek & Sojka, 2010) and follow the common procedure of stemming and lemmatizing using PorterStemmer (Porter, 1980) and WordNetLemmatizer of the Natural Language Tool Kit package (Fellbaum, 1998). We focus on the top 1, 2, and 3 grams for men and women separately by employing the CountVectorizer algorithm from the scikit-learn library (Pedregosa et al., 2011). We then assess the model fit area based on the out-of-sample predictive power using K-fold cross-validation (Hindman, 2015).

Insert Tables 1 and 2 about here

Communal vs. Agentic Styles

Table 3 shows the results of the regression in equation (3), where authenticity, emotional tone, and female linguistic styles are the dependent variables in Models 1, 2, and 3, respectively. In Model 1, the key variable, *female*, is highly significant with a positive sign for authenticity, indicating that female leaders are more likely to employ an authentic linguistic style. The results from Model 2 shows that female leaders have a more positive emotional tone in their language, indicated by the *female* coefficient being significant with a positive sign. Similarly, in Model 3, *female* is positive and highly significant. These three models suggest that female executives employ communal language more than male executives. Thus, Hypothesis 1a is supported.

The results for agentic style are examined in Models 4, 5, and 6 in the same table, where the dependent variables are clout, analytic thinking, and masculine linguistic style. The results illustrate female leaders employ a less agentic linguistic style. In Model 4, the *female* coefficient is highly significant with a negative sign, implying that female leaders use fewer clout words. Model 5 examines the gender difference in analytic thinking, and we find that female leaders' language is more in narrative style and less in a structured and formal way. In Model 6, the key variable, *female*, is highly significant with a negative sign for masculine style, revealing that female leaders are less likely to employ a masculine linguistic style. Overall, we find that female executives use less agentic linguistic style; thus, hypothesis H1b is supported.

Insert Table 3 and Table 4 about here

Industry Sector Comparisons

To test Hypothesis 2, we run the communal (i.e., authentic, emotional tone, feminine) and agentic (i.e., analytic thinking, clout, masculine) linguistic style models while accounting for the industry sectors. The results are presented in Table 4. Among the four industry sectors, *Consumer discretionary* sector leaders employ less agentic and more communal styles. Female leaders in this sector appear slightly less authentic than those in the communication sector. Similarly, we find that women in the consumer discretionary sector have lower feminine linguistic style words. When it comes to agentic style measures, these leaders are highly confident and masculine in their language. Technology leaders only show slight differences by being less authentic and more clout than communication leaders. Female leaders of other sectors (e.g., health care, utilities, government) are significantly less authentic and slightly less feminine while being more clout in their linguistic portrayal. Therefore, Hypothesis 2 is partially supported, showing those female executives in the consumer discretionary and other sectors talk more like male executives than other female leaders.

Compared to the general male and female language

The left columns of Table 5 show the mean scores reported by Newman et al. (2008) of the linguistic characteristics of "typical" men and women and the effect size, which measures the variation in word categories when the differences are significant. The p-value is insufficient to explain the magnitude of the difference between the usage of word categories. If the scores between the two groups highly overlap, this implies little difference in word categories between male and female executives, resulting in a relatively small effect size.

Insert Table 5 about here

Of the six masculine language categories, female executives score higher for *long words*, *first-person plural pronouns*, *articles,* and *prepositions* than the general female population. The comparisons to Newman et al. (2008) show that female executives have a more masculine style than the general female language. Thus, Hypothesis 3a is supported. Of the seven feminine linguistic style word categories, male executives score higher for *social processes*, *cognitive processes*, *tentative, verbs,* and *positive emotions* than the general male population. Male leaders employ fewer *first-person singular pronouns* and slightly fewer *negative emotion words*. The comparisons to Newman et al. (2008) show that male executives have a more feminine style than the general male language. Hence, Hypothesis 3b is supported.

Exploratory Content Analysis

So far, we have found significant differences in the agentic and communal linguistic styles adopted by male and female executives, indicating the different ways they communicate and that they are consistent with gender prescriptions. We now investigate whether the content of their communication differs by gender. Table 6 presents the most frequent 1, 2, and 3 grams as male and female executives' most often spoken terms. The most frequently spoken terms are almost identical between female and male executives, which we should expect, as these executives serve the same functions in their organizations. Thus, when they speak in the public domain, they naturally discuss their people, work, organizations, actions, and decisions.

Insert Table 6 about here

Next, we investigate the most salient words that distinguish gender. From the 747 tokens we obtain, we employ 10-fold cross-validation with an 80/20 split to run the random forest classification model. We rerun the model with different sets of parameters for the parameter tuning to obtain the optimal ones (i.e., number of trees in a forest, the maximum number of features considered for the node split, and depth of a tree). After the hyperparameter tuning, we set the number of trees in the forest to 1200 with a maximum depth of 460. The maximum number of features is the square root of the number of tokens for each split (i.e., “auto” in the sci-kit-learn library). The area under the curve value represents model performance, which is bounded between zero and one, where one indicates a perfect fit. After the 10-fold cross-validation, the resulting model has an average score of 0.86, which Bradley (1997) considers excellent performance.

Table 7 shows the 20 most important features used by the model to classify a given text into a gender category. For example, the model tends to classify text with high counts of terms such as “women” and “woman” produced by female executives. The N-gram comparisons and random forest classification reveal that while the content most frequently spoken by executives is similar, some features are distinctly different by gender. The terms that women tend to mention substantially more than men can be sorted into *female references*, *family*, *positive emotion*, and *home* LIWC categories. As these terms are linked to communion, our finding indicates that female executives use these terms to convey the communal image.

Insert Table 7 about here

Robustness Checks

Our estimation results strongly support that female executives use more of a communal and less of an agentic language and tend to mention more about women and family. However, as our sample constitutes interviews as the main component, it is unclear if the interviewer's questions drive these results. We examine our results' robustness by investigating whether the interviewer's questions shape the answers. We divide the sample into interview and non-interview sources and re-estimate the models separately. Additionally, estimate the joint model (interviews and non-interviews) with interview fixed effects. Table 8 presents the results (Models 1 and 2), showing that the gender effects remain virtually unchanged in both specifications.

Insert Table 8 about here

Second, we find that female executives speak more about women, family, home, and positive emotions from our content analysi*s*. However, past research shows that interviewers ask women different questions than men (Ormiston & Wong, 2019). Thus, the questions posed to female executives rather than what they voluntarily say might influence our results. To investigate this, we explore the content differences between men and women from the interview and non-interview sources. Again, Table 9 allows us to conclude that the questions do not drive the gender differences in agentic/communal linguistic styles.

Insert Table 9 about here

In the non-interview text, in which executives are likely to have more freedom about what they say, female leaders still use the LIWC categories of *family* ($p < 0.1$), *female references* ($p < 0.001$), and *home* ($p < 0.001$) more than men (Table 9). By studying the actual text, we find that, with family-related words, female executives discuss their upbringing the dilemmas in balancing family responsibilities and work. Female executives also mention their children when discussing how they make time despite a heavy work obligation. They typically use female-related words to discuss aspiring female figures in their lives. "Home" refers to family and where these executives come from, highlighting belonging. The results from the interview source subsample (Table 9) show much larger sets of categories that differ by gender. Thus, although the prompted questions do not completely drive our results, female executives seem to speak about different content more in interviews than on other occasions. This may reflect the audience's interests (Little, Major, Hinojosa, & Nelson, 2015; Ruderman, Ohlott, Panzer, & King, 2002). It is interesting that female executives are willing to share these personal details with the public audience, prompted or not.

Are female executives asked different questions?

We analyze the questions from all the interviews to examine whether female executives receive different questions from male executives. We then compare the frequencies of LIWC categories by gender; Table 10 shows those exhibiting significant gender differences. Interviewers more frequently ask female executives about *family* ($p < 0.1$), *achievement*-related questions ($p < 0.01$), *time* ($p < 0.01$), and *past focus* ($p < 0.01$). Regarding *female references* ($p < 0.001$), they ask female executives about various female-related topics; these include female-related products, female empowerment programs, and views on female leadership. They also tackle such subjects as the lack of female leaders in corporate America, the challenges of being a woman in a sector, and whether interviewees can offer any advice for women. Thus, interviewers ask female executives questions relating to women, which explains at least partially why female leaders use female reference categories significantly more than male leaders. On the contrary, we find no significant differences in terms of *home*, indicating that interviewers' questions do not influence female executives discussing these topics. Finally, *family*-related words appear more significantly in the interview content, implying that female executives talk more about family owing to the questions than male leaders.

Insert Table 10 about here

DISCUSSION AND CONCLUSIONS

Creating the image of a leader involves specific challenges for women. Social role theory highlights the causes of these obstacles by highlighting the clash between the communal female gender norms and agentic leader image. Due to the incongruity, female executives are likely to face a backlash when they act agentic but appear incompetent when communal. However, the literature also suggests that being communal is no detriment but can be advantageous to very top female executives.

Analyzing a sample of text data from 522 executives, we examine how female executives use agentic and communal language to manage their public impressions. They use less of an agentic and more of a communal linguistic style when presenting themselves in the public domain than male executives. This is consistent with the view that being communal is advantageous once women reach the top leadership positions.

However, this relationship is moderated by the industry in which they work. We find that female executives in a less male-dominated industry (consumer discretionary) employ more agentic styles. Having a higher number of women means the consistent integration of female members onto executive boards (Milliken & Martins, 1996), less prejudice against women (Ely, 1995; Hoobler, Wayne, & Lemmon, 2009), and easier access to social and industry networks (Joshi, Son, & Roh, 2015). Eagly et al. (2020) argue that with their increase in college education and labor force participation, women's aspirations and attitudes have started to resemble men's, suggesting a more openness toward nonstereotypical women. This finding also aligns with the literature that women take more risks (Booth, Cardona-Sosa, & Nolen, 2014) and occupy more leadership positions (Dasgupta & Asgari, 2004) in a female-dominated environment. Similarly, women appear as more effective leaders in female-dominated industries (Ko et al., 2015). Hence, these women are less likely to be restricted by gender norms.

When we extend our analyses to compare the language of executives to general male and female language reported by Newman et al. (2008), we find that female leaders have a stronger masculine linguistic style than general female language. Similarly, male leaders have a stronger feminine linguistic style than general male language. These findings may imply that female and male leaders have some androgyny in their linguistic style compared to the general gender language when they communicate to the general public. This is consistent with the general trend favoring transformational leadership style, that female leaders and male leaders combine agentic and communal in their image. Although Newman et al. (2008)'s results are based on the most comprehensive text sample currently available in the literature, it is possible that other factors, such as age and education, that are not explicitly accounted for in this analysis are driving our results. Thus, it would be interesting to conduct a similar comparison where these characteristics can be accounted for.

The higher usage of *first-person singular pronouns* by female executives, a pattern persistent even when they discuss their corporate roles and organizational tasks, might imply a sense of exclusion (i.e., not being part of the team). Shared values and common characteristics (i.e., gender) shape the group identity and determine who becomes part of an executive team. When men dominate an executive boardroom, women appear as outsiders (Hambrick, 2007; Hambrick & Mason, 1984).

In contrast to the significant differences in linguistic styles, we also find that the content of their communication—assessed by the most frequently spoken words—is strikingly similar between male and female executives. Combined with the significant style differences, this illustrates that *what* they say is similar, whereas *how* they say it is different. Both the n-grams and the random forest model reveal that the more frequent use of *female-related words*, *positive emotion words*, and *family-related words* are distinguishing features of the speech of female executives. Female executives show concern about the obstacles that women face to reach the upper echelons of corporate leadership and mention inspirational female figures in their lives. This overlaps with the literature on the importance of female role models to advance to leadership positions (Ibarra, 1993; Tharenou, 2001).

We provide a new insight that female top leader portrayal constitutes the display of family and women-regarding attitudes. Our robustness checks reveal that family and women-related topics are reinforced in both interviewed and non-interviewed content, showing that these are the topics that these women are willing to share and the public is interested to hear.

However, the existing literature on women's labor market conditions overwhelmingly portrays motherhood as an obstacle to advancement (Correll, Benard, & Paik, 2007; Cuddy et al., 2004; Heilman, 2001). By showing that top executive women talk about motherhood and family significantly more than male executives, both in and outside of interviews, we add topics of motherhood and family responsibilities to communal display that support the leadership image of women but only at top levels (Heilman, 2012; Little et al., 2015; Pillemer et al., 2014; Rosette & Tost, 2010). If top female executives can increase their perception of competence by clearing the double hurdles, family and motherhood constitute the third hurdle they could clear, thereby increasing the competence perception even more.

We acknowledge the following limitations of this work. First, we focus on how executives manage their public image outside their organizations. Since images that these leaders convey reflect the organization's desired image, what the leaders say could be carefully crafted through communication advisors. These outside images may differ from how they conduct themselves inside the organization or their true personality, and we cannot disentangle them. Still, the past literature indicates that spoken language is hard to deflate (Pennebaker, 2011), and it is cognitively consuming to create different personal images for different settings (Hewlin, Dumas, & Burnett, 2017). Therefore, our findings may not be too far from the actual behavior of the executives.

Second, we are not able to account for unobserved individual characteristics. For instance, these women at the top executive levels may have fundamentally different personalities or social preferences. For example, Adams and Funk (2012) report that female leaders have comparable personality traits to male leaders, only differing in benevolence, a stereotypically communal characteristic. Such underlying differences could be driving the language differences between female and male executives across industry sectors and styles different from the general female language. Due to the data limitation, we could not account for unobserved individual characteristics, e.g., via individual-fixed effects. However, if the top female executives have more agentic underlying characteristics, we might expect them to speak with more agency. We find the opposite is thus consistent with the literature supporting the female leadership advantage. However, executive women talking more agentic than general women might be due to their underlying characteristics. A deeper investigation of leadership and individual's underlying characteristics may provide further insights.

Third, the sample used in this study comprises executives featured by *The New York Times* and focuses on the executives based in the U.S. Our sample includes a large number of executives from diverse backgrounds, well-reflecting the demographically diverse U.S. corporate executives compared to other countries (Hambrick, 2007; Milliken & Martins, 1996). Even so, using other media sources or samples, including executives in other countries, may produce different results, as different countries operate in various corporate and organizational cultures and different degrees of gender equality. Therefore, our results may not apply to other countries. Further investigation of the country and cultural contexts may be fruitful.

Finally, we are not able to address how the language style might have evolved through the career path of these executives. Given their high status, we find that women are eager to display their communal sides, implying that this is beneficial for top female executives. However, these same women might have needed different strategies to climb the organizational hierarchy to demonstrate their competence while avoiding backlash. Even though manipulating and adjusting linguistic style seem easy and trivial, the research indicates that conveying the desired image with conviction requires certain honesty to their characters. Thus, it would be an interesting future research topic to assess how their true personality and the public language align and possibly evolve as they change their roles in the organization.

REFERENCES

Abele, A. E., & Wojciszke, B. 2014. Communal and agentic content in social cognition: A dual perspective model. In Mark P. Zanna & James M. Olson (Eds.), ***Advances in experimental social psychology***, 195-255. Academic Press.

Adams, R. B., & Funk, P. 2012. Beyond the glass ceiling: Does gender matter? ***Management science***, *58*: 219–235.

Amanatullah, E. T., & Morris, M. W. 2010. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. ***Journal of Personality and Social Psychology***, 98: 256–267.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. 2003. Gender, genre, and writing style in formal. ***Text & Talk***, 23: 321–346.

Bakan, D. 1966. ***The duality of human existence. An essay on psychology and religion.*** Chicago: Rand Mcnally.

Barber, B., & Odean, T. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. ***The Quarterly Journal of Economics***, 116: 261–292.

Bass, B. M. 1990. From transactional to transformational leadership: Learning to share the vision. ***Organizational dynamics***, 18: 19-31.

Booth, A., Cardona-Sosa, L., & Nolen, P. 2014. Gender differences in risk aversion: Do single-sex environments affect their development? ***Journal of Economic Behavior and Organization***, 99: 126–154.

Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. ***Pattern Recognition***, 30: 1145–1159.

Breiman, L. 2001. Random forests. ***Machine Learning***, 45: 5–32.

Brotheridge, C., Lee, R., Riggio, R. E., & Reichard, R. J. 2008. The emotional and social intelligences of effective leadership: An emotional and social skill approach. ***Journal of Managerial Psychology***, 23: 169–185.

Bryant, A. 2017. How to be the big boss. ***The New York Times***. October 29: 1.

Burns, J. M. 1978. ***Leadership***. New York: Harper & Row

Cabrera, S. F., Sauer, S. J., & Thomas-Hunt, M. C. 2009. The evolving manager stereotype: The effects of industry gender typing on performance expectations for leaders and their teams. ***Psychology of Women Quarterly*****,** 33: 419-428.

Carli, L. L. 2001. Gender and social influence. ***Journal of Social issues,*** 57: 725-741.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. 2019. Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. ***Strategic Management Journal*****,** 40: 1705–1732.

Cialdini, R. B., & Trost, M. R. 1998. Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), ***The handbook of social psychology,*** 151–192. McGraw-Hill.

Coates, J. 2015. ***Women, men and language: A socio-linguistic account of gender differences in language*** (3rd ed). New York, NY: Routledge.

Cohen, J. 1992. A power primer. ***Psychological Bulletin**,* 112: 155–159.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. 2004. Linguistic markers of psychological change surrounding September 11, 2001. ***Psychological science*****,** 15: 687-693.

Connor, R. A., & Fiske, S. T. 2018. Warmth and competence: A feminist look at power and negotiation. In C. B. Travis, J. W. White, A. Rutherford, W. S. Williams, S. L. Cook, & K. F. Wyche (Eds.), ***APA handbook of the psychology of women: History, theory, and battlegrounds,*** 321-342. American Psychological Association.

Correll, S. J., Benard, S., & Paik, I. 2007. Getting a job: Is there a motherhood penalty? ***American Journal of Sociology***, 112: 1297–1338.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. 2004. When professionals become mothers, warmth doesn't cut the ice. ***The Journal of Social Issues**,* 60: 701–718.

Dasgupta, N., & Asgari, S. 2004. Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. ***Journal of Experimental Social Psychology***, 40: 642–658.

Dezsö, C. L., & Ross, D. G. 2012. Does female representation in top management improve firm performance? A panel data investigation. ***Strategic Management Journal**,* 33: 1072–1089.

Eagly, A. H. 1987. ***Sex differences in social behavior*** (1st ed.). New York: Psychology Press.

Eagly, A. H., & Carli, L. L. 2003. The female leadership advantage: An evaluation of the evidence. ***The Leadership Quarterly***, 14: 807–834.

Eagly, A. H., & Carli, L. L. 2007. ***Through the labyrinth: The truth about how women become leaders***. Boston, MA: Harvard Business School Press.

Eagly, A. H., & Karau, S. J. 2002. Role congruity theory of prejudice toward female leaders. ***Psychological Review***, 109: 573–598.

Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. 2003. Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. ***Psychological bulletin*****,** 129: 569.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. 2020. Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. ***American Psychologist,*** 75: 301–315. https://doi.org/10.1037/amp0000494

Eagly, A., Wood, W., & Diekman, A. 2000. Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), ***The developmental social psychology of gender***, chapter 5. New York: Psychology Press.

Eckel, C. C., & Grossman, P. J. 2008. Men, women and risk aversion: Experimental evidence. In C. R. Plott & V. L. Smith (Eds.), ***Handbook of experimental economics***, 1061–1073. New York: North-Holland.

Eckes, T. 2002. Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. ***Sex Roles*****,** 47: 99-114.

Ely, R. J. 1995. The power in demography: Women's social constructions of gender identity at work. ***Academy of Management Journal***, 38: 589–634.

Fellbaum, C. 1998. ***WordNet: An electronic lexical database***. Cambridge, MA: The MIT Press

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. 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.

Fombrun, C., & Shanley, M. 1990. What's in a name? Reputation building and corporate strategy. ***Academy of Management Journal***, 33: 233–258.

Goffman, E. 1959. ***The presentation of self in everyday life***. Garder City, N.Y.: Doubleday.

Golan, G. 2006. Inter-media agenda setting and global news coverage: Assessing the influence of The New York Times on three network television evening news programs. ***Journalism Studies***, 7: 323–333.

Hambrick, D. C. 2007. Upper echelons theory: An update. ***Academy of Management Review***, 32: 334–343.

Hambrick, D. C., & Mason, P. A. 1984. Upper echelons: The organization as a reflection of its top managers. ***Academy of Management Review***, 9: 193–206.

Heilman, M. E. 2001. Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. ***The Journal of Social Issues***, 57: 657–674.

Heilman, M. E. 2012. Gender stereotypes and workplace bias. ***Research in Organizational Behavior***, 32: 113–135.

Heilman, M. E., & Okimoto, T. G. 2007. Why are women penalized for success at male tasks? The implied communality deficit. ***Journal of Applied Psychology***, 92: 81–92.

Hewlin, P. F., Dumas, T. L., & Burnett, M. F. 2017. To thine own self be true? Facades of conformity, values incongruence, and the moderating impact of leader integrity. ***Academy of Management Journal***, 60: 178–199.

Hindman, M. 2015. Building better models: Prediction, replication, and machine learning in the social sciences. ***Annals of the American Academy of Political and Social Science***, 659(1): 48–62.

Ho, T. K. 1995. **Random decision forests**. Paper presented at the 3rd international conference on document analysis and recognition.

Hoobler, J. M., Wayne, S. J., & Lemmon, G. 2009. Bosses' perceptions of family-work conflict and women's promotability: Glass ceiling effects. ***Academy of Management Journal***, 52: 939–957.

Ibarra, H. 1993. Personal networks of women and minorities in management: A conceptual framework. ***Academy of Management Review***, 18: 56–87.

Jones, J. J. 2016. Talk "like a man": The linguistic styles of Hillary Clinton, 1992-2013. ***Perspectives on Politics***, 14: 625–642.

Joshi, A., Son, J., & Roh, H. 2015. When can women close the gap? A meta-analytic test of sex differences in performance and rewards. ***Academy of Management Journal***, 58: 1516–1545.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. 2014. Pronoun use reflects standings in social hierarchies. ***Journal of Language and Social Psychology***, 33: 125–143.

Kark, R., Waismel-Manor, R., & Shamir, B. 2012. Does valuing androgyny and femininity lead to a female advantage? The relationship between gender-role, transformational leadership and identification. ***The Leadership Quarterly***, 23: 620-640.

Ko, I., Kotrba, L., & Roebuck, A. 2015. Leaders as males?: The role of industry gender composition. ***Sex Roles*****,** 72: 294-307.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. 2011. Are leader stereotypes masculine? A meta-analysis of three research paradigms. ***Psychological bulletin*****,** 137: 616.

Koppel, M., Argamon, S., & Shimoni, A. R. 2002. Automatically categorizing written texts by author gender. ***Literary and Linguistic Computing***, 17: 401–412.

Larsen, R. J., & Seidman, E. 1986. Gender schema theory and sex role inventories: Some conceptual and psychometric considerations. ***Journal of Personality and Social Psychology*****,** 50: 205.

Lemoine, G. J., & Blum, T. C. 2021. Servant leadership, leader gender, and team gender role: Testing a female advantage in a cascading model of performance. ***Personnel Psychology*****,** 74: 3-28.

Little, L. M., Major, V. S., Hinojosa, A. S., & Nelson, D. L. 2015. Professional image maintenance: How women navigate pregnancy in the workplace. ***Academy of Management Journal***, 58: 8–37.

Love, E. G., Lim, J., & Bednar, M. K. 2017. The face of the firm: The influence of CEOs on corporate reputation. ***Academy of Management Journal***, 60: 1462–1481.

Milliken, F. J. & Martins, L. L. 1996. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. ***Academy of Management Review***, 21: 402–433.

Mulac, A., Bradac, J. J., & Gibbons, P. 2001. Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. ***Human Communication Research***, 27: 121–152.

National Center for Education Statistics (2011). ***Profile of Graduate and First-Professional Students: Trends from Selected Years, 1995–96 to 2007–08***. Accessed online https://nces.ed.gov/pubs2011/2011219.pdf#page=10. October.

Newman, M. L., Groom, C. J., Handelman, L. D., and Pennebaker, J. W. 2008. Gender differences in language use: An analysis of 14,000 text samples. ***Discourse Processes***, 45: 211–236.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. 2003. Lying words: Predicting deception from linguistic styles. ***Personality and social psychology bulletin,*** 29: 665-675.

Ng, T. W. 2017. Transformational leadership and performance outcomes: Analyses of multiple mediation pathways. ***The leadership quarterly,*** 28: 385-417.

Niederle, M., & Vesterlund, L. 2007. Do women shy away from competition? Do men compete too much? ***The Quarterly Journal of Economics***, 122: 1067–1101.

Ormiston, M., & Wong, E. M. 2019. Ask me anything: The influence of gender bias on journalist questions of CEOs and CEO responses. ***Academy of Management Proceedings***. https://doi.org/10.5465/AMBPP.2019.11451abstract.

Pan, L., McNamara, G., Lee, J. J., Haleblian, J. J., & Devers, C. E. 2018. Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. ***Strategic Management Journal***, 39: 2204–2225.

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. 2014. Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. ***Journal of applied psychology*****,** 99: 1129.

Pedregosa, F. et al. 2011. Scikit-learn: Machine learning in Python. ***Journal of Machine Learning Research***, 85: 2825–2830.

Pennebaker, J. W. 2011. ***The secret life of pronouns what our words say about us***. New York: Bloomsbury.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. 2015. ***The development and psychometric properties of LIWC2015***. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. 2014. When small words foretell academic success: The case of college admissions essays. ***PLoS ONE***, 9(12): e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., & Lay, T. C. 2002. Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. ***Journal of Research in Personality***, 36: 271–282.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. 2003. The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. ***Human Brain Mapping***, 19: 170–182.

Pillemer, J., Graham, E. R., & Burke, D. M. 2014. The face says it all: CEOs, gender, and predicting corporate performance. ***The Leadership Quarterly***, 25: 855–864.

Pollach, I., & Kerbler, E. 2011. Appearing competent: A study of impression management in U.S. and European CEO profiles. ***The Journal of Business Communication***, 48: 355–372.

Porter, M. F. 1980. An algorithm for suffix stripping. ***Program: electronic library and information systems,*** 14: 130–137.

Powell, G.N., Butterfield, D.A., & Bartol, K.M. 2008, Leader evaluations: a new female advantage? ***Gender in Management*****,** 3: 156-174.

Rehurek, R., & Sojka, P. 2010. ***Software framework for topic modelling with large corpora***. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks: 46–50.

Rosenfeld, P., Giacalone, R. A., & Riordan, C. A. 1995. ***Impression management in organizations: Theory, measurement, practice***. New York, NY: Routledge.

Rosette, A. S., & Tost, L. P. 2010. Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. ***Journal of Applied Psychology***, 95: 221–235.

Ruderman, M. N., Ohlott, P. J., Panzer, K., & King, S. N. 2002. Benefits of multiple roles for managerial women. ***Academy of Management Journal***, 45: 369–386.

Rudman, L. A., & Glick, P. 1999. Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. ***Journal of personality and social psychology*****,** 77: 1004.

Rudman, L. A., & Glick, P. 2001. Prescriptive gender stereotypes and backlash toward agentic women. ***The Journal of Social Issues***, 57: 743–762.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. 2012. Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. ***Journal of Experimental Social Psychology***, 48: 165–179.

Rudman, L. A., & Phelan, J. E. 2008. Backlash effects for disconfirming gender stereotypes in organizations. ***Research in Organizational Behavior***, 28: 61–79.

Schlenker, B. R. 2012. ***Self-presentation***. In M. Leary & J. Tangney (Eds.), **Handbook of self and identity** (2nd ed.), 542–570. New York, NY: The Guildford Press.

Smith, D. G., Rosenstein, J. E., Nikolov, M. C., & Chaney, D. A. 2019. The power of language: Gender, status, and agency in performance evaluations. ***Sex Roles***, 80: 159–171.

Sosik, J. J., Avolio, B. J., & Jung, D. I. 2002. Beneath the mask: Examining the relationship of self-presentation attributes and impression management to charismatic leadership. ***The Leadership Quarterly***, 13: 217–242.

Spencer, S. J., Logel, C., and Davies, P. G. 2016. Stereotype threat. ***Annual Review of Psychology***, 67: 415–437.

Tausczik, Y. R., & Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. ***Journal of Language and Social Psychology***, 29: 24–54.

Tharenou, P. 2001. Going up? Do traits and informal social processes predict advancing in management? ***Academy of Management Journal***, 44: 1005–1017.

Toneva, Y., Heilman, M. E., & Pierre, G. 2020. Choice or circumstance: When are women penalized for their success? ***Journal of Applied Social Psychology*****,** 50: 651-659.

Tucker, S., Turner, N., Barling, J., Reid, E. M., & Elving, C. 2006. Apologies and transformational leadership. ***Journal of Business Ethics***, 63: 195–207.

Varian, H. R. 2014. Big data: New tricks for econometrics. ***The Journal of Economic Perspectives***, 28: 3–27.

Vinkenburg, C. J., Van Engen, M. L., Eagly, A. H., & Johannesen-Schmidt, M. C. 2011. An exploration of stereotypical beliefs about leadership styles: Is transformational leadership a route to women's promotion? ***The Leadership Quarterly*****,** 22: 10-21.

Vroman, S. R., & Danko, T. 2020. Against what model? Evaluating women as leaders in the pandemic era. ***Gender, Work & Organization*****,** 27: 860-867.

Williams, M. J., & Tiedens, L. Z. 2016. The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. ***Psychological Bulletin*****,** 142:165.

TABLE 1
Descriptive Statistics of the Text based on Sector, Gender, and Source

		Consumer Discretionary				Communication				Technology				Other Sectors			
		Male		Female		Male		Female		Male		Female		Male		Female	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1.	Essay	973	612.40	795	193.07	927.20	390.05	808.50	116.91	884.26	380.88	799.75	210.80	2,019.29	2913.64	1,373	1,215.5
2.	Caption	2,747.41	2,794.38	2,470.93	2,981.29	2,357.74	1,666.87	3,357.80	5,073.97	3,115.05	2,478.63	3,508.58	3,883.60	2,131.05	2,116.83	3,222.24	3,892.34
3.	Podcast	2,900.25	2,424.96	4,665.50	3,009.28	5,343.33	3.573,68	5,708.75	1,464.28	7,605.67	5,665.56	5,207.00	2,636.09	6,373.67	5,301.66	8,612.67	6,374.65
4.	Interview	1,536.20	1,523.96	1,239.21	642.00	1,193.06	804.16	1,154.13	620.67	1,755.43	1,698.51	1,016.50	504.99	1,493.96	1,401.24	1,157.05	633.10
5.	Other	407.00	328.97	984.57	906.92	1,166.25	827.63	651.67	49.17	609.00	384.42	83.67	47.96	750.50	569.59	588.75	963.75
6.	*New York Times*	1,239.91	418.17	1,181.57	364.35	1,174.87	338.29	1,260.15	381.45	1,126.13	314.62	1,042.12	274.26	1,249.35	374.21	1,155.27	354.43

Note: The table contains descriptive statistics of word count for each media outlet based on the industrial sector and the gender of the executive.

TABLE 2
Descriptive Statistics of the Individual and Company Characteristics

	Consumer Discretionary				Communication				Technology				Other Sectors			
	Male		Female		Male		Female		Male		Female		Male		Female	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1. MBA degree of the executive	0.32	0.47	0.21	0.41	0.17	0.38	0.11	0.31	0.35	0.48	0.31	0.47	0.38	0.49	0.36	0.48
2. Publicly traded status of the company	0.42	0.50	0.05	0.22	0.18	0.39	0.11	0.31	0.34	0.48	0.09	0.30	0.39	0.49	0.22	0.42
3. Native speaker status of the executive	0.83	0.38	0.90	0.30	0.80	0.40	0.80	0.40	0.64	0.48	0.81	0.40	0.71	0.46	0.93	0.25
4. Age of the executive	59.50	9.74	57.22	9.14	52.77	10.97	56.65	8.94	52.25	8.28	51.34	8.69	61.77	9.19	56.78	8.48

Note: The table contains descriptive statistics of the executive characteristics (MBA degree, native speaker status, and age) and a corporate feature (publicly traded status). Mean values represent the number of male/female MBA holders divided by the total number of male/female executives in a particular industry. The numbers for the publicly traded status of the company and the native speaker status of the executive are also calculated similarly. For each industrial sector, the age of the executive is the average number of years belonging to male and female executives, respectively.

TABLE 3
Linguistic Style Differences among the Executives

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Dependent Variable	**Authentic[1]**	**Tone[1]**	**Feminine Linguistic Style[2]**	**Analytic Thinking[1]**	**Clout[1]**	**Masculine Linguistic Style[2]**
Female	5.78^{**}	3.03^{*}	1.61^{**}	-5.14^{**}	-3.26^{**}	-1.01^{**}
	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
MBA degree of the executive	0.35	3.25^{**}	0.17	0.13	0.97	−0.04
	(0.78)	(0.01)	(0.76)	(0.92)	(0.35)	(0.91)
Native speaker status of the executive	−1.04	0.51	1.01	−2.10	−0.99	$-0.72^{\dagger}$
	(0.42)	(0.73)	(0.12)	(0.17)	(0.39)	(0.09)
Publicly traded status of the company	0.69	1.79	0.05	−1.90	1.34	−0.12
	(0.60)	(1.20)	(0.94)	(0.14)	(0.22)	(0.75)
Age of the executive	−0.04	−0.01	$0.04^{\dagger}$	−0.00	−0.04	−0.02
	(0.51)	(0.86)	(0.09)	(0.98)	(0.43)	(0.34)
Constant	53.45^{**}	67.82^{**}	61.31^{**}	48.77^{**}	77.05^{**}	40.21^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	522	522	522	522	522	522
R-squared	0.05	0.02	0.03	0.04	0.04	0.03

Note: The table contains the results from regression models using communal and agentic linguistic measures as dependent variables. Robust standard errors in parentheses. [1.]LIWC by Pennebaker et al. (2015). [2.]Jones (2016).

$^{\dagger}p < 0.10$

$^{*}p < 0.05$

$^{**}p < 0.01$

TABLE 4
Linguistic Style Differences among the Female Executives

Variable	Model 1 Authentic	Model 2 Tone	Model 3 Feminine Linguistic Style	Model 4 Analytic Thinking	Model 5 Clout	Model 6 Masculine Linguistic Style
Consumer discretionary	$-4.63^{\dagger}$	-1.18	-2.45^{*}	3.63	4.84^{*}	1.44^{*}
	(0.07)	(0.69)	(0.02)	(0.13)	(0.03)	(0.04)
Technology	$-4.95^{\dagger}$	-3.63	-1.94	2.90	$5.32^{\dagger}$	1.00
	(0.09)	(0.29)	(0.16)	(0.32)	(0.05)	(0.26)
Other sectors	-7.87^{**}	3.21	$-1.88^{\dagger}$	3.09	6.67^{**}	1.04
	(0.01)	(0.29)	(0.09)	(0.21)	(0.01)	(0.18)
MBA degree of the executive	-0.39	0.92	0.45	-2.05	0.56	-0.20
	(0.86)	(0.70)	(0.63)	(0.33)	(0.76)	(0.75)
Native speaker status of the executive	0.95	-1.67	2.13	-4.56	-1.03	-1.30
	(0.73)	(0.65)	(0.11)	(0.17)	(0.69)	(0.18)
Publicly traded status of the company	1.47	0.46	-0.28	-1.84	2.97	0.86
	(0.62)	(0.90)	(0.80)	(0.38)	(0.23)	(0.19)
Age of the executive	-0.13	-0.15	0.07	-0.16	-0.06	-0.06
	(0.23)	(0.21)	(0.11)	(0.14)	(0.54)	(0.10)
Constant	67.08^{**}	81.79^{**}	61.96^{**}	52.64^{**}	70.62^{**}	40.85^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	204	204	204	204	204	204
R-squared	0.04	0.03	0.05	0.04	0.06	0.05

Note: Table shows the results from regression models, including only the female executive sample. LIWC composite scores (Pennebaker, 2015) are the dependent variables of Models 1, 2, 4, and 5. Models 3 and 6 have dependent variables calculated with the algorithms from Jones (2016). Robust standard errors in parentheses.

$^{\dagger}p < 0.10$

$^{*}p < 0.05$

$^{**}p < 0.01$

TABLE 5
Main Effects of Gender on Linguistic Style

LIWC Category	Executives					General Public			
	Men		Women		Effect Size	Men		Women	
	Mean	*SD*	Mean	*SD*	*d*	Mean	*SD*	Mean	*SD*
1. Long words	16.76	2.66	16.29	2.66	0.18*	15.25	5.91	13.99	4.42
2. First-person plural pronouns	1.80	0.92	1.62	0.91	0.20*	1.07	2.12	1.17	2.15
3. Articles	6.76	0.81	6.37	0.73	0.50***	6.70	2.94	6.00	2.73
4. Prepositions	13.18	1.03	13.17	1.09	*n.s.*	12.88	2.64	12.46	2.44
5. Anger words	0.18	0.15	0.17	0.15	*n.s.*	0.65	0.92	0.61	0.81
6. Swear words	0.02	0.05	0.02	0.05	0.17†	0.17	0.44	0.09	0.25
7. Masculine linguistic style	38.71	3.68	37.64	3.84	0.29**	-	-	-	-
8. First-person singular pronouns	3.95	1.55	4.93	1.60	−0.62***	6.37	4.66	7.15	4.66
9. Social processes	11.91	1.75	12.08	1.78	*n.s.*	8.51	4.72	9.54	4.92
10. Cognitive processes	13.56	1.63	13.60	1.60	*n.s.*	7.17	2.82	7.35	2.57
11. Tentative	2.95	0.71	2.79	0.62	0.24**	2.54	1.57	2.54	1.43
12. Auxiliary verbs	9.88	1.13	10.05	1.09	−0.15†	-	-	-	-
13. Verbs	17.98	1.81	18.34	1.89	−0.19*	1.15	0.93	1.22	0.89
14. Positive emotions	3.28	0.67	3.47	0.81	−0.25**	2.41	1.40	2.49	1.34
15. Negative emotions	0.90	0.40	0.90	0.41	*n.s.*	1.89	1.56	2.05	1.65
16. Feminine linguistic style	64.43	5.55	66.15	5.64	−0.31***	-	-	-	-

Note: The table shows the results of the comparisons between male and female executives. LIWC categories belong to Masculine and Feminine linguistic style indices from Jones (2016). Effect size shows whether there is a meaningful difference between the executives. The general public column represents the scores from Newman et al. (2008), providing the source of comparison between the executives and the general public.

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

TABLE 6
Frequent N-Grams

	Uni-grams		Bi-grams		Tri-grams	
Rank	Male	Female	Male	Female	Male	Feale
1	*Said*	*Compani*	*Make_Sure*	*Make_Sure*	*Want_Make_Sure*	*New_York_Citi*
2	*Use*	*Busi*	*Littl_Bit*	*Littl_Bit*	*Spend_Lot_Time*	*Spend_Lot_Time*
3	*Kind*	*Said*	*Year_Ago*	*Feel_Like*	*Spent_Lot_Time*	*Want_Make_Sure*
4	*Actual*	*Got*	*Everi_Day*	*Everi_Day*	*Everi_Singl_Day*	*Spent_Lot_Time*
5	*New*	*Great*	*Feel_Like*	*New_York*	*New_York_Citi*	*Work_Realli_Hard*
6	*Everi*	*Team*	*Lot_Peopl*	*Year_Ago*	*Work_Realli_Hard*	*Veri_Young_Age*
7	*Custom*	*Talk*	*Peopl_Work*	*Lot_Peopl*	*Make_Sure_Peopl*	*Talk_Littl_Bit*
8	*Team*	*Women*	*Thing_Like*	*Realli_Import*	*Need_Make_Sure*	*Everi_Singl_Day*
9	*Great*	*Tri*	*Peopl_Want*	*Want_Know*	*Thi_Thi_Thi*	*Think_Realli_Import*
10	*World*	*New*	*High_School*	*Make_Decis*	*People_Feel_Like*	*Work_Everi_Day*
11	*Manag*	*Kind*	*Want_Know*	*Did_Know*	*Talk_Littl_Bit*	*Work_Life_Balanc*
12	*Product*	*Feel*	*New_York*	*Look_Like*	*New_York_Time*	*Yeah_Yeah_Yeah*
13	*Chang*	*Chang*	*Spend_Time*	*Peopl_Work*	*Think_Lot_Peopl*	*Women_Care_Global*
14	*Idea*	*Love*	*Make_Decis*	*High_School*	*Think_Realli_Import*	*Make_Sure_Peopl*
15	*Big*	*Actual*	*Like_Thi*	*Peopl_Think*	*Want_Spend_Time*	*Make_Thing_Happen*
16	*Better*	*Manag*	*Think_Thi*	*Ask_Question*	*Ask_Question_Like*	*Ask_Lot_Question*
17	*Happen*	*Mean*	*Year_Old*	*Veri_Import*	*Ask_Right_Question*	*Want_Feel_Like*
18	*Build*	*Mani*	*Look_Like*	*Realli_Want*	*Make_Lot_Money*	*Becaus_Feel_Like*
19	*Creat*	*Whi*	*Ask_Question*	*Peopl_Want*	*Tri_Make_Sure*	*Ask_Question_Like*
20	*Let*	*Life*	*Realli_Good*	*Know_Know*	*Peopl_Realli_Want*	*Think_Import_Thing*

Note: This table illustrates the most frequent top 20 word and word groups extracted with a count vectorizer. All words are in root form.

TABLE 7
Top 20 Features from the Random Forest Model

Rank	Token
1	*Women*
2	*Woman*
3	*Young*
4	*Guy*
5	*Love*
6	*Compani*
7	*Men*
8	*Feel*
9	*Use*
10	*Probabl*
11	*Custom*
12	*Mother*
13	*Softwar*
14	*Better*
15	*Famili*
16	*Someone*
17	*Role*
18	*Opportun*
19	*Children*
20	*World*

Note: This table illustrates the top 20 words that the random forest model uses to distinguish the linguistic samples of male and female executives. The words are in their root form.

TABLE 8
Robustness Checks

Variable	Model 1 Feminine Linguistic Style	Model 2 Masculine Linguistic Style
Female	1.55^{*}	-1.01^{**}
	(0.01)	(0.00)
MBA degree of the executive	-0.64	0.43
	(0.33)	(0.26)
Native speaker status of the executive	1.11	-0.61
	(0.14)	(0.17)
Publicly traded status of the company	-0.24	0.04
	(0.74)	(0.93)
Age of the executive	0.02	0.00
	(0.59)	(0.92)
Interview	9.90^{**}	-4.07^{**}
	(0.00)	(0.00)
Constant	56.63^{**}	43.60^{**}
	(0.00)	(0.00)
Observations	1,082	1,082
R-squared	0.17	0.10

Note: The table shows the models controlled for interview sources - robust standard errors in parentheses.

$^{\dagger}p < 0.10$
$^{*}p < 0.05$
$^{**}p < 0.01$

TABLE 9
Descriptive Statistics of the LIWC Content

LIWC Category	Interview Sources					Non-Interview Sources				
	Men		Women		Effect Size	Men		Women		Effect Size
	Mean	***SD***	**Mean**	***SD***	***d***	**Mean**	**SD**	**Mean**	**SD**	***d***
1. Common verbs	19.18	2.95	19.73	3.03	-0.19^{**}					
2. Comparisons	2.35	0.72	2.25	0.70	$0.13^{\dagger}$					
3. Interrogatives	2.33	0.76	2.52	0.78	-0.25^{***}					
4. Numbers	1.69	0.77	1.37	0.69	0.44^{***}					
5. Quantifiers	2.40	0.68	2.29	0.60	0.18^{*}					
6. Affective processes	4.32	1.06	4.60	1.11	-0.27^{***}					
7. Positive emotion	3.29	0.96	3.58	1.05	-0.28^{***}					
8. Anxiety	0.18	0.20	0.23	0.22	-0.23^{***}					
9. Anger	0.20	0.23	0.15	0.16	0.21^{**}					
10. Family	0.28	0.33	0.43	0.45	-0.40^{***}	0.27	0.44	0.43	0.80	$-0.26^{\dagger}$
11. Female references	0.18	0.37	0.53	0.62	-0.71^{***}	0.19	0.37	0.72	1.06	-0.75^{***}
12. Insight	2.99	0.93	3.10	0.88	$-0.12^{\dagger}$					
13. Causation						2.21	0.98	2.00	0.84	$0.23^{\dagger}$
14. Tentative	3.16	1.01	2.91	0.96	0.25^{***}					
15. Certainty	1.61	0.49	1.70	0.58	-0.17^{*}					
16. Perceptual processes	1.88	0.72	2.00	0.69	-0.16^{*}	1.85	0.94	1.61	0.89	0.27^{*}
17. See						0.81	0.57	0.66	0.52	$0.26^{\dagger}$
18. Feel	0.36	0.26	0.40	0.28	-0.16^{*}					
19. Biological processes	0.78	0.65	0.95	0.71	-0.26^{***}	0.90	0.84	1.51	1.53	-0.53^{***}
20. Body	0.20	0.20	0.23	0.20	$-0.14^{\dagger}$					
21. Health	0.31	0.40	0.37	0.52	-0.14^{*}	0.40	0.53	0.80	0.96	-0.55^{***}
22. Sexual	0.05	0.11	0.08	0.12	-0.18^{**}	0.05	0.14	0.09	0.20	$-0.26^{\dagger}$
23. Ingestion						0.17	0.24	0.40	1.24	-0.30^{*}
24. Risk	0.49	0.34	0.43	0.32	0.19^{**}	0.51	0.63	0.34	0.37	0.30^{*}
25. Past focus	4.18	1.70	4.61	1.85	-0.25^{***}					
26. Future focus	1.38	0.54	1.27	0.52	0.21^{**}	1.28	0.68	1.12	0.61	$0.26^{\dagger}$
27. Motion						2.42	1.07	2.12	0.81	0.31^{*}
28. Space	6.71	1.22	6.47	1.22	0.20^{**}					
29. Home	0.23	0.24	0.31	0.34	-0.28^{***}	0.27	0.45	0.51	0.81	-0.40^{**}
30. Money	1.28	1.13	0.97	0.86	0.30^{***}	1.90	1.42	1.55	1.35	$0.25^{\dagger}$
31. Death	0.05	0.09	0.04	0.08	$0.12^{\dagger}$					
32. Swear words	0.02	0.08	0.01	0.05	$0.14^{\dagger}$					
33. Nonfluencies						0.15	0.17	0.21	0.28	-0.28^{*}

Note: Table shows the descriptive statistics of LIWC content words. The effect sizes are calculated separately for the interview and non-interview sources.

$^{\dagger}p < 0.10$
$^{*}p < 0.05$
$^{**}p < 0.01$
$^{***}p < 0.001$

TABLE 10
Descriptive Statistics and the Effect Size of the Question Content with the LIWC Categories

LIWC Category	**Executives**				
	Male		**Female**		**Effect Size**
	Mean	***SD***	**Mean**	***SD***	***d***
1. Quantifiers	2.61	1.44	2.36	1.25	0.18*
2. Social processes	17.92	5.35	19.26	4.90	−0.26**
3. Family	0.32	0.62	0.41	0.67	−0.15†
4. Female references	0.07	0.29	0.53	0.84	−0.81***
5. Tentative	3.33	1.67	3.06	1.61	0.16†
6. Differentiation	2.65	1.49	2.38	1.52	0.18*
7. Drives	8.06	2.71	8.49	2.38	−0.17†
8. Achievement	2.86	1.54	3.25	1.58	−0.25**
9. Risk	0.28	0.46	0.20	0.37	0.18*
10. Past focus	4.42	2.17	4.97	2.10	−0.26**
11. Time	4.78	2.38	5.33	2.15	−0.24**
12. Money	0.95	1.13	0.67	0.92	0.26**

Note: The table shows the significantly different content words of interview questions asked to male and female leaders.

†$p < 0.10$

*$p < 0.05$

**$p < 0.01$

***$p < 0.001$

Personality Expression by Language among Business Executives

Nur Yaldiz

Abstract

The Social Role Theory postulates that leaders' public expression of images is shaped by what society expects of leaders, such as inspirational, charismatic, and reliable. These characteristics are linked to personality traits of extraversion and agreeableness. Accordingly, the images they choose to convey and how they communicate them are likely to reflect both the impression management strategies to fulfill the social expectations and their personalities. The literature points out that men and women express these personality traits differently, but it is unclear how the gender differences in personality expression materialize in leadership contexts. This study contributes to the interdisciplinary literature by studying a large sample of top business executives' personality expressions by employing language-based personality assessment, a new and expanding method in psychology and management literature. I test gendered differences in personality expression among top executives, thereby providing insights beyond the student and laypeople samples typical in the personality literature. The main findings are that among the relevant traits for leadership, female executives express more extraversion in general than male executives, and they do this by being authentic. Male executives tend to be more conscientious and open by using clout and analytical thinking words. The findings are consistent with gender expectations that align men with agency and women with communion.

Personality Expression by Language among Business Executives

Personality traits of business executives have been capturing the interest of the public and researchers. As the faces of the company, how business executives present themselves matters to both inside and outside of the organization (Schlenker, 2012). Leaders' public expression of images is shaped by what society expects of leaders, such as inspirational, charismatic, and reliable. These characteristics are linked to personality traits of extraversion and agreeableness. Accordingly, the images they choose to convey and how they communicate them are likely to reflect both the impression management strategies to fulfill the social expectations and their personalities. As personalities are latent constructs, surveys are the commonly employed methods to assess one's personality (e.g., Goldberg, 1990). There are two drawbacks of survey-based assessment. First, participants can manipulate their answers to create the desired profile, resulting in biased assessment (Roberts, 2009). Second, surveying a large number of top business leaders would be a challenge, so there are not many studies implementing surveys among executives (e.g., Adams & Funk, 2012; Wille et al., 2018).

As an alternative source of personality information, management literature has recently started to employ text samples, such as conference calls (e.g., Harrison et al., 2019) and social media status (e.g., Wang & Chen, 2020). Compared to surveys, spoken language is hard to deflate, likely to provide a more accurate picture of the genuine personality (Mehl et al., 2006; Pennebaker, 2011). The approach of using written and spoken text for personality assessment is by no means a new idea—Allport and Odbert already claimed in 1936 that linguistic data contain a person's most distinctive traits (Allport & Odebert, 1936). This notion later contributed to establishing the Big Five Personality approach (Goldberg, 1990).

The sociolinguistic literature provides the theoretical and empirical bases for linking language to personality traits (Mehl et al., 2006; Pennebaker & King, 1999; Tackman et al., 2020). For instance, extraverted people use more social words and first-person singular pronouns to talk about their stories. The diligent mindset of conscientious people reveals itself with proper language that excludes swear and sexual words. Formal and analytical thinking (e.g., articles, prepositions) indicates openness to experience. As a communal trait, agreeableness heavily relies on linguistic indicators describing others (e.g., first-person plural pronouns, third-person singular pronouns). Finally, various emotion-word groups display neuroticism (e.g., affective processes).

The recent advancement in computational power, availability of the massive amount of digital text data, and various aids in natural language processing made it possible to fully exploit the link between language and personality. Many social science studies currently employ a sociolinguistic approach combined with machine learning algorithms to investigate psychological processes, personal orientation, and personality traits. Although this literature is expanding, current studies samples from students (e.g., Mehl et al., 2006; Sun & Vazire, 2019) and adults (e.g., Tackman et al., 2020) situated in daily routine lives or through written data from social media (e.g., Schwartz et al., 2013; Wu & Zheng, 2019). Because the context in which the written and spoken texts are formulated matters (Stryker, 2007; Roberts, 2009), there is still a need for investigations into more diverse sources and samples.

This study utilizes text-based personality assessment to investigate the expression of personalities of top business executives. I collected unique data from 522 executives (318 men and 204 women) interviewed by the renowned *New York Times* columnist Adam Bryant in *the Corner Office* interview series[1] from March 2009 to October 2017. The *New York Times* interview data was supplemented with additional text data written or spoken by the same executives from other media sources. These executives operate in the United States but from

[1] https://www.nytimes.com/column/corner-office

different nationalities and socio-economic upbringing, and spanning across diverse industry sectors.

I employ the Big Five personality traits (extraversion, conscientiousness, openness to experience, agreeableness, and neuroticism), arguably the most well-used framework to understand the personality of leaders in the applied psychology literature (Bono & Judge, 2004; Judge & Bono, 2000; Judge et al., 2002; Judge et al., 2009). In general, extraversion, conscientiousness, openness, and emotional stability are found relevant for leader emergence and effectiveness (Judge et al., 2002), while agreeableness is a fundamental trait of transformational leadership (Judge & Bono, 2000).

I hypothesize that how executives express personalities is moderated by gender. Gender is one of the most salient social categories and is often the first categorization applied when meeting a new person, even before age or race (Eagly et al., 2000). The Social Role Theory postulates that leaders are subject to societal norms expected of leaders, but that does not exclude them from being subject to gender norms (e.g., men being agentic and women being communal) (Eagly & Karau, 2002). Studies show how the expression of the Big Five traits is perceived differently by gender. For example, family-related words (e.g., mother, daughter, dad) are linked to agreeableness for women but conscientiousness for men (Tackman et al., 2020). Similarly, discrepancy words (e.g., should, would) indicate extraversion for women, but extraversion for men is linked to more informal and sexual words and fewer discrepancy words. In Mehl et al. (2006), third-party judges perceive the expression of negative emotions (e.g., ugly, nasty, hurt) from women as neuroticism, wherein the same expressions by men are perceived emotionally stable. However, these studies are conducted using students and laypeople samples, so how these gender differences materialize in leadership contexts is unclear. Management and social psychology literature investigate how female leaders meet the societal gender expectations through communal display in a nonstereotypical context (e.g., Amanatullah & Morris, 2010; Heilman & Okimoto, 2007; Williams & Tiedens, 2016). This study supplements this literature by employing an alternative approach focusing on the expression of the Big Five personality traits.

The rest of the paper consists of the following sections. First, the Theoretical Framework introduces linguistic insights and personality expression in the leadership domain, followed by the hypotheses. Then, the Method section begins with the linguistic indicators of the Big Five Personality traits from Tackman et al. (2020). The Results section introduces the findings of the analyses. Further, the Discussion elaborates on the findings, and the Conclusion provides a general overview.

Theoretical Framework

Linguistic Expression of Personality

According to Allport and Odbert's (1936) lexical hypothesis, written and spoken data contain a person's most distinctive traits and further helped establish the Big Five approach (Goldberg, 1990). Pennebaker and King (1999) explored how people present themselves by analyzing 1.9 million words belonging to 2,479 essays from psychology students. Findings showed that written texts give a generic picture of personality. For example, people with neurotic personalities have higher usage of present-focused words (e.g., now, is), talk more about others, and use more social words (e.g., family, friends), implying the need for social support. Mehl et al. (2006) explored personality clues from spoken language by recording 96 students' interactions in short snippets through an Electronically Activated Recorder (EAR) that they wore for approximately two days. Third-party student judges rated the participants' personality traits by going through the transcripts of the recordings. They find that judges, in general, accurately assessed the subjects' personalities.

These two studies have shown that written and spoken language provide reliable clues about personality. However, it was unclear which data produced the most accurate indicators for personality assessment. Mairesse et al. (2007) filled the gap by employing various machine learning algorithms to understand whether written (data from Pennebaker & King, 1999) or spoken text (data from Mehl et al., 2006), and self-assessment (Mehl et al., 2006; Pennebaker & King, 1999) or observer ratings (Mehl et al., 2006) best predict personality. Results showed that observed personality features and spoken text were the most reliable.

Tackman et al. (2020) were the first to focus on gender differences in the Big Five personality traits and language use with an adult sample. The study aimed to replicate Mehl et al.'s (2006) findings with a more extensive sample of 462 participants from the United States *(N = 462, 59.3% female, age range 24 – 94 years, and 29.8% non-white)*. There was little evidence to support gender differences in how people exercise their personalities in everyday life. Still, they found 36 categories where personality expression is linked to different language use. Among those, family-related issues, avoiding inappropriate language, and the use of diverse emotion-words to express gratitude and affection are significant indicators of female agreeableness. However, the same indicators imply conscientiousness (not agreeableness) for men. Inappropriate and informal language is linked to male extraversion, but those for female extraversion are modest and polite language.

Expression of Personality in the Context of Leadership

Current literature shows that male and female leaders do not substantially differ regarding the relevant Big Five personality traits. These are extraversion (Adams & Funk, 2012; Eagly & Carli, 2007), conscientiousness (Eagly & Carli, 2007; Wille et al., 2018), and openness to experience (Eagly & Carli, 2007). However, female leaders appear to be more agreeable, as they are more sympathetic, helpful, and approachable than male leaders (Adams & Funk, 2012). In the same way, general public women tend to be more agreeable than men (Feingold, 1994). Similar to self-reported personality traits, the expression of personality in public may have traces of gendered expectations. Like ordinary men and women, how male and female leaders present themselves is still subject to descriptive norms (e.g., stereotypes; established expectations on what men and women do) and prescriptive norms (what men and women should do, Eagly & Karau, 2002).

However, these arguments may not directly imply a large gap between male and female executives regarding personality expressions. As current literature points out, female executives may integrate feminine traits into their image to meet social expectations (e.g., Kark et al., 2012; Koenig et al., 2011; Vroman & Danko, 2020). Hence, these women may mostly have comparable traits to male leaders while also displaying stereotypically feminine characteristics.

Hypotheses

Current literature finds no substantial gender differences in their Big Five personality traits (Adams & Funk; 2012; Eagly & Carli, 2007; Willet et al., 2018). However, different linguistic components build up the overall personality traits. Therefore, although the overall expression of personality traits might be similar, male and female executives can emphasize linguistic markers in specific personality expressions differently. Below, I elaborate on the possible differences between male and female executives for each personality trait and expression.

Extraversion is a core leadership characteristic (Judge et al., 2002; Judge et al., 2009) that plays a significant role in communication. Senior executives represent their organizations in public and impact the intra-organizational atmosphere. Thus, they are likely to have better social skills (Bono & Ilies, 2006) and charisma (Bono & Judge, 2004). Since being sociable comes as a natural leadership characteristic, executive men and women are less likely to differ. Indeed, Adams and Funk (2012) find that male and female managers are comparable in extraversion. However, Eagly and Carli (2007) suggest that female leaders express extraversion by being more approachable and warmer. In contrast, male leaders are likely to appear extravert by being more assertive and competitive.

Conscientiousness means setting clear expectations, having strong discipline, a clear direction (Bono & Judge, 2004) with an achievement-oriented view (Bono et al., 2014). Thus, conscientious leaders prefer to dominate their environment by being in charge (Judge & Bono, 2000). Although expressing strict conscientiousness may cause backlash towards female leaders, it is not likely that female leaders fall behind because of having a conscientious mindset (Eagly & Carli, 2007). Indeed, survey studies also show that male and female leaders are not dramatically different regarding conscientiousness (e.g., Adams & Funk, 2012; Wille et al., 2018).

Openness to experience in leadership means to have an unconventional attitude (Judge et al., 2002) and a transformational leadership style (Bono & Judge, 2004). Hence, open leaders are likely to support the intellectual growth of their followers. Similarly, these leaders aim to establish an environment conducive to creative mindsets and innovative ideas. Female leaders are relatively more divergent thinkers (Wille et al., 2018), and innovation outcomes are associated with female representation on executive boards (Dezsö & Ross, 2012). However, different study finds that boards with all female members focus less on research and development (Apesteguia et al., 2012). Therefore, indicators of openness may be related to diversity and not fixated on a specific group. Male and female leaders are likely to be comparably open.

Agreeableness is not strictly relevant to leadership (Judge et al., 2002; Eagly & Carli, 2007). However, Judge and Bono (2000) argue that the right level of agreeableness positively impacts the leader-follower relationship. Furthermore, agreeable leaders are idolized role models who promote cooperation over conflict and competition (Judge et al., 2009) with a friendly working environment (Eagly & Johnson, 1990). Female executives support fewer layoffs (Matsa & Miller, 2013), motivate female managers to become executives (Matsa & Miller, 2011), and invest more in social sustainability (Apesteguia et al., 2012). Furthermore, female leaders are frequently associated with leadership styles (e.g., servant, transformational) that emphasize stakeholder relationships (Bass & Avolio, 1994; Lemoine & Blum, 2021). Hence, female leaders tend to be more agreeable than male leaders.

Lack of **Neuroticism** is emotional stability, and emotionally stable leaders are being calm and collected in the face of a crisis, patient with employee adaptation, and rapidly recovering from failures (Judge et al., 2009). There may not be a dramatic difference between male and female leaders as emotional stability is an integral part of leadership (Judge et al., 2002; Kirkpatick & Locke, 1991). However, a display of anger is endorsed in male leaders (Brescoll & Uhlmann, 2008), wherein displaying emotion may appear as a sign of weakness for female leaders (Brescoll, 2016). Thus, the gender difference in expressing emotional stability could be about women showing less emotion than men to appear emotionally stable.

Method

Linguistic Dimensions

Each executive's Big Five personality traits are assessed based on the collected texts following Tackman et al. (2020). The cleaned texts are submitted to the Linguistic Inquiry and Word Count (LIWC), an extensively tested and validated dictionary. A text (sentences, paragraphs, or entire document) is cleaned and sorted into LIWC's 90 existing word categories based on 6,400 terms. Subsequently, LIWC summarizes the submitted text as the percentage of the total word count for each word category. In addition to the word categories, LIWC (2015) offers four linguistic summary variables; *analytic thinking*, *clout*, *authenticity,* and *emotional tone* (Pennebaker et al., 2015), which are also included in Tackman et al. (2020)'s personality traits analysis (excluding emotional tone). According to Tackman et al. (2020), among the 90 LIWC-generated variables, 57 are correlated to one or more of the Big Five personality traits and are theoretically justifiable. Below, I discuss linguistic markers for each Big Five personality trait, and the summary of this information is provided in the left columns (Big Five) with hypothesized directions (positive or negative) in Table 2.

Extraversion. Extraverts employing function words indicate higher cognitive processing and complex social skills (Chung & Pennebaker, 2008). All personal pronouns (e.g., her, mine), combined with conjunctions, common verbs (e.g., carry, go, drink), and auxiliary verbs, imply a person talking about their story (i.e., what has been going on in their lives; Pennebaker, 2011; Tackman et al., 2020). Extraverts focus on the present results in a simple linguistic style, explaining lower analytical thinking (Tackman et al., 2020). In contrast, long words indicate a sophisticated mindset (Tausczik & Pennebaker, 2010). Even though extraverts do not powerfully employ complex language, they own their story (i.e., first-person singular pronouns, personal pronouns) and share genuine details (Newman et al., 2003), supporting authenticity (Tackman et al., 2020).

Extraverts are upbeat and talkative individuals (Judge et al., 2009). They employ a diverse range of words (Mairesse et al., 2007; Pennebaker & King, 1999), pronouns, and perceptual process words (e.g., feeling, look, hear) to describe events and people (Tackman et al., 2020). They also tend to share a group identity, e.g., via social processes (e.g., friend, talk, us; Gill & Oberlander, 2002; Oberlander & Gill, 2006; Iacobelli et al., 2011; Tackman et al., 2020; Yarkoni, 2010), family (Mairesse et al., 2007; Mehl et al., 2006; Nowson & Gill, 2014; Pennebaker & King, 1999; Yarkoni, 2010), female references (e.g., mother, women, girl; Mairesse et al., 2007; Mehl et al., 2006), friends (e.g., neighbor, buddy; Mairesse et al., 2007; Pennebaker & King, 1999; Yarkoni, 2010), and affiliation (e.g., social, ally, friend; Tackman et al., 2020). Sad words are the only emotion-related category that appears significantly more in extroverts' speech (Tackman et al., 2020).

Conscientiousness. Conscientious people avoid word groups that might appear improper, including sex and sexuality (Mehl et al., 2006; Mairesse et al., 2007; Pennebaker &

King, 1999; Yarkoni, 2010), swear words (Mehl et al., 2006; Mairesse et al., 2007; Pennebaker and King, 1999; Tackman et al., 2020; Yarkoni, 2010) negative emotions (e.g., hurt, ugly) and anger words (Mehl et al., 2006; Mairesse et al., 2007; Pennebaker & King, 1999; Tackman et al., 2020; Yarkoni, 2010). This uptight style of conscientious people results in more work-related words with less interest in the details (i.e., adjectives). High clout score (Tackman et al., 2020) and focus on the task at hand resembles a leader attitude (i.e., fewer first-person singular pronouns with more first-person plural pronouns; Kacewicz et al., 2014; Pennebaker, 2011) with a focus on followers (i.e., third-person plural pronouns; Chung & Pennebaker, 2007). Nonfluencies (e.g., uh, umm; Mehl et al., 2006; Mairesse et al., 2007; Tackman et al., 2020) were used less by conscientious people. Instead of talking about different topics and engaging in a conversation, conscientious individuals share less (Tausczik & Pennebaker, 2010) and only disclose information about previous events (i.e., past focus (went, ago); Tackman et al., 2020).

Openness to experience. Although open people have various interests and have enthusiasm for new experiences (Costa & McCrae, 1992), linguistic indicators from Tackman et al. (2020) have shown that open people are not necessarily sociable individuals. Lack of total pronouns (e.g., his, she, us; Mehl et al., 2006; Mairesse et al., 2007; Pennebaker & King, 1999; Yarkoni, 2010), second-person pronouns (Yarkoni, 2010), social processes (Mehl et al., 2006; Mairesse et al., 2007; Pennebaker & King, 1999; Yarkoni, 2010), family words (Pennebaker & King, 1999; Yarkoni, 2010), and female references indicate that open people are less other-oriented (Tackman et al., 2020). However, lack of social orientation may not imply that open individuals have strong self-focus. Openness is negatively associated with personal pronouns (Tackman et al., 2020) and first-person singular pronouns (Pennebaker & King, 1999; Tackman et al., 2020; Yarkoni, 2010). The implication is that open people's intellect (Costa & McCrae, 1992) might not allow them to build social connections unless the social context consists of high intellectuality (McCrae, 1996). Consequently, open people have lower authenticity while scoring higher in analytical thinking (Tackman et al., 2020).

Analytical thinking, work, and death-related (e.g., kill, coffin) content words (Pennebaker & King, 1999; Yarkoni, 2010) reveal intellectual curiosity, formal thinking, and questioning of existential issues (Tackman et al., 2020). Articles (Pennebaker & King, 1999; Yarkoni, 2010) and conjunctions with fewer common verbs, auxiliary verbs, and negations (Tackman et al., 2020) also signal higher cognitive processing (Chung & Pennebaker, 2008). This philosophical side of open people might also push them not to focus on the present (Pennebaker & King, 1999; Yarkoni, 2010) or the future (e.g., will, may) with no reward (e.g., benefit, prize) or time orientation (Tackman et al., 2020). Biological processes (e.g., pain, blood, eat) and ingest (e.g., pizza, eat) appear less in open people's language, which may not directly indicate anything meaningful about openness (Tackman et al., 2020).

Agreeableness. Agreeable people employ words associated with affiliation and achievement (e.g., success, win) with fewer power words (e.g., bully, superior; Tackman et al., 2020). An indication might be that agreeable people enjoy being part of a group (i.e., first-person plural pronouns; Pennebaker & King, 1999; Tackman et al., 2020; Yarkoni, 2010) while not exercising power over group members (Pennebaker, 2011). Combined with achievement, a higher focus on the past (Tackman et al. 2020; Yarkoni, 2010) indicates that agreeable people might tend to reevaluate their previous accomplishments. Conjunctions serve as an ingredient to tell a coherent story (Tausczik & Pennebaker, 2010). When combined with hear-words (e.g., hearing, listen), third-person singular pronouns, and male references (e.g., dad, his, boy)[2], agreeable people show their prosocial and communal inclinations by focusing on others. Accordingly, agreeableness negatively relates to analytical thinking (Tackman et al., 2020).

[2]Male references are the only significant group in the agreeable language among all social process words. This outcome is most likely related to the sample employed in Tackman et al.'s (2020) study.

Higher usage of assents (e.g., yes, agree, OK) and nonfluencies (Tackman et al., 2020) implies agreement within an environment (Tausczik & Pennebaker, 2010). Indeed, agreeable people do not stick out with any improper language that might upset the harmony of a social context (i.e., swear words, anger words; Mairesse et al., 2007; Mehl et al., 2006; Pennebaker & King, 1999; Yarkoni, 2010; and sexual words; Mairesse et al., 2007; Mehl et al., 2006; Tackman et al., 2020).

Neuroticism. Various emotion words (i.e., affective processes) appear in the neurotic language (Pennebaker & King, 1999), including negative emotion words, anxiety (e.g., fearful, worry), anger (Pennebaker & King, 1999; Yarkoni, 2010), and feel-words (e.g., feel, touch; Tackman et al., 2020). Pronouns, including personal pronouns, second-person pronouns, and family words (Mairesse et al., 2007; Mehl et al., 2006; Pennebaker & King, 1999), appear more in the neurotic language, except for fewer third-person plural pronouns. Indeed, neurotic people rely on others for social and emotional support by spending more time talking with a higher word count and diverse content (i.e., dictionary words; Tackman et al., 2020).

Neurotic people have a high self-focus (i.e., first-person singular pronouns; Mairesse et al., 2007; Mehl et al., 2006; Pennebaker & King, 1999; Tackman et al., 2020; Yarkoni, 2010) and are prone to depression (Chung & Pennebaker, 2007; Pennebaker, 2011). These associations also explain lower analytic thinking (Tackman et al., 2020). Neurotic language may sound inappropriate and informal due to body-related words, including sexual words, netspeak (e.g., lol, thx, btw), and swear words. This casual style is also reflected in the less employed content words (i.e., work, time; Tackman et al., 2020), echoing life’s serious aspects. Despite the immature mindset, present and future focus (Pennebaker & King, 1999; Tackman et al., 2020) resembles older people’s linguistic styles. With age, people’s linguistic styles might become more complex or at least stay the same. Auxiliary verbs, negations (Tackman et al., 2020), and discrepancy (e.g., would, should; Tackman et al., 2020; Yarkoni, 2010) appear more in the neurotic language (Tausczik and Pennebaker, 2010). However, this imperfect picture of sophisticated language does not positively correlate with analytical thinking. A negative relationship with conjunctions and higher usage of common verbs and motion words (e.g., go, arrive, car; Tackman et al., 2020) indicate an incoherent style when discussing personal issues.

Sample

Executive Data. Of a total of 522 executives, 61% are identified as men ($N = 318$) and 39% as women ($N = 204$) based on the pronouns used in texts, information in Wikipedia, and common knowledge (famous person). The executives' ages ranged from 33 to 93 years, with a mean of 56.15 ($SD = 9.77$).[3] Seventy-three percent of male ($SD = 0.44$) and 87% ($SD = 0.33$) of female executives were natives of the United States.[4] The consumer discretionary sector accounted for 51% of female executives ($N = 81$). In contrast, female executives made up 22% of the technology leaders ($N = 32$). The communication sector has the second-largest share of female executives (43%, $N = 46$), followed by 39% from other industries ($N = 45$).

Text Data. The text data consisted of 1082 samples with 1,780,042 words ($M = 1{,}645.14$, $SD = 1{,}920.71$). From those 62% belonged to male executives ($SD = 1896.50$) and 38% to female executives ($SD = 1958.19$). The text sample comprised of 95% spoken ($M = 1{,}678.73$, $SD = 1{,}951.50$) and 5% written samples ($M = 1{,}052.14$, $SD = 1{,}111.88$).

Structure. The descriptive summary of the data is presented in Table 1. Interview data are collected from interviews in *The New York Times*, lifestyle magazines (e.g., *Cosmopolitan*), and business-related sources (e.g., *Forbes*). YouTube captions include talks from formal environments (e.g., TED Talks, keynote speeches) and casual interviews (e.g., podcasts videos). Podcast data were mainly from Kara Swisher's *Recode Decode*, Reid Hoffman's *Masters of Scale*, and *The Tim Ferriss Show*. Some executives do not appear in sources other than written

[3] When no birth year information was available, executives' age was determined based on the years of work experience and their most recent photo.

[4] The executives gave information about their native-US background during the interviews and talks. Industrial sectors were coded according to the Bloomberg Industry Classification system.

opinion pieces or published personal blogs. When these sources do not provide sufficient data, quotes from news reports and answers from survey-type interviews were added as additional sources.

Analysis

Following Tackman et al. (2020), I test the gender difference in LIWC word categories (including four summary variables) that correlate to each Big Five personality category by independent sample t-test. The results were interpreted in line with Cohen's (1992) suggestion for small (d = .20), medium (d = .50) and large (d = .80) effect sizes.

Results

Table 2 presents the results of gender comparisons in each LIWC category. The Dictionary words category in Table 2 shows that about 90% of the texts are categorized in at least one of the word categories. The numbers under columns M show the percentage of word count for that word category in relation to the total word counts in the document, except for the composite scores of analytical thinking, clout, and authenticity (these categories take a value between 0 and 100), word count, and dictionary words refer to the percentage of the words in a text that matches with LIWC dictionary. Among word categories, for example, total function words comprise 60% and 59% of female and male texts. The difference is statistically significant, with a moderate effect size (d = .31, between small and medium effect size). Note that positive effect size indicates women employ more total function words than men.

Female executives employ total pronouns (d = .34), personal pronouns (d = .43), and third-person singular pronouns (d = .23) that indicate extraversion. In contrast, male executives only use first-person plural pronouns (d = -.19) slightly more than female executives in their language. In terms of linguistic summary variables, female executives are moderately more authentic (d = .43), while male executives are more analytic thinkers (d = -.38). The hypothesis postulates that male and female executives do not differ in extraversion. The results indicate that female executives employ more extraversion-related word categories than male executives in their language, judging by the number of significant effects consistent with the higher extraversion.

For conscientiousness, as shown in Table 2, male and female executives differ in clout (*d* = -.35) and word groups positively associated with conscientiousness; first personal plural pronouns (*d* = -.19), past focus (*d* = .17), and negatively associated word groups (first-person singular pronouns (*d* = .62), biological processes (*d* = .28), sexual words (*d* = .18), swear words (*d* = -.16). These results show that female leaders employ negatively associated word-groups slightly more and positively associated word-groups slightly less than male leaders.

Among significantly different negative correlates of openness to experience, male executives only use slightly more future tense verbs (*d* = -.17). In contrast, female executives increasingly employ the rest of the categories. As shown in Table 2, male and female executives differ in the linguistic indicators of openness based on positively associated categories (analytical thinking (*d* = -.38), articles (*d* = -.50), conjunctions (*d* = .33)), and negatively associated categories (authenticity (*d* = .43), dictionary words (*d* = .40), total pronouns (*d* = .34), personal pronouns (*d* = .43), first person singular pronouns (*d* = .62), auxiliary verbs (*d* = .14), common verbs (*d* = .19), family (*d* = .33), female references (*d* = .86), biological processes (*d* = .28), future focus (*d* = - .17)). Other than future focus, female leaders employ most of the negatively associated indicators more than male leaders.

Agreeableness is the only category where male and female executives do not differ in most word groups. However, male executives use slightly more swear words (*d* = -.16) and have a moderate analytical tone (*d* = -.38), which is negatively associated with agreeableness. Among positive correlates, male executives only employ first-person plural pronouns more (*d* = -.19). Female executives use the rest of the significant and positive indicators in their language. As shown in Table 2, female executives employ the indicators of agreeableness somewhat more based on dictionary words (d = .40), third-person singular pronouns (d = .23), conjunctions (d = .33), past focus (d = .17).

Female executives employ most of the positively correlated and significant word groups of neuroticism (e.g., affective processes (d = .23), anxiety (d = .20), family (d = .48)).

Discussion and Conclusion

Assessing one's personality has mainly been done by survey studies, but recent advancements in natural language processing dramatically increase the possibility of using written and spoken samples to explore personality traits. Linguistic data provides a more realistic picture than self-administered personality inventories since survey questions are prone to self-bias. Linguistic data has long been subject to personality and social psychology research. Indeed substantial research in this field has produced the Big Five personality traits, which also serves as the main framework in applied psychology research exploring leader traits. These studies show that personality traits have relevance to leader emergence, effectiveness, and transformational leadership style. Therefore, the established research shows that personality traits have importance for leadership. However, no study has been done exploring linguistic data of top executives.

Sociolinguistic literature gives insights into linguistic indicators of personality traits from the daily and spontaneous recordings of student and adult samples. However, there is still a need for applications on different groups. This study aims to fill this gap by asking, "how do executives express their personality traits" and "how does gender moderate these traits?" Since obtaining private speech samples of top executives serves as a challenge, this study utilizes publicly available interview and speech samples with written text data to explore personality *expression*.

Substantial literature shows that gender mediates how people perceive and express personality traits. Stereotypical expectations encoded in the social subconscious interpret similar word groups differently from men and women. The findings from this study show that male and female leaders are no exception.

In expressing extraversion, female executives talk about their stories while focusing on other people (e.g., second-person pronouns, third-person singular, and plural pronouns). Lack of first-person plural pronouns suggests that female executives are more vulnerable during their interviews and talks. This finding indicates that female executives tend to be more sincere and warm with their interviewers (Pennebaker, 2011). Furthermore, female executives have lower clout scores. The clout-related words and word groups of male executives only contain first-person plural pronouns (e.g., we, us). In contrast, female executives talk about other people (e.g., she, he, women, someone) with authenticity, using first-person plural pronouns.

Similar to the expression of extraversion, findings of conscientiousness show that female executives have strong self-narrative with higher past focus. In addition, first-person singular pronouns, biological processes, and sexual words indicate that female executives tend to be less conscientious than male executives. However, these word groups may not necessarily mean inappropriate content borne out of negligence. The consumer discretionary sector has many female executives who work in the cosmetics industry. Consequently, these executives employ more biological processes while discussing their product lines and services. Sexual words may have different indications. For example, the term “sexy” could refer to something interesting. The word “sexual” appeared with “harassment” and “orientation” in female speech.

Text data may not provide thorough clues about openness (Mehl et al., 2006). In addition, talking about personal stories and focusing on others do not contribute to the expression of openness. Hence, female executives appear less open than male executives.

From social process words, family words may appear more in the female executives’ speech owing to the discussion of family responsibilities. Furthermore, female executives discuss inspirational family figures in their lives (e.g., mothers, grandmothers) while telling their stories. As indicated by conjunctions and past tense verbs, a simple linguistic style may also make female executives more approachable. Female executives appear to be more agreeable than male executives, focusing on other people (e.g., third-person singular pronouns). Lower analytical thinking scores also support that female executives focus less on tasks and objects but have a higher self-focus.

In addition to agreeableness, lower analytical thinking implies neuroticism. People low in emotional stability seek support from others to ease negative emotions and a depressed psyche, indicated by more pronouns and social words. Still, these implications may not directly interpret female executives' thinking. Anxiety and overall emotions are part of female executives' speech, reflecting their struggles while being part of a nonstereotypical environment. Moreover, female executives mainly discuss career challenges, being a working mother, or being the only female team member. This type of content also adds to the social words. Finally, first-person singular pronouns appear mostly when female executives describe their accomplishments; male executives mostly talk about similar subjects with first-person plural pronouns. Female executives may not consider themselves as part of an executive team. However, these indicators do not directly imply a depressed psyche (e.g., Mehl et al., 2012).

Overall, these findings indicate that female executives' personality expression comprises being more extraverted, agreeable but less conscientious, open, and emotionally stable. This personality expression partially overlaps with communal traits stereotypically associated with women.

Surely, these findings are not free from limitations with context and social expectations. Firstly, interviewers may bring up more female-related issues with female executives (e.g., paid leave, mentorship programs, female products). Words and word groups that describe these issues contribute to extraversion and agreeableness. Furthermore, female executives may receive more personal questions, increasing self-focus and thereby adding narrative style. In addition, female executives may speak and write about specific issues as part of their corporate image. Communication advisors are likely to craft linguistic samples which may produce a targeted image.

Future research may investigate more data to explore the differences between male and female executives. This approach would also provide insights into what extent executives' linguistic expressions overlap with or deviate from those of non-leader men and women. Another research outlet could be exploring personality expression over the course of a career. An executive's image may evolve, and it would be interesting to investigate a set of linguistic samples over an extended period of time.

References

Adams, R. B., & Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management science*, *58*(2), 219–235. https://doi.org/10.1287/mnsc.1110.1452

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*(1), i–171. https://doi.org/10.1037/h0093360

Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of personality and social psychology*, *98*(2), 256.

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, *58*(1), 78–93. https://doi.org/10.1287/mnsc.1110.1348

Bass, B. M., & Avolio, B. J. (1994). Shatter the glass ceiling: Women may make better managers. *Human resource management*, *33*(4), 549–560. https://doi.org/10.1002/hrm.3930330405

Bono, J. E., & Ilies, R. (2006). Charisma, positive emotions and mood contagion. *The Leadership Quarterly*, *17*(4), 317–334. https://doi.org/10.1016/j.leaqua.2006.04.008

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of applied psychology*, *89*(5), 901–910. https://doi.org/10.1037/0021-9010.89.5.901

Bono, J. E., Shen, W., & Yoon, D. J. (2014). Personality and leadership: Looking back, looking ahead. In D. V. Day (Ed.), *Oxford library of psychology. The Oxford handbook of leadership and organizations* (pp. 199–220). Oxford University Press.

Brescoll, V. L. (2016). Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *The Leadership Quarterly*, *27*(3), 415–428. https://doi.org/10.1016/j.leaqua.2016.02.005

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead?: Status conferral, gender, and expression of emotion in the workplace. *Psychological science*, *19*(3), 268–275. https://doi.org/10.1111/j.1467-9280.2008.02079.x

Chung, C., & Pennebaker, J. W. (2007). *The psychological functions of function words.* In K. Fiedler (Ed.), *Frontiers of social psychology. Social communication* (p. 343–359). Psychology Press.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of research in personality*, *42*(1), 96–132. https://doi.org/10.1016/j.jrp.2007.04.006

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Dezsö, C. L., & Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic management journal*, *33*(9), 1072–1089. https://doi.org/10.1002/smj.1955

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Harvard Business Press.

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological bulletin*, *108*(2), 233–256. https://doi.org/10.1037/0033-2909.108.2.233

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A., Wood, W., & Diekman, A. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender*, chapter 5. New York: Psychology Press.

Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological bulletin*, *116*(3), 429.

Gill, A. J., & Oberlander, J. (2002). Taking care of the linguistic features of extraversion. In W.D. Gray & C. D. Schunn (Eds), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pp. 363–368. Lawrence Erlbaum Associates.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, *40*(8), 1316–1330. https://doi.org/10.1002/smj.3023

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: the implied communality deficit. *Journal of applied psychology*, *92*(1), 81.

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In S. D'Mello, A. Graesser, B. Schuller, & J. C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 568–577). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24571-8_71.

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, *85*(5), 751–765. https://doi.org/10.1037/0021-9010.85.5.751

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780. https://doi.org/10.1037/0021-9010.87.4.765

Judge, T. A., Piccolo, R. F., & Kosalka, T. (2009). The bright and dark sides of leader traits: A review and theoretical extension of the leader trait paradigm. *The Leadership Quaterly*, *20*(6), 855–875. https://doi.org/10.1016/j.leaqua.2009.09.004

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kark, R., Waismel-Manor, R., & Shamir, B. (2012). Does valuing androgyny and femininity lead to a female advantage? The relationship between gender-role, transformational leadership and identification. *The Leadership Quarterly*, *23*(3), 620-640.

Kirkpatick, S. A., & Locke, E. A. (1991). Leadership: Do traits matter? *Academy of Management Perspectives*, *5*(2), 48–60. https://doi.org/10.5465/ame.1991.4274679

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological bulletin*, *137*(4), 616.

Lemoine, G. J., & Blum, T. C. (2021). Servant leadership, leader gender, and team gender role: Testing a female advantage in a cascading model of performance. *Personnel Psychology*, *74*(1), 3-28.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

Matsa, D. A., & Miller, A. R. (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review*, *101*(3), 635–639. https://doi.org/10.1257/aer.101.3.635

Matsa, D. A., & Miller, A. R. (2013). A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics*, *5*(3), 136–169. https://doi.org/10.1257/app.5.3.136

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, *120*(3), 323–337. https://doi.org/10.1037/0033-2909.120.3.323

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Nowson, S., & Gill, A. J. (2014). Look! Who's Talking?: Projection of Extraversion Across Different Social Contexts. *Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition* (pp. 23–26). https://doi.org/10.1145/2659522.2659530

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes, 42*(3)*,* 239–270. https://doi.org/10.1207/s15326950dp4203_1

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us.* Bloomsbury Press.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS one*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Schlenker, B. R. (2012). *Self-presentation.* In M. Leary & J. Tangney (Eds.), Handbook of self and identity (2nd ed.), 542–570. New York, NY: The Guildford Press.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Stryker, S. (2007). Identity theory and personality theory: Mutual relevance. *Journal of Personality*, *75*(6), 1083–1102. https://doi.org/10.1111/j.1467-6494.2007.00468.x

Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment?. *Psychological science*, *30*(3), 405-414.

Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley, S. A., Polsinelli, A. J., & Mehl, M. R. (2020). 'Personality in its Natural Habitat' Revisited: A Pooled, Multi–sample Examination of the Relationships between the Big Five Personality Traits and Daily Behaviour and Language Use. *European Journal of Personality*, *34*(5), 753–776. https://doi.org/10.1002/per.2283

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Vroman, S. R., & Danko, T. 2020. Against what model? Evaluating women as leaders in the pandemic era. *Gender, Work & Organization, 27*(5), 860-867.

Wang, S., & Chen, X. (2020). Recognizing CEO personality and its impact on business performance: Mining linguistic cues from social media. *Information & Management*, *57*(5), 103173. https://doi.org/10.1016/j.im.2019.103173

Wille, B., Wiernik, B. M., Vergauwe, J., Vrijdags, A., & Trbovic, N. (2018). Personality characteristics of male and female executives: Distinct pathways to success? *Journal of Vocational Behavior*, *106*, 220–235. https://doi.org/10.1016/j.jvb.2018.02.005

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin, 142*(2), 165–197. https://doi.org/10.1037/bul0000039

Wu, T., & Zheng, Y. (2019). Is Impression Management Through Status Updates Successful? Meta-Accuracy and Judgment Accuracy of Big Five Personality Traits Based on Status Updates From Social Network Sites in China. *Frontiers in psychology*, *10*, 1192.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Table 1 ***Mean Text Data Characteristics with the Number of Samples for each Media Outlet***

Source	Word count of female executives			Word count of male executives		
	N	*M*	*SD*	*N*	*M*	*SD*
The New York Times (March 2009–October 2017)	204	1,171.61	357.50	318	1,189.97	361.96
Alternative interviews (January 2003–June 2019)	105	1,167.96	612.99	130	1,508.00	1,415.98
YouTube captions (December 2004–May 2019)	69	2,965.03	3,660.82	107	2,694.68	2,375.64
Podcasts (November 2009–November 2019)	15	5,805.33	3,506.37	28	6,074.29	4,724.36
Quotes and single responses (February 2002–June 2019)	17	673.71	769.83	31	690.06	543.80
Essays (September 1998–May 2019)	18	927.50	584.81	40	1,108.22	1,283.50

Note: The table reports the descriptive statistics of word count based on gender. *N* represents the number of text samples from a specific source with a mean (*M*) and standard deviation (*SD*).

Table 2 ***Linguistic Categories of Big Five Personality Traits from Tackman et al. (2020)***

LIWC Category	Big Five					Female		Male		*t* (520)	*p*	Cohen's *d*
	E	C	O	A	N	*M*	*SD*	*M*	*SD*			
Word Count	+			+	+	3,499.191	2,995.305	3,479.349	2,635.043		n.s.	
Analytical Thinking	-		+	-	-	41.540	12.759	46.539	13.119	4.293	.000	-0.385
Clout		+				71.141	11.489	74.920	10.228	3.923	.000	-0.351
Authentic	+		-			56.382	13.627	50.934	11.889	-4.821	.000	0.432
Words > 6 letters	+					16.291	2.659	16.759	2.655	1.966	.049	-0.176
Dictionary words	+		-	+	+	91.944	2.058	91.113	2.044	-4.521	.000	0.405
Total function words	+					59.522	2.695	58.641	2.807	-3.555	.000	0.318
Total pronouns	+		-		+	18.144	2.075	17.422	2.090	-3.859	.000	0.346
Personal pronouns	+		-			11.110	1.746	10.358	1.686	-4.902	.000	0.439
1st person singular	+	-	-			4.925	1.601	3.950	1.545	-6.936	.000	0.622
1st person plural	+	+		+		1.620	.911	1.803	.920	2.218	.026	-0.199
2nd person			-			2.445	.877	2.560	.985		n.s.	
3rd person singular	+			+		.520	.438	.426	.369	-2.621	.009	0.235
3rd person plural		+				1.596	.596	1.617	.617		n.s.	
Articles			+			6.367	.728	6.759	.812	5.588	0.000	-0.501

Table 2 ***Linguistic Categories of Big Five Personality Traits from Tackman et al. (2020)*** **(continued)**

LIWC Category	Big Five					Female		Male		*t* (520)	*p*	Cohen's *d*
	E	C	O	A	N	*M*	*SD*	*M*	*SD*			
Prepositions			+			13.172	1.092	13.182	1.025		n.s.	
Auxiliary verbs	+		-		+	10.048	1.091	9.883	1.126	-1.657	.098	0.148
Common adverbs					+	5.876	.855	5.605	.852	-3.542	.000	0.317
Conjunctions	+		+	+	-	7.644	.866	7.364	.804	-3.768	.000	0.338
Negations			-		+	1.477	.416	1.458	.456		n.s.	
Common verbs	+		-		+	18.337	1.890	17.983	1.809	-2.139	.032	0.191
Common adjectives		-			+	4.115	.690	4.161	.716		n.s.	
Comparisons					+	2.231	.479	2.309	.500	1.783	.075	-0.159
Affective processes					+	4.469	.886	4.270	.793	-2.666	.007	0.239
Negative emotion		-			+	.902	.411	.903	.404		n.s.	
Anxiety					+	.214	.145	.182	.160	-2.330	.020	0.209
Anger		-		-	+	.168	.154	.183	.145		n.s.	
Sadness	+					.156	.143	.158	.122		n.s.	
Social processes	+		-			12.076	1.776	11.914	1.745		n.s.	
Family			-		+	.387	.349	.253	.221	-5.364	.000	0.481
Friends	+					.160	.145	.172	.153		n.s.	

Table 2 ***Linguistic Categories of Big Five Personality Traits from Tackman et al. (2020)*** **(continued)**

LIWC Category	Big Five					Female		Male		*t* (520)	*p*	Cohen's *d*
	E	C	O	A	N	*M*	*SD*	*M*	*SD*			
Female references	+		-			.509	.483	.190	.275	-9.595	.000	0.860
Male references				+		.503	.447	.512	.400		n.s.	
Discrepancy					+	1.761	.451	1.790	.455		n.s.	
Perceptual processes	+					1.917	.505	1.834	.499	-1.843	.065	0.165
Hear				+		.676	.316	.634	.319		n.s.	
Feel					+	.368	.182	.338	.167	-1.891	.059	0.169
Biological processes		-	-		+	.973	.576	.809	.559	-3.223	.001	0.289
Body		-			+	.223	.149	.216	.193		n.s.	
Health					+	.404	.391	.333	.362	-2.109	.035	0.189
Sexual		-			+	.067	.083	.052	.077	-2.051	.040	0.184
Ingestion			-			.194	.307	.158	.256		n.s.	
Affiliation	+			+		2.873	1.011	2.943	.996		n.s.	
Achievement				+		2.660	.722	2.622	.762		n.s.	
Power				-		2.744	.746	2.737	.721		n.s.	
Reward			-			1.845	.446	1.907	.510		n.s.	
Past focus		+		+		4.323	1.290	4.100	1.277	-1.934	.053	0.173

Table 2 ***Linguistic Categories of Big Five Personality Traits from Tackman et al. (2020)*** **(continued)**

LIWC Category	Big Five					Female		Male		*t* (520)	*p*	Cohen's *d*
	E	C	O	A	N	*M*	*SD*	*M*	*SD*			
Present focus			-		+	12.139	1.719	11.984	1.783		n.s.	
Future focus			-		+	1.227	.338	1.291	.368	1.987	.047	-0.178
Motion					+	2.136	.433	2.151	.467		n.s.	
Time			-		-	4.288	.735	4.203	.738		n.s.	
Work		+	+		-	5.085	1.438	5.119	1.382		n.s.	
Death			+			.042	.064	.045	.058		n.s.	
Swear words		-		-	+	.017	.045	.024	.047	1.864	.062	-0.167
Netspeak					+	.119	.121	.121	.134		n.s.	
Assent				+		.162	.176	.150	.165		n.s.	
Nonfluencies		+		+		.194	.153	.197	.154		n.s.	

Note: The table provides all LIWC word categories and their association with the Big Fiver personality traits, listed in their relevance to leadership (Extraversion, E; Conscientiousness, C; Openness, O; Agreeableness, A, Neuroticism, N). Positive associations are indicated with a + sign and vice versa. Empty spaces indicate no association. *d* scores are reported when there is a significant difference between male and female executives. *n.s.* refers to a nonsignificant result.