Gender bias through production about and memory for names

By

Bethany Gardner

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

In

Psychology

August 31, 2020

Nashville, Tennessee

Approved:

Sarah Brown-Schmidt, Ph.D.

Megan Saylor, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Out of the many ways we can refer to people—first names, last names, nicknames, and titles—why do we pick the ones that we do? Different ways of referring to people signal different information about status, distance, and opinion (Slobin et al., 1968; Atir & Ferguson, 2018; Takiff et al., 2001; Cowan & Kasen, 1984). Here, we investigate how the forms of reference used to describe a person guide inferences about gender.

We talk about women less

Despite knowing that about 50% of people and 40% of doctors are women (Misersky et al., 2014), our language use does not reflect these statistics. Boyce et al. (2019) presented participants with short stories that included gender-stereotyped role nouns, and then asked participants to write a continuation of the story (i.e. “After the shop on High Street closed for the night, a baker stayed to tidy up. Before the baker took out the trash…”). Using a measure of how often participants used she/her pronouns, Boyce et al. found that participants were less likely to refer to feminine referents in their sentence continuations than the distributional statistics about the role nouns (estimated in a separate norming study) would predict. A complementary finding probed memory for these referents, and likewise, they found that participants were less likely to recall the referents as female than would be expected, given participants’ estimates of the gender distributions.

In a related study during the 2016 US presidential election cycle, von der Malsburg et al. (2020) asked participants to complete a sentence about the next US president and measured participants' use of pronouns (i.e. "The next US president will be sworn into office in January 2017. After moving into the Oval Office, one of the first things that…"). They found that participants were less likely to write *she* than their beliefs that Hillary Clinton would be elected in 2016 would have predicted. At different time periods in 2015, participants estimated a 50-60% chance Clinton would win, but used *she* only around 10% of the time and *they* around 50% of the time. In a reading time study, von der Malsburg et al. found that despite these high beliefs that the next president would be female, participants also showed significant delays when reading sentences that contained *she* as compared to *he* and *they*. An auxiliary experiment found no reading time penalty for *she* vs. *he*, indicating that these results were driven by a higher difficulty to have *she* co-refer with *the president*, as opposed to *she* being intrinsically slower to process. Additionally, Hamilton (1988) found that when asked to write about a generic person (i.e. "Before a pedestrian crosses the street…") and then describe the person they imagined, participants imagined men 2x as often as women but were 2.5x as likely to use masculine names to refer to the characters. This again suggests that participants identify referents as female at significantly lower rates than they believe they are female.

Similar evidence of a bias to under-infer female referents has been reported when gender is unspecified. Davis Merritt & Kok (1995) asked participants to read about a gender-unspecified person ("Chris") in a text that did not contain pronouns. When asked at the end what gender they imagined the character as, over 75% answered male, despite other findings that “Chris” was the best example of a 50% masculine and 50% feminine name (see also Davis Merritt & Wells Harrison, 2006).

Separate evidence suggests that people’s estimates of how gender is distributed within different contexts generally reflects real world distributions. Misersky et al. (2014) asked participants to estimate the gender ratios in different occupations, in order to create norming data for studies about role nouns. Garnham et al. (2015) then compared these estimates to real-world statistics from UK government data and found a strong, positive correlation. While the positive correlation indicates that participants successfully used knowledge about the relative proportions of men and women in different professions, the data indicated that participants also overestimated men, particularly when the estimated and actual gender ratios strongly diverged.

One consideration is that the results discussed so far did not arise from participants being less likely to refer to women, but from using generic masculine language, where *he/him* refers to a generic person of any gender. The prescriptive use of the generic masculine form was contested by second-wave feminists, who argued that this language was not inclusive and perpetuated biases of masculine as the default (Bodine, 1975). Indeed, reading or hearing generic *he/him* (vs. *he or she* and *they)* exacerbates the tendency to overestimate men in tasks such as describing what images come to mind from a sentence or writing a story about a character from a prompt (Gastil, 1990; Moulton et al., 1978; for a review see Silveira, 1980). Producing generic *he/him* also has this effect: Hamilton (1998) guided participants to use generic *he/him* or inclusive alternatives like *he or she* and *they* with instructions about how formally they should write. Participants who used generic masculine language were more likely to imagine men in their stories and even more likely to label the characters as male than participants who used *he or she* or *they*. Constructions such as *he or she* and *they* lead to less over-representation of men, but results do not always approach 50% inclusion of women (e.g. neutral conditions are balanced in Moulton et al.,1978; but not in Gastil, 1990 and Hamilton, 1998). These effects are part of what Silveira calls the "people=male bias," an instance of generic=specific bias where the default person is male, and men are more default (belonging to an unmarked category). She argues that this cognitive bias is both reflected in and perpetuated by our language use, where generic masculine forms are used to represent all people (but masculine ones more fittingly) and feminine forms are strongly marked (Silveira, 1980).

We talk about women differently

These findings indicate that we are less likely to talk about women than the distributional statistics about names and occupations predict, and less likely to talk about women than our beliefs about those distributions predict. In addition, when people do talk about women, they do so in different ways than they talk about men. People are more likely to refer to men in professional contexts by their last names. Emerging findings indicate that how we choose to talk about women impacts how we think about women. For example, when scientists were referred to by last name, they were subsequently judged as more eminent, famous, and deserving of awards (Atir & Ferguson, 2018). Thus, the ways that we talk about people of different genders make men sound more important and successful, regardless of explicit beliefs about women's ability in science.

Other findings indicate that women are less likely to be referred to with titles (Dr., Professor) across a number of different contexts. In speaker introductions at medicine grand rounds, women used titles more overall and equally for men and women, while men introduced other men with titles and women by first name (Files et al., 2017). During the 2008 Democrat primary elections in the US, a corpus analysis of the first time Hillary Clinton was mentioned in TV news segments found that Clinton was more likely to be referred to by first name only than Barack Obama and other male politicians, who were typically introduced using titles and last names. This effect was primarily driven by male speakers and was observed regardless of polling status, news station ideology, or if politicians branded themselves by first rather than last name (Uscinski & Goren, 2011). In the classroom, male professors are more likely to be addressed by title (Rubin, 1981; Takiff et al., 2001; Stewart al., 2005). This has concrete effects on their perception: when students evaluated a transcript of a class introduction that manipulated the gender and form of address (e.g. “Jordan” vs. “Professor Smith”), professors who were referred to by title were afforded higher status (Takiff et al., 2001; Stewart et al., 2005). However, when female professors were referred to by title, they were perceived as less accessible; this double-bind between respect and accessibility was not found for male professors (Takiff et al., 2001).

It is, however, worth nothing that more informal terms of address do not always indicate less respect. Cowan and Kasen (1984), in a study of recommendation letters, found that letter writers used titles more often for women and first name more often for men, but the intent varied by the gender of the letter writer. Men using first names for other men was interpreted as solidarity, and men using titles for women was interpreted as distance. However, women using titles interpreted them as signaling status and respect regardless of gender.

The present study

One explanation for why referring to people by last name or title makes them seem more successful and important (Atir & Ferguson, 2018) is that it makes them, overall, seem more masculine. Since people make inferences about a referents' gender as soon as they are introduced into the discourse (Carreiras et al., 1996; Duffy & Keir, 2004; Kennison & Trofe, 2003), and revise inferences as information is added (Oakhill et al., 2005; Osterhout et al., 1997), it is possible that referring to someone by last name and/or title weakens the cue of femininity that a first name could provide. Since people tend to interpret gender-unspecified people as masculine (Davis Merritt & Kok, 1995; Davis Merritt & Wells Harrison, 2006), referring to people by their last names would bias the gender inference towards masculinity. Further, this may extend to referring to people by full names, in that the last name would decrease the level of femininity associated with the first name. If this is the case, we speculate that particular forms of reference highlight a referent's femininity (first name), while other forms decrease femininity (titles, last names, and possibly full names).

To begin to answer this question, we ask if the way that we refer to people affects inferences about that person's gender, first in a task that measures use of gendered pronouns (Experiment 1) and second in a task where participants are explicitly asked about the referent's gender (Experiment 2). We hypothesize that people will infer referents as male at higher rates than the gender distribution of the names would predict, and that inferences about gender will be shaped by the way in which those referents are introduced.

CHAPTER 2

Experiment 1: Production

The aim of Experiment 1 was to examine the relationship between how a character in a sentence is introduced (e.g. by their first, last, or full name) and inferences about that character's gender. We based our design closely on von der Malsburg et al. (2020), asking participants to read a sentence that introduced a character with a name (e.g. Jordan, Smith, or Jordan Smith) and then continued with a sentence fragment that invited a completion with a pronoun. We then used the gender information that was (or was not) carried in the pronoun as a measure of the participants' inferences about that character's gender.

Methods

Participants. 450 participants who completed the task on Amazon Mechanical Turk were included in the dataset, for 150 participants in each of the 3 between-subjects conditions. The sample size was selected a-priori based on von der Malsburg et al. (2020). Participants were required to be over the age of 18, be located in the US, have completed more than 100 Turk tasks with an acceptance rate of over 95%, and have started learning English before the age of 5. Participants were paid $1.50 for a task that took 10-15 minutes. A total of 574 participant responses were collected, and participants were excluded for having completed one of the study tasks before (6.62% of total responses), responding nonsensically (e.g., entering "good" for every question, 4.01%), reporting that they were not native English speakers (2.96%), or writing at the end that they guessed the study was about names and gender (7.67%).

Norming Study. In order to select a set of first names that range from feminine to androgynous to masculine, we conducted a norming study on a set of 90 names. 30 masculine and 30 feminine names were selected from lists of the most common names for assigned male at birth (AMAB) and assigned female at birth (AFAB)[1] babies in the US (US Social Security Administration, 2019). An additional 30 androgynous names were selected from Flowers (2015), who used US Social Security Administration data to identify names that were given at least one-third of the time to AFAB children and also at least one-third of the time to AMAB children. 50 participants on Amazon Mechanical Turk, following the same inclusion criteria as Experiment 1, were asked the rate the 90 names on a scale of 1-7, with 1 being "definitely masculine" and 7 being "definitely feminine." From these results, we selected 21 names to represent a range of names from masculine to feminine, with different levels of androgyny in between. The names and their average ratings are listed in the Appendix. The norming data were compared to US census data from 1930-2015 (US Social Security Administration, 2020; aggregated in Howard, 2016). The percentage given to AFAB children in the census data and the percent feminine (converted from 1-7 scale) from the norming data showed a strong positive correlation, $r(19) = .92$, $p < .001$.

Stimuli & Procedure. In addition to the 21 first names selected from the norming study, 21 last names were selected from a list of the most common surnames in the US (US Census Bureau, 2016). These names were presented to participants in short sentences (Appendix).

[1] We use assigned male at birth (AMAB) and assigned female at birth (AFAB) to indicate that these datasets only have information about what sex children were assigned at birth, not their gender identities later. For more information about current best practices for talking about gender, see GLAAD (2020).

The participants' task was to read a prompt containing a sentence and then complete the continuing fragment. We created 21 prompts that introduced a human character with a name and continued with a second (incomplete) sentence that was easiest to complete with a subject pronoun. The prompts did not include gendered pronouns, other names, or additional human characters. There were 3 between-subjects conditions that manipulated the type of name used to refer to the human character: *First Name, Last Name,* and *Full Name*:

(1) First: *Jordan woke up early to walk the dog. After making coffee...*

(2) Last: *Smith woke up early to walk the dog. After making coffee...*

(3) Full: *Jordan Smith woke up early to walk the dog. After making coffee...*

The mappings between names and production prompts were counterbalanced between participants by creating 3 lists within each condition. Each list had a different combination of first and last names; due to experimenter error, there was one combination that appeared in two lists (this item was included in the analysis) and one first name missing from one list in the Full Name condition. In addition to the critical stimuli, each participant saw 8 filler items that featured names of 26 US presidential 2020 candidates in May 2019. These fillers (8-9 per list) served two purposes: first, they were used as a distraction from the focus of the study. Second, they were used to pilot items for an unrelated study about forms of reference in political language.

Each participant saw 21 critical prompts and was asked to finish the sentence in a way that made sense to them. After completing the production task, participants were asked for basic demographic information: gender, age, race/ethnicity, and education level. The participant gender question was written as an open-ended response, following best practices for trans-inclusive study design (Zimman, 2017; Vincent, 2018). The study preregistration and stimuli can be found at https://osf.io/aypu2/.

Predictions

Prior studies indicate that when an utterance does not unambiguously convey gender, participants are more likely than not to assume that the character is male (Davis Merritt & Kok, 1995; Davis Merritt & Wells Harrison, 2006; Gastil, 1990; Hamilton, 1998; Moulton et al., 1978; Silveira, 1980). Yet in many situations, people are introduced using a first name, which provides probabilistic information about the person's gender. Thus, we asked if introducing a person using their first name would shape inferences about gender, and in particular if this bias to over-infer men would be attenuated. If probabilistic gender information given by first names is integrated into the gender inference, we would predict that (1) the rate of “She” responses would be higher when a first name was provided (First and Full Name conditions) compared to the Last Name condition. Within the First and Full Name conditions, we additionally predict that (2) the more female-biased the first name, the more likely that the character would be assumed to be female. We are particularly interested in the cases with androgynous first names, and predict that, like Boyce et al. (2019), participants will need a higher level than chance (names that are 50% masculine and 50% feminine) in order to assume the character is female.

A secondary question was whether (3) introducing a character with the full name (e.g. Jordan Smith) rather than the first name only (e.g. Jordan) would attenuate the influence of the gender information carried by the first name. This secondary question was motivated by the observation that, in English, it is more common to refer to men than women by their last names, and thus adding the last name may act as a cue to masculinity (Atir & Ferguson 2018).

Results

Responses that used he/him/his pronouns to refer to the named character were categorized together; hereafter we will refer these as “He” responses. Likewise, responses that used she/her/hers pronouns to refer to the named character were categorized together and will be referred to as “She” responses. Responses that did not use a gendered pronoun were coded as “Other”.

Table 1 shows the rates of “He”, “She”, and “Other” responses across all three conditions. For the First Name and Full Name conditions, the rates of “He” and “She” responses were roughly equal, following the balanced distribution of first names in our stimuli. In the Last Name condition, responses overwhelmingly bias towards “He.” Notably, in the Last Name condition, responses that gendered the character as female were slightly *less* common than “Other” responses that did not gender them.

	She	He	Other	Ratio of She vs He+Other
First	1381	1546	223	0.781
Full	1489	1478	130	0.907
Last	249	2580	321	0.086

Table 1. Numbers of “He”, “She”, and “Other” responses and ratios of “She” vs “He” and “Other” responses across all three conditions.

Responses were analyzed using logistic mixed-effect regression models using lme4 in R (Bates et al., 2015; R Core Team, 2018), predicting the likelihood of "She" responses as opposed to "He" and "Other" responses. Since our hypotheses involved the rate of "She" responses, these were coded as 1. "He" responses were coded as 0; "Other" responses were also coded as 0, as they were not frequent enough to be placed in a third category. Participant and Item were included as random intercepts, with items defined as the unique first, last and first + last name combinations. Because the condition manipulations were fully between-subject and between-item, fitting a random slope model was not possible. The fixed effect of Condition was coded with orthogonal Helmert contrasts (Table 2). The intercept in the model was significant ($\beta = -1.43$, $z = -4.65$, $p<.001$), due to overall more "He" than "She" responses. A significant effect of Condition was found for the comparison between Last and First/Full ($\beta= 2.82$, $z = 4.04$, $p<.001$), such that participants in the Last condition were less likely to produce "She" responses than participants in the First and Full conditions. The comparison between First and Full conditions was not significant.

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	**-1.423**	**0.306**	**-4.648**	**<.001**
Condition: Last (-.66) **vs First** (+.33) **+ Full** (+.33)	**2.817**	**0.697**	**4.043**	**<.001**
Condition: First (-.5) vs Full (+0.5)	0.6010	0.696	0.863	0.388
Random Effects	Variance	SD		
Participant	1.029	1.014	*9397 observations, 104 items, 450 participants*	

Table 2. Model results for effect of Condition on likelihood of "She" responses.

We conducted a separate analysis that included each first name's Gender Rating (based on the aforementioned norming study). This analysis included the First and Full Name conditions only, as we did not have gender ratings for the Last Name condition. Figure 1 and Figure 2 show the proportions of “He”, “She”, and “Other” responses for the First Name condition and the Full Name condition by the gender rating of the first name. As the rating of the name becomes more feminine, “She” responses increased and “He” responses decreased. Notably, however, “She” responses do not surpass “He” responses until the first name in the prompt is biased somewhat feminine, rather than at the midpoint on the scale. In the mostly feminine range of first names (5-6 on the scale), “He” responses outnumbered “Other” responses. In the mostly masculine range (2-3 on the scale), “She” responses occurred at similar rates as “Other” responses.

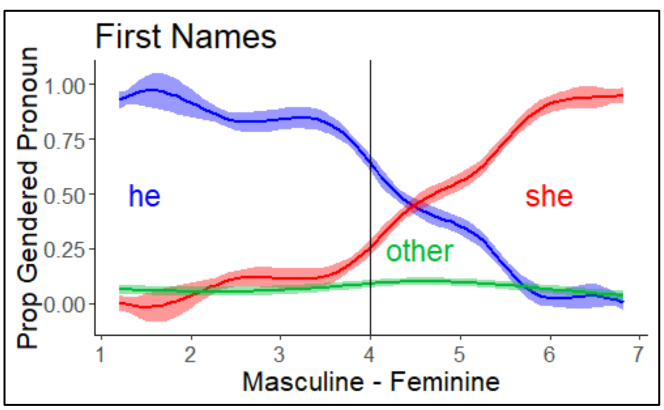

Figure 1. Proportions of “He”, “She”, and “Other” responses in the First Name condition by the gender rating of the first name.

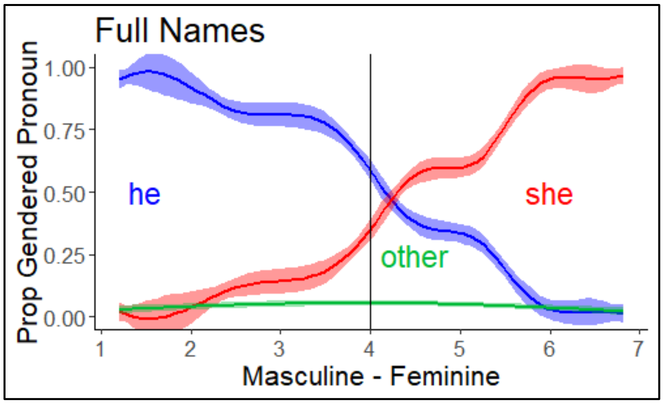

Figure 2. Proportions of "He", "She", and "Other" responses in the Full Name condition by the gender rating of the first name.

The Gender Rating for each first name was mean-centered, with positive numbers being more feminine and negative numbers being more masculine (Table 3). Condition was coded with mean-centered contrasts as before. The intercept was significant (β = -0.51, *z* = -4.30, p<.001), due to overall more "He" than "She" responses. The main effects of Condition (β = 0.51, *z* = 2.14, p<.05) and Gender Rating (β = 1.58, *z* = 22.07, p<.001) were both significant, such that participants produced more "She" responses in the Full Name condition and as the rating of the names became more feminine. The interaction between Condition and Gender Rating was not significant.

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	**-0.510**	**0.119**	**-4.299**	**<.001**
Condition (First=-.5, Full=.5)	**0.508**	**0.238**	**2.138**	**<.05**
Gender Rating (centered, fem +, masc -)	**1.584**	**0.072**	**22.069**	**<.001**
Condition * Gender Rating	-0.179	0.138	-1.300	0.19
Random Effects	Variance	SD		
Participant	0.888	0.942	*6247 observations, 83 items, 300 participants*	
Item	0.483	0.695		

Table 3. Model results for effects of Condition and Gender Rating on likelihood of "She" responses in the First and Full Name conditions.

Supplemental Analyses. A supplemental analysis that included participant gender as a covariate revealed that neither the effect of participant gender nor its interaction with Condition were significant after correcting for multiple comparisons.

Of all participant responses, 7.17% (674) were categorized as "Other" (Figure 3). While these responses were not numerous enough to analyze using inferential statistics, we describe their distribution across categories to characterize the dataset. "Other" responses fell into several categories: Repeated Name responses repeated the name of the character and thus did not provide further information about the participant's inference about the character's gender (e.g. *Jordan woke up early to walk the dog. After making coffee...Jordan sat down to read the news*), Null Subject responses had no grammatical subject (e.g. *Jordan woke up early to walk the dog. After making coffee…sat down to read the news*), Other Subject responses talked about other characters or the environment (e.g. *Jordan woke up early to walk the dog. After making coffee...it started to rain*), and Singular They responses used they/them pronouns to refer to the named character (e.g. *Jordan woke up early walk the dog. After making coffee...they sat down to read the news*). Singular They responses were distinguished from uses of plural they (Other Subject) by the context of the prompt.

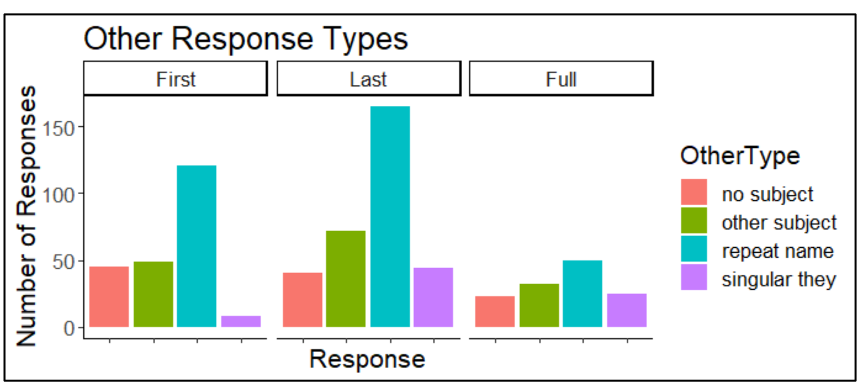

Figure 3. Number of "Other" responses (674) by type across conditions.

Discussion

We investigated whether the form of reference (first name, last name, or full name) affected people's inferences about a character's gender, measured through the gendered pronouns they used to complete a sentence about the character. We predicted that (1) the rate of "She" responses would be higher when a first name was provided (First and Full Name conditions) compared to the Last Name condition, and the data were consistent with this prediction. We also observed that when participants were not given gender information (Last Name condition), participants overwhelming assumed that the referent was male. Moreover, in this condition participants were approximately equally likely to *not* use a gendered pronoun at all ("Other" responses), as they were to use she/her pronouns.

We also predicted that (2) the more female-biased the first name, the more likely that the character would be assumed to be female. The data were also consistent with this prediction: probabilistic cues to the referent's gender did shape inferences, with many more "She" responses when a first name was given. However, the "He" bias persisted such that a name needed to be more strongly feminine (i.e. rated as female 60% in the norming study) for participants to preferentially refer to that referent with "She." In addition to this bias, participants also showed a pattern of asymmetry for mostly masculine and mostly feminine names. Across all three conditions, mostly masculine names (i.e. Chris) are most often referred to as "He", but mostly feminine names (i.e. Jackie) are "Other" and not "She."

Lastly, we had hypothesized that (3) introducing a person with a first and last name attenuate the gender cue from the first name, such that that the bias towards "He" responses would be greater in the Full Name condition as compared to the First Name condition. The data were not consistent with this prediction; instead, in the primary analysis the bias towards “He” responses was numerically larger in the First Name condition.

These patterns suggest that other factors besides knowledge of the real-world distribution of gender and names are at play in participants’ production decisions here. One possibility is that there is a masculine bias in knowledge of the gender associations of first names, such that participants believe names are more masculine than the actual distributions are. This could cause participants to produce “She” responses at lower rates than predicted by the actual distribution from census data, but would not imply that there is a difference between how often people believe a referent is female and how often they call a referent female. This is unlikely to explain the current data, however, as comparing the data from our norming study (used to establish the Gender Rating measure of the names we used) to the US census data indicated the two had a strong positive correlation, $r(19) = .92$, $p < .001$. Moreover, when the norming data differed from the census data, it was not always in the direction of over-estimating the masculinity of a name.

Another possibility is that identifying referents as male is easier than identifying them as other genders. This is one of the implications of the “people=male” hypothesis (Silveira 1980): if the generic person is a man, then producing a “He” response (the unmarked category) might be faster, easier, or require a lower threshold of evidence than producing a “She” response (the marked category) or an “Other” response (avoiding a categorization). Similarly, if men are more prototypical people, it may be easier and faster to categorize a referent as male.

An alternative explanation of the “He” response bias is the generic masculine usage, where speakers of American English were taught to use *he/him* to refer to referents of unknown or unspecified genders (Bodine, 1975). This has been replaced in formal language policies by *he or she* and *they* constructions (e.g. for APA standards: “Guidelines”, 1997; “Singular They”, 2020), but some speakers may retain the generic masculine usage. If so, some instances of “He” responses in the data may reflect this generic use. Of note is that the generic masculine is interpreted as specifically masculine, not gender neutral (Silveira, 1980; Hamilton, 1998; Gastil, 1990; Moulton et al., 1978). In Experiment 2, we ask participants to make explicit inferences about gender in order to address this alternative interpretation of the observed “He” response bias.

CHAPTER 3

Experiment 2: Memory

The aim of Experiment 2 was to examine the relationship between how a character in a story is referenced (e.g. by their first, last, or full name) and later explicit judgements about that character's gender. Participants read a series of short stories that introduced a human character with a name. Then after a brief delay, participants were asked about the gender of the characters in each of the stories from memory. If the results of Experiment 1 are driven by generic masculine language, where participants produced "He" responses but did not necessarily infer the referent as male, then the gender inferences in Experiment 2 would not show a male response bias as compared to underlying gender distribution of the names. If the "He" responses in Experiment 1 were representative of participants' actual gender inferences, we would expect to see a similar male response bias when participants are specifically asked to make gender inferences.

Methods

Participants. Participants were recruited on Amazon Mechanical Turk. The sample size (1350 planned) was determined a-priori based on Boyce et al. (2020), who used a similar task. Participants were required to be over the age of 18, be located in the US, have completed more than 100 tasks with an acceptance rate of over 95%, and have started learning English before the age of 5. They were paid \$1.50 for a task that took 10-15 minutes. A total of 1534 participants were recorded across the three conditions. Participants were excluded for having completed one of the study tasks before (4.63% of total responses), indicating that they were not native English speakers (1.30%), failing the attention check by writing nonsense text (2.48%), or not understanding the task and answering each recall question with the name instead of a gender (3.32%). Unlike Experiment 1, participants were not excluded for guessing the study was about gender, since this task explicitly asked about it. After these exclusions, the final sample (N=1352) included 451 participants in the First Name condition, 449 in the Last Name condition, and 452 in the Full Name condition.

Stimuli & Procedure. The names were combined into three conditions as in Experiment 1 (*First Name, Last Name, Full Name*). Unlike Experiment 2, participants saw two-sentence stories that referred to a character by name twice; the stories did not contain any gendered pronouns, e.g.:

> *Jordan was walking from the train station to work when it started to rain. Jordan had forgotten an umbrella and was annoyed to get so wet.*

Participants saw a total of 7 critical stories, which described everyday actions selected to avoid strong gender stereotypes (e.g., making coffee, walking a dog). To counterbalance the names and stories across conditions, three subsets of names and three mappings of names and stories were created, for a total of nine lists within each of the three conditions (First Name, Full Name, Last Name). Each participant was randomly assigned to one experimental list. In the First and Full Name conditions, names were distributed evenly between lists across the gender ratings from masculine to feminine. For the Last Name condition, since there was no gender rating data associated with the names, lists were randomly created. The combinations of first and last names for the Full Name condition were identical to Experiment 1, with the exception that we corrected an error in the Experiment 1 lists where a duplicate name appeared. As in Experiment 1, the names of 26 US presidential 2020 candidates acted as filler items to pilot a separate study, with each participant seeing 1 of these of these items.

After reading each story, participants were asked to type the name of the character as an attention check. Participants then completed 16 simple math questions as a distraction task. Participants were given a summary of the main action in each story and asked to type the gender of the character into a free response box (i.e. *What was the gender of the person who got caught in the rain?*). The free response box allowed participants to express uncertainty (i.e. "gender wasn't specified" or "I can't remember"). Critically, the memory prompt referenced the action and not the name. Finally, participants answered basic demographic questions about their gender (open-ended question design), age, race/ethnicity, and education level. The procedure is shown in Figure 4. The study preregistration and stimuli can be found at https://osf.io/aypu2/.

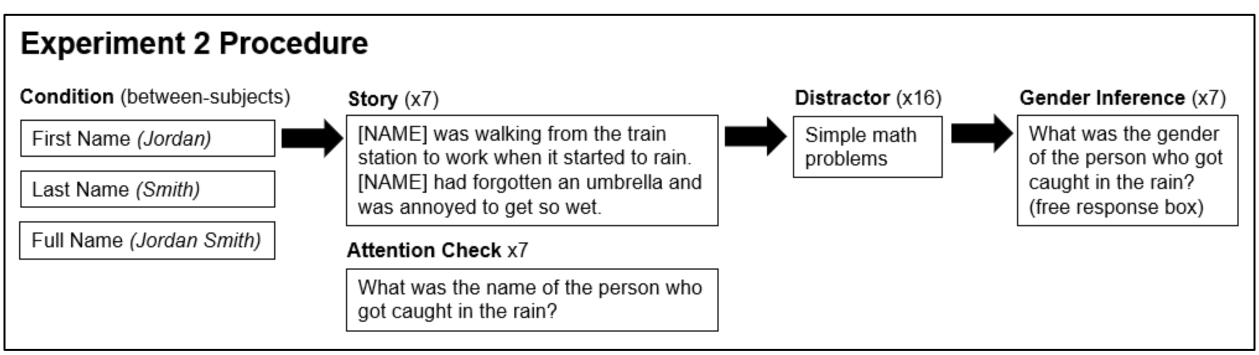

Figure 4. Experiment 2 procedure.

Predictions

If the “He” response bias in Experiment 1 was driven by a tendency to overestimate the presence of male referents and underestimate the presence of female referents, the following predictions are generated for Experiment 2: (1) as in Experiment 1, the rate of “She” responses will be higher when a first name is provided (First and Full Name conditions) compared to the Last name condition; (2) this bias will be attenuated when probabilistic information about gender is provided by a First Name. We are again particularly interested in the cases with androgynous first names and predict participants will need a higher level than chance (names that are rated 50% masculine and 50% feminine) in order to assume the character is female.

Alternatively, if the results of Experiment 1 were driven primarily by the use of generic masculine language, “He” responses did not necessarily reflect an inference that the referent was male. If so, we would expect to find limited evidence of a male response bias in Experiment 2, where gender is asked about explicitly.

Results

Responses were coded as recalling the named character as Male (e.g. “m”, “man”, “male”), Female (e.g. “f”, “woman”, “female”), or Other (e.g. “It wasn’t specified”, “I don’t remember”). As in Experiment 1, responses were analyzed using logistic mixed-effect regression models using lme4 in R (Bates et al., 2015; R Core Team, 2018), predicting the likelihood of Female responses as opposed to Male and Other responses. Since the conditions were fully between-participant and between-item, Participant and Item (each first, last, and first-last combination) were included as random intercepts only

	Female	Male	Other	Ratio of Female vs Male + Other
First	1561	1543	53	0.978
Full	1427	1632	105	0.821
Last	399	2479	265	0.145

Table 4. Numbers of “Female”, “Male”, and “Other” responses and ratios of “Female” vs “Male” and “Other” responses across all three conditions.

in the statistical models. The fixed effect of Condition was coded with orthogonal Helmert contrasts.

Table 4 shows the rates of gender recall as Male, Female, and Other across the three conditions. As in Experiment 1, the rates of recall as male and female are roughly equal in the First and Full Name conditions, following the balanced distribution of the first names, but there was a bias towards recalling the character as male in the Last Name condition. The intercept in the model (Table 5) was significant ($\beta = -0.89$, $z = -5.90$, $p<.001$), due to a bias towards Male responses overall. The main effect of Condition was significant for the contrast between Last vs First/Full ($\beta = 1.99$, $z = 5.83$, $p<.001$), such that participants were less likely to call the named character Female in the Last name condition compared to the First and Full Name conditions. The contrast between First and Full conditions was not significant.

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	**-0.888**	**0.151**	**-5.904**	**<.001**
Condition: Last (-.66) **vs First** (+.33) **+ Full** (+.33)	**1.993**	**0.342**	**5.833**	**<.001**
Condition: First (-.5) vs Full (+0.5)	-0.233	0.344	-0.676	0.50
Random Effects	Variance	SD		
Participant	0.205	0.452	*9464 observations, 105 items, 1352 participants*	

Table 5. Model results for effect of Condition on the likelihood of recalling the character as female.

As in Experiment 1, the effect of Gender Rating was analyzed for the First and Full Names conditions (see Table 4). The First Name condition is shown in Figure 5 and the Full Name condition in Figure 6. As the rating of the name becomes more feminine, "She" responses increased and "He" responses decreased. However, "She" responses do not surpass "He" responses until the first name in the prompt is biased somewhat feminine, rather than at the midpoint on the scale.

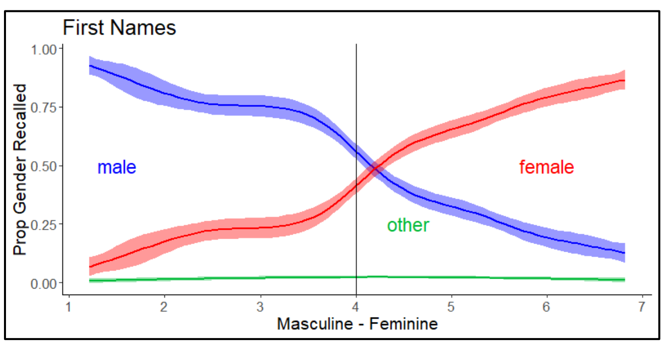

Figure 5. Proportions of recall as Male, Female, or Other in the First Name condition by the gender rating of the first name.

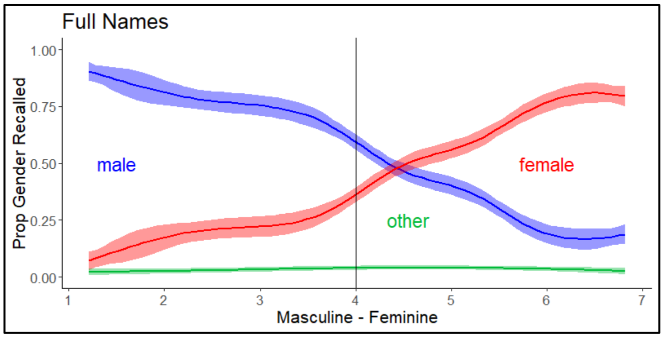

Figure 6. Proportions of recall as Male, Female, or Other in the Full Name condition by the gender rating of the first name.

The model results are shown in Table 6. The Gender Rating for each first name was centered, with positive numbers being more feminine and negative numbers being more masculine. The intercept term was significant (β = -0.21, *z* = -3.45, p<.001), indicating that participants were less likely to recall the referent gender as Female than as Male & Other in the First and Full Name conditions. The main effect of Condition was trending (β = -0.22, *z* = -1.88, p=.06), such that rates of recall as Female were lower in the Full Name condition than First. The main effect of Gender Rating was significant (β = 0.78, *z* = 21.99, p<.001), with participants being more likely to call the named character Female as the name ratings became more feminine. The interaction between Gender Rating and Condition was not significant.

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	**-0.206**	**0.060**	**-3.453**	**<.001**
Condition (First=-.5, Full=.5)	-0.223	0.112	-1.875	0.06
Gender Rating (centered, fem +, masc -)	**0.751**	**0.036**	**21.988**	**<.001**
Condition * Gender Rating	-0.074	0.070	-1.062	0.288
Random Effects	Variance	SD		
Participant	0.125	0.354	*6321 observations, 83 items, 903 participants*	

Table 6. Model results for the effects of Condition and Gender Rating on likelihood of recalling the character as Female in the First and Full Name Conditions.

Supplemental Analyses. The effect of participant gender and its interactions were not significant after correcting for multiple comparisons. The rate of Other responses (4.47%) was lower than Experiment 1, resulting in too few observations to analyze.

Discussion

One possible explanation of the “He” response bias in Experiment 1 was that participants were using generic masculine language, in which case using he/him pronouns would not necessarily imply inferring the referent as male. To evaluate this interpretation, we investigated if participants would continue to show a bias to overestimate male referents and underestimate feminine referents when the inference about gender was made explicitly.

As in Experiment 1, participants’ judgements about the gender of characters introduced in short narratives exhibited a male bias. We predicted that (1) this male bias would be strongest in the Last Name condition and (2) attenuated when the character was introduced including a first name (First and Full conditions), and the results were consistent with these predictions. Again, participants did not begin recalling the character as female 50% of the time at the midpoint androgynous names, but when the names were biased feminine. The results were overall similar to those in Experiment 1, though with smaller effects. In Experiment 1, participants were 4.15 times more likely to produce a “He” than a “She” response. In Experiment 2, participants were 2.43 times more likely to recall characters as male than female. Comparing the Last Name condition to the First and Full Name conditions, participants were 16.73 times more likely to produce a “She” response in the First and Full Name conditions (in Experiment 1) and 7.34 times more likely to recall characters as female in the First and Full Name conditions (in Experiment 2). This suggests that participant responses were not entirely driven by generic masculine production rules; instead, inferences about the gender of the referent per se influenced how the characters were referenced and remembered.

CHAPTER 4

General Discussion

Prior studies have shown that people are less likely to refer to feminine referents than the distributional statistics would predict, instead requiring extra evidence that a referent could be feminine before they begin to refer to them as such (Boyce et al., 2019; von der Malsburg et al., 2020). We also know that gender impacts the form of reference: women are less likely to be referred to by last name or title, and this impacts how they are perceived (Files et al., 2017; Uscinski & Goren, 2011; Rubin, 1981; Takiff et al., 2001; Stewart al., 2005). In particular, when scientists were referred to by their last name, they were judged more eminent, successful, and deserving of awards (Atir & Ferguson, 2018). We hypothesize that these two findings may be related: one reason people referred to by their last name are evaluated more highly may be because, on average, they are perceived as more masculine. We asked if the form of reference—first, last, or full name—affects participants' inferences about a referent's gender through an implicit sentence production task (Experiment 1) and an explicit memory task (Experiment 2). We find that that referents are less likely to be inferred as feminine than the distributional statistics about names predict. Moreover, we find that this bias is modulated by the form of reference used to name the characters.

When participants were only provided with the referent's last name, participants overwhelming refer to and recall the referent as male. This parallels other results showing that a gender-unspecified or generic person is largely inferred as male (Davis Merritt & Kok, 1995; Davis Merritt & Wells Harrison, 2006; Gastil, 1990; Moulton, 1978;

Silveira, 1980). It is also worth noting the magnitude of this "people=male bias" (Silveira, 1980) is roughly the same as results from 20-30 years ago, despite recent social advances in women's rights and nonbinary visibility. When cues to gender are given, as in the First and Full Name conditions, participants underestimate the frequency of women as compared to the actual gender distribution of the first names. This finding is consistent with prior findings from Boyce et al. (2019) on role nouns and von der Malsburg et al. (2020) on presidential candidates.

What is the source of these biases? Multiple causal pathways are possible, including speaker knowledge of gender distributions, speaker inference about a

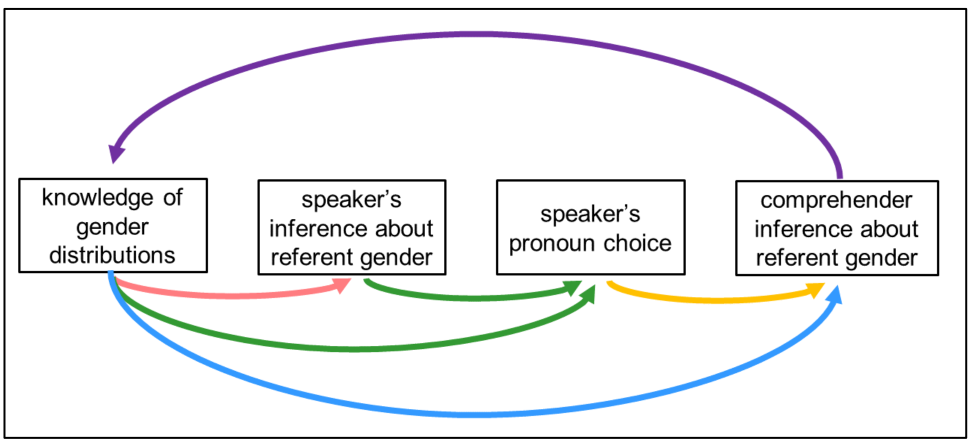

Figure 7. Sources of bias to refer to women less often than the gender distribution would predict.

referent's gender, speaker pronoun choice, and comprehender inference about referent gender (Figure 7).

First, we can assume that speakers know from experience that around half of people are women. Speakers also have information about the rates of women in different professions: Garnham et al. (2015) found a strong, positive correlation between the relative rates of women in the UK employment data and participants' estimates of about role nouns in Misersky et al. (2014). Garnham et al. do note, however, that in the cases where actual and estimated data did strongly diverge, participants overestimated the presence of men in a given occupation, rather than overestimating women. Thus, knowledge of underlying gender distributions in contexts like occupations and names may be biased to underestimate women.

Knowledge of the gender distribution for a first name or an occupation is used to make an inference about the gender of a specific referent (red arrow in Figure 7). Studies examining gender-unspecified referents show that people do not infer women 50% of the time and are instead biased to assume a generic or gender-unspecified referent is male (Silveira, 1980; Gastil, 1990; Moulton, 1978). This parallels the results from the Last Name condition in both experiments here: when our stimuli provided no explicit information about gender, about 80% of responses in both experiments infer the referent as male. Similarly, prior results suggest that when some gender information is given, either through an occupation or a first name, participants consistently underestimate the frequency of feminine referents, particularly for occupations that are not strongly gender-stereotyped and for names that are androgynous (Boyce et al., 2019; Davis Merritt & Kok, 1995; Davis Merritt & Wells Harrison, 2006). Thus, the results from the First and Full Name conditions extend these results to show that participants need more than a 50% chance that a referent is female in order to infer them as female 50% of the time.

From the knowledge of gender distributions in general and the inference about the gender of a specific referent comes the choice of what gendered pronouns a speaker will use. The results of von der Malsburg et al. (2020) indicated that 50-60% of participants believed Hillary Clinton would win the 2016 election, but only 10% of them used *she* to refer to the next president. Similarly, Hamilton (1998) finds that participants infer a generic referent as female at higher rates than they chose a feminine name to refer to them. These results suggest that inference about referent gender and gendered language choice are separable phenomena. The experiments here cannot distinguish between the contributions of knowledge of the gender distribution and inference about a specific referent's gender (green arrows in Figure 7), only conclude that this process introduces a bias to underestimate women.

It is worth noting, however, that the choice to use gendered language does not have to reflect certainty about a referent's gender. Participants in both experiments show low rates of overt hedging, such as using they/them pronouns or avoiding pronouns entirely in the production task and responding "I don't know" or "It wasn't specified" in the memory task. This does not mean, however, that responses using gendered language came from a place of uniform certain about the referents' genders. Participants may have still retained some uncertainty about the gender inference, or been slower to come to a decision about what language to use. This is especially likely for speakers of dialects where some manners of overtly expressing uncertainty, such as singular they/them pronouns for a referent with an unknown or unspecified gender, are not available. Future work could explore how the same language produced may reflect underlying levels of confidence.

Another potential source of bias is in the use of knowledge of gender distributions to form inferences about the referent's gender. It is possible that comprehenders know that speaker pronoun choice is biased masculine and correct for this bias (weighting the blue arrow over the yellow one). Boyce et al. (2019) suggests that this is not the case. They report data from a sentence completion task using role nouns, and showed a masculine bias in pronoun use as compared to the normed gender distributions of those occupations. A separate gender recall task showed participants stories that included two repetitions of the role noun and one gendered pronoun. When making gender inferences in the recall task, participants did not correct for masculine bias in pronoun use, and instead continued to recall the referents as feminine at lower rates than the normed gender distribution of the role nouns.

Finally, it is possible that inferences about gender in comprehension influence underlying beliefs about gender distributions (purple arrow). As such, speaker's choices about how to refer to entities in the world (orange arrow) may drive patterns in language comprehension (MacDonald, 2013). Thus, if speakers consistently under-refer to women and comprehenders do not correct for this bias, beliefs about general gender distributions may then become biased to underestimate women.

These results have potential implications for how we talk about women, particularly in professional and academic spheres. When we refer to people, we choose between different combinations of forms including pronouns, first names, last names, gendered titles (Mr./Mrs./Ms.), and nominally ungendered titles (Doctor, Professor). If certain forms of reference make feminine referents less likely to be inferred as feminine, should this influence how we choose to refer? On one hand, given that prior results find that people are judged more competent and successful when they are talked about using masculine-coded forms (Atir & Ferguson, 2018), a strategic speaker or writer could reference a female referent using masculine-coded forms to encourage a more masculine interpretation of the referent. This could mean reaping potential advantages (e.g. in perceived “eminence”), but potentially at the cost of having someone’s femininity be diminished or unacknowledged, and in perpetuating language production patterns that in turn may shape biases in comprehension. Alternatively, it may be preferable to work to change the underlying tendency to underestimate the presence of women, to emphasize a referent’s femininity especially in contexts where women are less visible.

Appendix

Stimuli

Table 7. First names, listed from most masculine to most feminine, with the mean and SD of each name's gender rating (1 as "definitely masculine", 7 as "definitely feminine") from the norming study (N=50).

Name	Mean	SD
Matthew	1.21	0.74
Brian	1.24	0.75
James	1.28	0.61
Chris	2.12	1.27
Tommie	2.41	1.63
Emerson	2.61	1.44
Stevie	3.16	1.53
Quinn	3.75	1.60
Reese	3.87	1.67
Taylor	4.22	1.14
Riley	4.34	1.35
Jessie	4.39	1.27
Kerry	4.73	1.29
Blair	5.22	1.53
Jackie	5.34	1.13
Jodie	5.59	1.22
Elisha	5.86	1.83
Ashley	6.24	1.15
Mary	6.73	0.86
Rebecca	6.78	0.85
Emily	6.82	0.73

Table 8. Last names.

Baker	Cooper	Smith
Bell	Green	Turner
Brooks	Hill	Walker
Brown	King	Ward
Campbell	Miller	White
Collins	Moore	Wright
Cook	Parker	Young

REFERENCES

Atir, S., & Ferguson, M. J. (2018). How gender determines the way we speak about professionals. *Proceedings of the National Academy of Sciences*, *115*(28), 7278–7283. https://doi.org/10.1073/pnas.1805284115

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01

Boyce, V., von der Malsburg, T., Poppels, T., & Levy, R. (2019). Remember him, forget her: Gender bias in the comprehension of pronominal referents. *CUNY Conference on Human Sentence Processing*.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A(3), 639–664.

Davis Merritt, R., & Kok, C. J. (1995). Attribution of gender to a gender-unspecified individual: An evaluation of the people = male hypothesis. *Sex Roles*, *33*, 3–4.

Davis Merritt, R., & Wells Harrison, T. (2006). Gender and ethnicity attributions to a gender and ethnicity-unspecified Individual: Is there a people = white male bias? *Sex Roles*, *54*, 787–797. https://doi.org/10.1007/s11199-006-9046-7

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge, *32*(4), 551–559.

Files, J. A., Mayer, A. P., Ko, M. G., Friedrich, P., Jenkins, M., Bryan, M. J., … Hayes, S. N. (2017). Speaker introductions at internal medicine grand rounds: Forms of address reveal gender bias. *Journal of Women’s Health*, *26*(5). https://doi.org/10.1089/jwh.2016.6044

Flowers, A. (2015). The most common unisex names in America: Is yours one of them? Retrieved from https://fivethirtyeight.com/features/there-are-922-unisex-names-in-america-is-yours-one-of-them/

Flowers, A. (2015). Unisex Names Data. Retrieved from https://github.com/fivethirtyeight/data/tree/master/unisex-names

Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters’ gender made elaboratively? *Memory & Cognition*, *30*(3), 439–446.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, *23*(11–12), 629–643. https://doi.org/10.1007/BF00289252

GLAAD. (2020). GLAAD Media Reference Guide - Transgender. Retrieved May 21, 2020, from https://web.archive.org/web/20200522040917/https://www.glaad.org/reference/transgender

"Guidelines for Nonsexist Language in APA [American Psychological Association] Journals," *American Psychologist*, June, 1977, pp. 487-94.

Gygax, P., Garnham, A., & Doehren, S. (2016). What do true gender ratios and stereotype norms really tell us? *Frontiers in Psychology*, *7*(JUL), 2019. https://doi.org/10.3389/fpsyg.2016.01036

Hamilton, M. C. (1988). Using masculine generics: Does generic he increase male bias in the user's imagery? *Sex Roles*, *19*(11–12), 785–799. https://doi.org/10.1007/BF00288993

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *23*(3).

Lieberson, S., Dumais, S., & Baumann, S. (2000). The instability of androgynous names: The symbolic maintenance of gender boundaries. *American Journal of Sociology*, *105*(5), 1249–1287.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00226

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., … Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871. https://doi.org/10.3758/s13428-013-0409-z

Moulton, J., Robinson, G. M., & Elias, C. (1978). Sex bias in language use: "Neutral" pronouns that aren't. *American Psychologist*, 1032–1036. https://doi.org/10.1037/0003-066X.33.11.1032

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, 33(6), 972–983.

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition* (Vol. 25).

Pilcher, J. (2017). Names and doing gender: How forenames and surnames contribute to gender identities, difference, and inequalities. *Sex Roles*, 77, 812–822. https://doi.org/10.1007/s11199-017-0805-4

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rubin, R. B. (1981). Ideal traits and terms of address for male and female college professors. *Journal of Personality and Social Psychology*, 41(5), 966–974. https://doi.org/10.1037/0022-3514.41.5.966

Silveira, J. (1980). Generic masculine words and thinking. *Women's Studies International Quarterly*, 3(2–3), 165–178. https://doi.org/10.1016/S0148-0685(80)92113-2

Singular "They." APA Style (2020). Retrieved from https://apastyle.apa.org/style-grammar-guidelines/grammar/singular-they

Slobin, D. I., Miller, S. H., & Porter, L. W. (1968). Forms of address and social relations in a business organization. Journal of Personality and Social Psychology, 8(3), 289–293. https://doi.org/10.1037/h0025657

Stewart, T. L., Berkvens, M., Engels, W. A. E. W., & Pass, J. A. (2003). Status and likability: Can the "mindful" woman have it all? *Journal of Applied Social Psychology*, *33*(10), 2040–2059. https://doi.org/10.1111/j.1559-1816.2003.tb01874.x

Takiff, H. A., Sanchez, D. T., & Stewart, T. L. (2001). What's in a name? The status implications of students' terms of address for male and female professors. *Psychology of Women Quarterly*, *25*, 134–144.

United States Social Security Administration (2019). Top names over the last 100 years. Retrieved from https://www.ssa.gov/oact/babynames/decades/century.html

United States Social Security Administration (2020). Beyond the Top 1000 Names. Retrieved from https://www.ssa.gov/oact/babynames/limits.html

Uscinski, J. E., & Goren, L. J. (2011). What's in a name? Coverage of Senator Hillary Clinton during the 2008 Democratic primary. *Political Research Quarterly*, *64*(4), 884–896. https://doi.org/10.1177/1065912910382302

Vincent, B. W. (2018). Studying trans: recommendations for ethical recruitment and collaboration with transgender participants in academic research. Psychology and Sexuality. https://doi.org/10.1080/19419899.2018.1434558

von der Malsburg, T., Poppels, T., & Levy, R. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 US and 2017 UK election. *Psychological Science*, 31(2), 115–128.

Zimman, L. (2017). Transgender language reform: some challenges and strategies for promoting trans-affirming, gender-inclusive language. *Journal of Language and Discrimination*, 1, 84–105. https://doi.org/10.1558/jld.33139