Estimating the Prevalence of Gender-Biased Language in Undergraduates' Everyday Speech

Heather J. MacArthur

Furman University

Jessica L. Cundiff

Missouri University of Science and Technology

Matthias R. Mehl

The University of Arizona

Author Note

Heather J. MacArthur, Department of Psychology, Furman University; Jessica L. Cundiff, Department of Psychological Science, Missouri University of Science and Technology; Matthias R. Mehl, Department of Psychology, The University of Arizona

Correspondence concerning this manuscript should be addressed to Heather J. MacArthur, 206D Johns Hall, Department of Psychology, Furman University, Greenville, SC. Email: heather.macarthur@furman.edu

Abstract

Research has shown that language can be gender-biased; however, little research has investigated the prevalence of this bias in everyday speech. Using recordings sampled from undergraduates' daily conversations, we investigated two forms of gender bias: paternalism through use of the infantilizing label *girl* to refer to women and androcentrism through a tendency to use more masculine (e.g., *man, guy*) than feminine (e.g., *girl, woman*) labels in everyday speech. U.S. participants ($n = 175$) wore the Electronically Activated Recorder (EAR), a device that recorded sound samples from their environments for 30 seconds every 12.5 minutes, for up to 4 days. Verbatim transcripts were then analyzed for instances of commonly used labels for females and males (e.g., *girl, woman, boy, man*). Results indicated that the label *girl* surpassed all other labels for women, as well as *boy* labels for men. We also found evidence of a masculine-label bias: Participants used masculine labels more frequently than feminine labels overall. These findings indicate the need for future research to investigate the potential consequences of infantilizing and androcentric language as well as the need for teachers, professors, clinicians, and practitioners of all types to be mindful of how their speech may include, exclude, or infantilize people based on gender.

Keywords: gender, gender studies, gender inequality, language & languages, masculinity, sexism in language, non-sexist language, gender-fair language, male bias, androcentrism

Estimating the Prevalence of Gender-Biased Language in Undergraduates' Everyday Speech

Scholars have identified several ways in which language can be gender-biased, noting how words, phrases, or discourse can demean, ignore, or stereotype women or men (for reviews, see Bigler & Leaper, 2015; Crawford, 2001). Given that language has the potential not only to reflect thought, but also to shape perceptions of the self and others (Bruckmüller, Hegarty, & Abele, 2012; Gaucher, Friesen, & Kay, 2011; Henley, Gruber, & Lerner, 1988), it is critical to understand the language landscape in which people operate on a daily basis. Yet there is a paucity of research examining how often people actually use gender-biased language in their everyday conversations. In the present paper, we examine gender bias in everyday language, focusing on two key forms of modern linguistic bias: paternalism and androcentrism.

Research has shown that the use of gender-biased language shifts with social norms over time (Bodine, 1975; Twenge, Campbell, & Gentile, 2012). Because more overt and hostile forms of prejudice have generally declined over the past several decades (Donnelly et al., 2016; Spence & Hahn, 1997; Twenge, 1997) while implicit sexism has continued to be problematic (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Swim, Aikin, Hall, & Hunter, 1995), it is particularly important to examine how subtle forms of sexism may remain embedded in our everyday language use. Two key forms of sexism that have been identified as harmful, yet that often go unnoticed by both users and targets, are paternalism and androcentrism (Bem, 1993; Eagly & Kite, 1987; Glick & Fiske, 1996).

Paternalism is one component of what Glick and Fiske (1996) refer to as benevolent sexism, a subtle form of bias in which people feel positive emotions toward women, yet still view them as weak and in need of men's assistance. *Androcentrism* refers to the assumption that men and masculinity represent the default form of humanity, whereas women and femininity are

secondary (Bem, 1993). Research on gender bias in language has highlighted the existence of both paternalism and androcentrism in common English language practices (Bailey & LaFrance, 2017; Gastil, 1990; Halbert & Latimer, 1994; Lerner, 1976; Messner, Duncan, & Jensen, 1993; Weatherall, 2002). Yet, with a few notable exceptions (Halbert & Latimer, 1994; Messner et al., 1993; Prentice, 1994; Twenge et al., 2012), research has rarely documented these forms of bias in everyday speech. Fewer still have addressed such questions across a range of contexts in everyday life, with most prior naturalistic studies addressing very specific uses of language, such as language in written classwork (Prentice, 1994) or in sports commentary (Halbert & Latimer, 1994; Messner et al., 1993).

We address this issue by examining the prevalence of two types of gender-biased language in everyday speech. First, we investigated paternalism in language by examining the prevalence of an infantilizing label to describe women, specifically, use of the label *girl* to refer to adult women. Second, we investigated androcentrism by examining whether there is an overarching bias toward masculine language, such that people use masculine labels (e.g., *man*, *boy*) more often than feminine labels (e.g., *woman*, *girl*) in everyday speech.

Although we suspected that these two types of gender-biased language occur often in people's daily conversations, no research to our knowledge has yet documented their actual prevalence. The purpose of the current research was to empirically document the frequency of their occurrence and to provide a lay of the land, so to speak, of the language landscape that young adults navigate in their daily lives. Our goal is to provide a broader context to research on gender-biased language by empirically documenting its prevalence in people's everyday conversations. We review research on these two types of gender-biased language in the following.

Paternalism and Infantilizing Labels

As we described previously, benevolent sexism refers to ideologies about gender that involve positive emotional feelings toward women, yet rest on assumptions about women being very different from, and ultimately weaker than, men (Glick & Fiske, 1997). One component of benevolent sexism is *protective paternalism*, the notion that because of their greater strength, power, and status, men should protect and provide for women, who might otherwise be incapable of fending for themselves. Research shows that benevolent sexism (e.g., a man assuming that a woman needs help carrying items) is often not recognized as sexist (Barreto & Ellemers, 2005; Swim, Mallett, & Stangor, 2004), and may even be seen favorably by women (Bohner, Ahlborn, & Steiner, 2010). Yet benevolent sexism and protective paternalism nevertheless have harmful effects. Research has shown, for example, that being exposed to benevolent sexism can increase women's body surveillance and feelings of body shame (Shepherd et al., 2011), decrease the importance that women ascribe to their own academic achievement and feelings of competence (Barreto, Ellemers, Piebinga & Moya, 2010), decrease women's performance on job-relevant problem-solving tasks (Dardenne, Dumont, & Bollier, 2007), and decrease women's willingness to challenge discrimination and unjust social systems (Becker & Wright, 2011; Jost & Kay, 2005).

One form of paternalism that has previously been identified in research on linguistic sexism is the infantilization of women through use of the label *girl*. Indeed, *girl* is defined as "a female child from birth to adulthood" (Meriam-Webster, 2017), yet examples spanning multiple decades suggest that *girl* is commonly used to refer to female adults as well as female children. For example, the popular television program "Girls" features four college graduates in their mid-twenties, and the prime-time show "New Girl" refers to an elementary school teacher in her early

thirties. In the 1970s, Lerner (1976) noted that psychotherapists frequently used the term *girl* to

describe female clients in clinical settings, and research on televised sports commentary in the

1990s similarly found that female athletes were commonly referred to as *girls*, whereas male

athletes were never referred to by the infantilizing male equivalent, *boys* (Halbert & Latimer,

1994; Messner et al., 1993). Other research has shown that people tend to believe that the upper

age limit for *girl* targets is older than the upper age limit for *boy* targets (Bebout, 1984), making

it likely that women are labeled *girl* more often than men are labeled *boy*.

      Colloquial labels such as *girl* may seem inconsequential, yet even minor changes in

wording can critically influence people's perceptions of themselves and others (Bruckmüller et

al., 2012; Gaucher et al., 2011; Henley et al., 1988). Scholars have argued that referring to

women as *girls* rather than *women* signifies that women are immature and incapable of taking on

responsibility or making important decisions (Lakoff, 1973; Richardson, 1993), and initial

research supports this conclusion. For example, participants rated an applicant for a high-level

executive position as less mature, tough, dignified, and brilliant, and less deserving of a high

salary when labeled *girl* rather than *woman* (Brannon, 1978). Some evidence also suggests that

women are aware of the stigma associated with being called *girl*. In one study (MacArthur,

2015), undergraduate women received bogus leadership feedback that referred to them as either

girl or *woman*. Compared to participants who received the *woman* feedback, those who received

the *girl* feedback reported that the feedback made them feel less mature, adult, and professional,

and also felt less confident that evaluators who read the feedback would see leadership qualities

in them.

      Although there is reason to believe that the label *girl* is used commonly to describe

women, and that its use could potentially have negative consequences for women, no research to

our knowledge has reliably estimated its prevalence in natural daily conversations. Therefore, we analyzed samples of undergraduate students' language use across daily life contexts to examine instances of this infantilizing label. Importantly, we focused our investigation on undergraduate students because we assumed that women are most likely to be called girls when they are transitioning into adulthood (late teens to late twenties), and that being referred to as *girl* may be more impactful and potentially damaging to women at a time of important identity and career development (Arnett, 2007).

Psychology research has been criticized for focusing on decontextualized laboratory studies at the expense of naturalistic studies. To address this criticism, we used a method for recording naturally occurring conversations, described in more detail in the Methods section. Additionally, to better understand the relative experience of being called *girl*, we compared how often women are called *girl* to how often women are called other labels, and how often men are called *boys*. Based on the research we reviewed here, we predict that women will be called *girl* more often than any other label (Hypothesis 1a) and more often than men are called the infantilizing label *boy* (Hypothesis 1b).

Because it could be argued that labels for men other than *boy* are also infantilizing, in particular the common label *guy*, we conducted a pilot study in which we asked participants ($n =$ 77; 61 women, 16 men) to report in an open-ended manner the label they would be most likely to use to refer to males and females in three age categories: 12 and under, 13–17, and 19–25. Results indicated that for children age 12 and under, participants were most likely to use the label *girl* (71%) to refer to females, and the labels *boy* (50%) and *kid* (39%) to refer to males. For children aged 13–17, *girl* (79%) was again most commonly reported for females, whereas the labels *kid* (37%), *boy* (26%), and *guy* (24%) were most commonly reported for males. For those

aged 18–25, participants reported being most likely to use the label *girl* (69%) for females, and *guy* (84%) for males. Thus, our data show that whereas *girl* is unambiguously the most common label used to refer to a female in both early and late childhood, and *boy* is the most common label for males in early childhood, *guy* is a label that is applied increasingly to males as they enter adolescence and adulthood. Consequently, the label *guy* cannot be seen as infantilizing to the same degree as *girl*, a point that is supported by the dictionary definition of *guy* as a "man or fellow" (Merriam-Webster, 2018).

The pilot data we described here also provide justification for our focus on the label *girl* rather than other potentially infantilizing labels for women: *girl* was the only label in our pilot data that participants applied to both young female children and adult women. We reasoned, therefore, that *girl* is the most pervasive (and thus potentially harmful) infantilizing label used to describe women. Other labels that could be considered infantilizing, such as *gal* and *chick,* were not observed in our data, and for this reason, they did not form part of our primary hypotheses about infantilizing labels. We did, however, include these labels among those to be searched in our conversational data, allowing us to gather baseline information about their use in everyday speech.

Androcentrism and Masculine-Label Bias

Androcentrism, or people's tendency to define the standard person as male and to treat men and masculinity as the norm, has been a primary topic of feminist study for decades (Bem, 1993; Eagly & Kite, 1987). Early findings, for example, showed that when thinking of the category *human*, people tend to bring to mind male rather than female exemplars (Bailey & LaFrance, 2017; Hamilton, 1991) and that stereotypes of nationalities tend to resemble stereotypes of men from those nations more than women (Eagly & Kite, 1987). Studies of gender

in the popular media have further shown that androcentrism tends to be the norm in North

American society. For instance, men are featured more often than women in news stories and

photos (Len-Rios, Rodgers, Thorson, & Yoon, 2005), prime-time television shows (Lauzen,

Dozier, & Horan, 2008), and prime-time advertisements (Ganahl, Prinsen, & Netzley, 2003).

Male characters are also represented more often than female characters in video games (Burgess,

Stermer, & Burgess, 2007), children's picture books (Gooden & Gooden, 2001; Hamilton,

Anderson, Broaddus, & Young, 2006), and computer clip art (Milburn, Carney, & Ramirez,

2001).

Given the above-mentioned research on androcentrism in people's thinking and in the

media, it would not be surprising to find such biases also embedded in daily language use, and

research on gender bias in language confirms this suspicion. In particular, many studies have

focused on a type of gender-biased language called *masculine generics*, a type of speech in

which people use masculine words (e.g., *he, his, man*) generically to refer to both women and

men. Studies on masculine generics have shown that they are not truly generic, but instead lead

people to think of men more than gender-neutral words (e.g., *they*, *their*, *person*) do (Bailey &

LaFrance, 2017; Crawford, 2001; Gastil, 1990; Sczesny, Formanowicz, & Moser, 2016;

Weatherall, 2002), with unintended consequences. For instance, compared to gender-neutral

pronouns, using a generic *he* in job descriptions leads women to avoid applying for the job (Bem

& Bem, 1973) and to express less interest in it (Briere & Lanktree, 1983; Stout & Dasgupta,

2011), and leads others to rate women as less competent for the job (Hyde, 1984). Masculine

generics can also influence self-perceptions. Henley and colleagues (1988) found that boys' self-

esteem increased more after hearing stories with masculine generic language rather than gender-

neutral language or feminine generics, whereas girls' self-esteem increased more after hearing stories with gender-neutral language.

Although the use of masculine generics has increasingly fallen out of favor (Parks & Robertson, 1998a, 1998b, 2000), with organizations like the American Psychological Association (APA, 1994) cautioning against their use, some forms seem to persist in English spoken language—for example, using the label *guys* to refer to a group of any gender composition. Research has not yet examined whether these more modern phrases (e.g., "you guys") bring to mind images of men more readily than images of women, as other forms of masculine generic language do. However, given that *guy* has historically referenced male targets (Bebout, 1984), and according to our pilot data, continues to be used to reference males more often than females, it is probable that such phrases represent modern forms of masculine generic language that persist in everyday conversation.

Considering the tendency of both language and society in general to represent men and boys more frequently than women and girls, and to represent maleness as the default, we predict that people will use more masculine labels (e.g., *man, boy, guy*) than feminine labels (e.g., *woman, lady, girl*) in their everyday speech (Hypothesis 2). We also explore whether the predicted masculine-label bias will appear consistently across a variety of target types. For instance, targets of gendered labels can be: mixed-gender groups (calling a group of women and men *guys*), gender-consistent with the label (calling a man *man*), gender-inconsistent with the label (calling a man *girl*), non-person objects (*men's restroom*), or an expression with no specific target (*oh man*). Given the prevalence of androcentric bias (Bem, 1993; Eagly & Kite, 1987) and masculine generics (Weatherall, 2002), we suspect that a masculine-label bias will appear for each target type (Hypothesis 3).

Gender Differences

As an exploratory analysis, we examined whether the predicted pattern of infantilizing labels and masculine-label bias varied across speaker gender. On the one hand, it is reasonable to suspect that both women and men use gender-biased language to the same extent, given that people of all genders are susceptible to unconscious gender bias (Moss-Racusin et al., 2012). On the other hand, because it has been found that college students tend to socialize more with same-gender peers than with other-gender peers (Mehta & Strough, 2010), it is also plausible that men use masculine labels more often than feminine labels, whereas women use feminine labels more often than masculine labels. We therefore tested whether the predicted patterns of infantilizing labels and masculine-label bias were present for both female and male speakers.

Method

To address our research questions about the prevalence of women being called *girls*, and about an overarching androcentric bias in the labels used to describe women and men, we analyzed sound recordings sampled from undergraduate students' daily interactions.

Participants

We derived our sample from two previously published samples of 96 participants (Mehl, Gosling, & Pennebaker, 2006) and 79 participants (Vazire & Mehl, , 2008), for a total of 175 participants. Eighty-nine of these participants (51%) identified as female; 86 (49%) identified as male. Sample size was determined based on the recorded conversational data available when the research began, and as described in more detail in the Results section, data from these participants led to the inclusion of 812 instances of gendered labels for our first set of analyses (infantilization) and 1,114 instances of gendered labels for our second set of analyses (androcentrism). A sensitivity analysis using G*Power 3 (Fraul, Erdfelder, Lang, & Buchner,

2007), with power set at .80 and alpha set at .05, indicated that a sample of 812 labels was large enough to detect effects in which an observed frequency (in this case, instances of gendered labels) and its expected frequency (the number of labels expected by chance) differed by 15% or greater. For example, if women were called *girl* twice as often (e.g., $n = 200$) as they were called all other feminine labels (*woman, lady, and chick*) combined (e.g., $n = 100$), this would represent a 166.66% difference from what would be expected by chance, wherein chance is defined as the number of instances expected if all labels were to be used equally. That is, statistically, we would expect all four labels for women examined in this study (*girl, woman, lady*, and *chick*) to be equally frequent (via the null hypothesis), meaning that out of a total of 300 instances, each label would be expected to occur 75 times (300/4 = 75). Consequently, if we observed 200 instances of the label *girl*, this would mean *girl* occurred 166.66% more than statistically expected (200/75 = 2.66). Our previously discussed pilot data showed that in the 18–25 age range, participants applied the label *girl* to women more than twice as often as they applied other labels (20 instances of *girl* vs. 9 instances of other labels), and that men were never referred to as *boys* in that age range.; Therefore, our sample of 812 was more than adequate to detect the effects expected for our primary research questions around infantilizing labels. Similarly, a sensitivity analysis indicated that a sample of 1,114 labels (androcentrism analyses) was large enough to detect effects in which observed and expected frequencies differed by 12.5% or greater. Although we are not aware of any previous literature that compares the frequency of masculine to feminine language use, research suggests that the representation of male-to-female speaking characters in media tends to hover around 2:1 (Smith, Choueiti, & Pieper, 2017), a 33.33% difference from what would be expected by chance. Using this ratio as a baseline estimate of

androcentric thinking more broadly, our sample was more than adequate to detect effects for

analyses of androcentric bias in language.

Participants in the study were recruited from introductory psychology classes and fliers

posted in the psychology department at a large university in the southwestern United States., and

they ranged in age from 17 to 23 ($M = 18.70$, $SD = 1.13$). Most identified as White ($n = 111$,

63%), followed by Asian ($n = 32$, 19%) and Latino/a ($n = 24$, 14%).

Procedure

Participants wore the Electronically Activated Recording (EAR) device (Mehl, 2017;

Mehl, Pennebaker, Crow, Dabbs, & Price, 2001) for either two (47 women, 49 men) or four (42

women, 37 men) consecutive days during waking hours. The EAR consisted of a digital audio

recorder, external microphone, and controller microchip carried in a shock-protective case,

which participants were given upon reporting for the study and after having been thoroughly

briefed about the EAR procedure. The EAR was programmed to record ambient sounds for 30

seconds every 12.5 minutes, producing 4.8 sound recordings per hour, and participants were

unaware of when the device was recording. Participants were encouraged to wear the device as

much as possible and to only take it off at night, when they did not want to be recorded, and

when the proper functioning of the device was at risk (e.g., during exercise). They were told that

the device operated automatically, and that to ensure its proper functioning, all they needed to do

was ensure that the microphone was accessible. Finally, participants were provided with a phone

number they could call if they had questions (no participant called).

These procedures yielded a total of 36,180 valid waking recordings across all

participants, an average of 207 recordings per participant ($SD = 65.6$). The number of valid

recordings differed based on how long participants wore the EAR, translating to a mean of 130

waking recordings (*SD* = 34) for those who wore the EAR for 2 days and 300 (*SD* = 104) for those who wore it for 4 days. The average number of valid waking recordings reflects a high compliance rate, close to what would be expected if participants wore the EAR at all times during waking hours. That is, if participants wore the EAR for 16 hours each day while awake, we would expect a total of 153 recordings for participants who wore the EAR for 2 days and 307 recordings for those who wore it for 4 days, figures that are close to the averages of 130 and 300, respectively, that we observed in our samples.

Self-report surveys administered following the EAR sessions were consistent with this high compliance rate and indicated limited intrusion of the EAR in participants' everyday lives. Specifically, participants on average indicated that they wore the EAR 75% of their waking time (*SD* = 16%). They also reported on 5-point scales from 1 (*not at all*) to 5 (*a great deal*) that the EAR was only moderately intrusive for themselves (*M* = 2.91, *SD* = .96) and the people around them (*M* = 3.22, *SD* = .98); that they felt only slightly constrained by the device (*M* = 1.88, *SD* = 0.91) and uncomfortable wearing it (*M* = 2.04, *SD* = 1.02); and that the EAR caused only minor changes in their speech (*M* = 1.43, *SD* = 0.75), behavior (*M* = 1.57, *SD* = 0.77), and the behavior of people around them (*M* = 1.98, *SD* = 1.03). Furthermore, participants rarely mentioned the EAR in their daily conversations (*M* = 3.2%, *SD* = 4.4%).

Importantly, although some individuals had higher compliance wearing the EAR than others, and therefore had a greater number of sound files included in the analyses, there is no a priori reason to believe that those who demonstrated less compliance with the EAR would use the labels of interest at differing rates from participants with high compliance (the studies that resulted in the two samples used here were originally introduced to participants as being about personality; Mehl et al., 2006, 2006; Vazire & Mehl, 2008). Therefore, we do not expect that

differences in compliance had a significant bearing on our results. As we discuss in greater detail

in the Results section, analyses of the speech patterns of individual participants in our study

revealed that a similar pattern of infantilizing and androcentric speech was observed in the

majority of participants rather than simply a handful who used certain labels more frequently

than others. Therefore, we believe that these patterns of gender-biased language are pervasive in

everyday speech rather than being driven by any artifact of our data collection.

The unobtrusive recording of conversations raises important ethical and legal questions.

These are discussed in detail in Mehl and Holleran (2007), Manson and Robbins (2017), and in

Robbins (2017). The two EAR studies reported here implement several safeguards to protect the

privacy of participants and their conversation partners. First, the net audio recording was limited

to a small fraction of the day (approximately 5% of the time), effectively leaving 95% of

participants' conversations private. Second, the short 30-sec recordings ensured that minimal

personal information was captured beyond what was needed for the coding. Third, participants

had the opportunity to review their recordings and delete any they wanted to remain private; very

few were deleted (approximately 0.01%; Mehl et al., 2006). Fourth, participants were instructed

to inform their conversation partners about the possibility of being recorded, thereby minimizing

their expectation of privacy. Finally, from a legal perspective, both studies were conducted in a

state where the recording of private conversations required the consent of only one party of the

conversation and both studies were approved by an Institutional Review Board.

Following data collection, all of participants' utterances recorded by the EAR were

transcribed, and the transcriptions were searched for instances of commonly used labels for

women (*girl/s*, *woman/women*, *lady/ladies*, *gal/s*, *chick/s*, *sis*) and men (*boy/s*, *man/men*,

gentleman/men, *guy/s*, *dude/s*, *bro/s*), resulting in 1,206 instances. Gendered labels were

determined through the pilot study we described previously ($n = 77$; 61 women, 16 men), in which undergraduates indicated the gendered labels they use to describe females and males in different age categories, as well as another portion of the same pilot in which they indicated the labels that others use to describe them. We included all relevant labels mentioned by participants, as well as the other-gender equivalents of each label so that the number of feminine and masculine labels were equivalent.

For coding purposes, research assistants listened to the original audio clips that contained gendered labels. Some audio clips did not work, did not match the transcribed text (and thus did not contain the gendered label), or could not be found ($n = 92$). The remaining instances of gendered labels ($n = 1,114$) came from 926 different audio clips and from 157 participants (77 women, 80 men; 86 of whom wore the EAR for 2 days, 71 of whom wore the EAR for 4 days), with an average of 7.10 ($SD = 6.04$) gendered label utterances per participant. The average number of gendered label utterances differed based on the sample: 6.97 ($SD = 6.77$) for participants who wore the EAR for 2 days and 7.25 ($SD = 5.06$) for those who wore it for 4 days. These 157 participants ranged in age from 17 to 23 ($M = 18.69$, $SD = 1.14$) and identified primarily as White ($n = 100$, 63.7%), Asian ($n = 27$, 17.2%), and Latina/o ($n = 22$, 14.0%).

Coding Scheme

Two research assistants who were unaware of the hypotheses of the study initially coded the audio recordings for whether each gendered label was a masculine or feminine word. Next, they coded each label according to target type (Kappa $= 0.60$; agreement $= 89.68\%$), that is, whether the label was targeted toward a person/s ($n = 964$) or non-person/s ($n = 150$). Most disagreements for target type involved confusion over whether the label targeted a person/s or was an expression. Several English expressions use masculine words that could also refer to a

person. For example, *man* is sometimes used as a synonym for *geez* or *my goodness*. Take, for instance, this phrase uttered by one participant: "Man I haven't done lunges in how long and he makes us do 120 of them?" It is unclear in this case whether *man* is an expression or whether the participant is addressing a male conversation partner.

Labels coded as targeting non-person/s were further coded either as an expression ($n = 110$) or as something else (e.g., a movie/song/book title, animal, or object; $n = 40$; Kappa $= 0.78$; agreement $= 90.83\%$). Labels coded as targeting person/s were further coded according to target age (Kappa $= 0.37$; agreement $= 98.20\%$): adult (defined as 18 or older; $n = 947$), child ($n = 13$), unknown ($n = 4$), or mixed ages ($n = 0$). Given that most undergraduates are age 18 or over, coders were instructed to categorize any undergraduate target as an adult. Interrater reliability was low for target age because both coders overwhelmingly rated the labels as targeted toward adults, and Kappa heavily weights disagreements in such cases. That is, categories that have unequal probabilities, as is the case here, tend to produce low Kappa values (Bakeman, McArthur, Quera, & Robinson, 1997). Finally, labels coded as targeting person/s were further coded according to target gender (Kappa $= 0.82$; agreement $= 93.82\%$): female ($n = 302$), male ($n = 585$), unknown ($n = 61$), or mixed gender group ($n = 16$). All discrepancies (target type: $n = 115$; target age: $n = 16$; target gender: $n = 87$) were resolved by the first author; excluding these instances does not change the significance or direction of the results.

Given that coders made determinations about target age and gender based on somewhat limited information (30-second clips), it is possible that some instances could have been mis-categorized, particularly when coders assessed target gender. For example, if speakers used a label (e.g., *girl*) to refer to someone not present in the conversation, the information given about that individual was sometimes sparse; therefore, in ambiguous cases, coders were instructed to

classify the instance as "unknown." It is important to note that interrater reliability for gender

was high (Kappa = .82); however, even if some instances were mis-categorized, we would not

expect this to influence the validity of our results. That is, there is no reason to believe that

coders would be more likely to improperly code labels referring to female speakers than labels

referring to male speakers.

Similarly, upon coding instances of labels used to refer to people present in the

conversation, when coders heard both male and female voices responding to a particular label

(e.g., "guys"), the instance was coded as "mixed gender," whereas the instance was coded as

"male" or "female" if the voices of only males or females were heard. As mentioned previously,

if there was any ambiguity, coders would classify the instance as "unknown." Importantly,

coders would not have been able to ascertain whether there were individuals present (and who

may have been included in the reference group when a term such as *guys* was used) but not

speaking in the clip. As we discussed previously, we do not have any reason to expect that one

gender or another would have been more likely to be inaudible in the clips, therefore, we do not

believe that possible mis-categorizations influenced the pattern of results we obtained.

Results

Preliminary analysis. Given that we derived our data from two separate samples

collected at different times, we tested whether these samples differed in the number of labels

available for analysis. Chi-square analyses revealed that, after adjusting for the number of

participants in each sample (86 in one; 71 in the other), the two samples did not differ in rates of

gendered labels, $\chi^2(1, n = 1114) = .46$, $p = .500$. Therefore, in the analyses that follow, we

collapsed across the two samples; analyzing the samples separately does not change the direction

or significance of the results.

Infantilizing labels. We predicted that women would be called *girl* more often than any other label (Hypothesis 1a) and more often than men are called *boy* (Hypothesis 1b). To accurately test these hypotheses, we analyzed only those instances of gendered labels that were targeted toward adult person/s and were gender-consistent (i.e., the person being called *woman* was determined by coders to be an adult woman). That is, we excluded instances of labels that referred to non-persons ($n = 150$), non-adults ($n = 13$), targets of unknown age ($n = 4$) or gender ($n = 61$), mixed gender groups ($n = 16$), or that were gender-inconsistent (e.g., a man being called *girl*, $n = 58$). The final sample for Hypothesis 1 included 812 instances of the labels, which came from 150 participants (72 women, 78 men).

An analysis of the frequencies of female and male labels used by participants in the sample revealed that out of a total of 240 instances of labels for women, *girl* was the most common ($n = 167$; 69.6%), followed by *woman* ($n = 31$; 12.9%), *lady* ($n = 30$; 12.5%), and *chick* ($n = 12$; 5%). For the data referring to men, 572 instances of the labels of interest were found. Of these, *guy* was the most frequently used ($n = 272$; 47.6%), followed by *dude* ($n = 164$; 28.7%), *man* ($n = 123$; 21.7%), and *boy* ($n = 13$; 2.3%). There were no instances of *gal*, *sis*, or *gentleman* observed in the data. Table 1 (top rows) shows the Chi-square analyses for Hypothesis 1a and Hypothesis 1b. Consistent with both predictions, women were called *girl* significantly more often than all other labels, as well as significantly more often than men were called *boy*.

To ensure that results were not driven by a small number of participants who were simply using the label *girl* much more than other participants, we also analyzed how often the hypothesized patterns of label use emerged within the speech of individual participants. Results indicated that 79 participants (50.32%) used *girl* to describe women more often than any other label, whereas only 23 (14.65%) used other labels for women more often than they used *girl*.

Similarly, 91 participants (57.96%) used *girl* to describe women more often than they used *boy*

to describe men, whereas the reverse pattern was observed in only 5 participants (3.18%).

To test whether the use of infantilizing labels was present for both female and male

speakers, we reran all analyses separately for women and men. Results revealed that both female

and male speakers exhibited gender-biased language overall. Specifically, both female and male

participants called women *girl* more often than any other label and more often than they called

men *boy* (see middle and bottom rows of Table 1).

Masculine-label bias. Table 2 displays the frequencies of masculine and feminine labels.

We predicted that masculine labels would be used more frequently than feminine labels overall

(Hypothesis 2). To test this prediction, we used the full dataset including instances of the labels

that referred to a non-person, child, mixed or unknown gender or age group, and those coded as

gender-inconsistent. The final sample for Hypothesis 2 included 1,114 instances of gendered

labels, coming from 157 participants (80 male, 77 female). A Chi-square analysis revealed that,

consistent with Hypothesis 2, masculine labels were used significantly more frequently than

feminine labels (see top row of Table 2). To ensure, once again, that a small number of

participants were not responsible for these results, we assessed how frequently androcentrism

was observed in the speech of individual participants. We found that 116 participants (73.89%)

used masculine labels more often than feminine labels, whereas only 23 participants (14.65%)

used feminine labels more than masculine labels.

As we discussed previously, the greater number of masculine compared to feminine

labels may vary across target type. To test this possibility, we divided the dataset into specific

target types: mixed gender groups, expressions, non-persons, gender-inconsistent, gender-

consistent, and unknown gender. We performed a Chi-square analysis on each target type to test

whether more masculine than feminine labels were used in each. We predicted that masculine labels would be used more often than feminine labels in each target category (Hypothesis 3).

Table 2 displays the results. Mostly consistent with Hypothesis 3, masculine labels were used more often than feminine labels in all but one target type, including gender-inconsistent labels, gender-consistent labels, expressions, targets of unknown gender, and mixed gender groups (all 16 instances involved masculine labels, so no Chi-square statistics are reported for this target type). The only target type for which there was no significant difference in the frequency of masculine and feminine labels was non-person objects.

To test whether both female and male speakers exhibited masculine-label bias, we conducted separate Chi-square analyses for both groups of speakers. As shown in the middle and bottom rows of Table 2, both female and male speakers used masculine labels more often than feminine labels overall, as well as in each of the target categories, except for two notable exceptions. First, in the gender-consistent category, men used masculine labels significantly more than feminine labels, whereas women used masculine and feminine labels equally. Second, in the non-person category, neither female nor male speakers used masculine labels more than feminine labels.

Discussion

In the present study, we examined audio-recordings sampled from undergraduates' everyday speech to investigate two common forms of subtle gender bias that have been identified in previous research: paternalism and androcentrism (Bem, 1993; Eagly & Kite, 1987; Glick & Fiske, 1996). Specifically, we assessed paternalism in language by examining use of the infantilizing label *girl* to refer to adult women, and androcentrism by investigating whether or not there is an overarching masculine bias in the labels used to describe women and men. Results

showed support for the idea that young women are frequently referred to as *girls* in everyday

social interactions: *girl* was the most commonly used label for women and was used to describe

them 70% of the time. Indeed, results revealed that *girl* was used to describe women

significantly more often than any other label (including the age-appropriate label *woman*), as

well as more often than *boy* was used to describe men.

The fact that *boy* was rarely used to describe men (occurring in only 2% of label uses)

reveals an asymmetry in the use of masculine and feminine infantilizing labels, and suggests the

presence of gender bias in everyday conversation. Interestingly, the labels *chick* and *gal* were

rarely used in our sample, indicating that these terms may have fallen out of favor in

undergraduates' everyday speech. Given that the most common masculine label in our sample

was *guy*, it seems that there has been a shift over time in the equivalence of particular terms, with

girl, rather than *gal*, now representing the equivalent of *guy*. As revealed in our pilot data,

however, *girl* has infantilizing connotations, whereas *guy* is used primarily to refer to male

adolescents and adults.

Results of the study also revealed an overarching androcentric bias, in that both female

and male speakers used masculine labels significantly more often than feminine labels overall.

Importantly, androcentric bias revealed itself in almost all of the target categories we analyzed.

First, masculine labels (e.g., *guys, dude, man*) were used more often to refer to a mixed gender

group than were feminine labels, which were never used. In particular, the label *guys* was used

frequently to refer to groups of men and women, thus representing a modern-day masculine

generic: its original male target has shifted to include both males and females. Previous research

on masculine generics focused mainly on words like *he* and *his*, showing that when used

generically, these terms brought to mind men more than gender-neutral language did (and thus

were not truly generic; Weatherall, 2002). It is unknown, however, whether the term *guys* also

brings men to mind more than women or mixed-gender groups. Therefore, it is important that

future research test the possibility that *guys* brings to mind groups of men more frequently than

gender-neutral terms do. Importantly, however, even if future research reveals that phrases like

"you guys" do not result in a male construal (i.e., that they bring to mind both men and women),

our research nevertheless reveals a gendered asymmetry in label usages: multiple labels

historically associated with men (*man, dude, guy*) are applied generically, whereas very few

labels historically associated with women are applied in the same way.

Second, masculine labels were applied to women more than feminine labels were applied

to men: *guy*, *dude*, and *man* were frequently used to refer to women and girls, but feminine labels

were rarely applied to men and boys. When feminine labels were used to refer to men or boys,

they often seemed to be used as an insult. For instance, one participant remarked: "I'm going to

have to come out and be like, Andre [pseudonym], don't be a girl. One time I told him, 'you're

being a girl.' I'm like, 'why are you like a girl, I am dating a guy, please." This is consistent with

other research on sexism (Alksnis, Desmarais, & Curtis, 2008; Crawford, 2012; Levanon,

England, & Allison, 2009) suggesting that whatever is associated with women tends to be

devalued relative to whatever is associated with men. In fact, a key component of masculinity is

to avoid anything feminine (Brannon, 1976; Kimmel, 1996), and men risk losing access to the

resources and privilege associated with manhood status when they act in a feminine manner (for

review, see Bosson, Vandello, & Caswell, 2014).

Third, expressions using masculine labels were much more common than expressions

using feminine labels. Indeed, expressions such as "oh man," and "boy, was that good" were

frequent, whereas there were no similar expressions using feminine labels. Although our data do

not allow us to draw conclusions about how such expressions came into existence, examining

this result along with our other findings indicates a pattern consistent with larger androcentric

biases in which men are considered the default human (Eagly & Kite, 1987).

Fourth, we found that for gender-consistent labels, masculine labels were used more

frequently than feminine labels, but only among male speakers. This result suggests that men

may talk about other men more often than they talk about women, whereas women talk about

women and men equally. Alternatively, men may be more likely than women to use gendered

labels to refer to each other (e.g., "hey *bro*"), or men in our study may have simply interacted

more with men and boys than with women and girls, a possibility that we could not assess with

our data.

Fifth, results showed that there were significantly more masculine than feminine labels

applied to targets whose gender could not be determined by coders. Given that *guy* was the most

frequent label in this category (77%), this finding likely reflects the androcentric bias seen in

previously discussed categories, such as using *guy* to refer to a woman (gender-inconsistent) or a

mixed gender group. It is also possible that men were simply being discussed more in cases

where a target's gender was unknown, reflecting the masculine bias previously discussed for

gender-consistent targets. In either case, our findings seem to suggest an androcentric bias in

gendered label use.

Finally, there was one category in which feminine and masculine labels were used

equally: non-person objects. Labels in this category were mainly used to describe the titles of

movies (e.g., *Bad Boys*), television shows (e.g., *Family Guy*), and songs (e.g., *My Girl*), as well

as non-human objects (e.g., *women's restroom*) and animals (e.g., saying "good boy" to a dog). It

is unclear why no difference between masculine and feminine labels was found here, particularly given that male characters are over-represented in media, as we discussed previously.

Limitations and Future Directions

The present research provides evidence that both paternalistic and androcentric language are prevalent in U.S. undergraduates' everyday speech; however, limitations to our research should be considered. First, given that our sample was restricted to undergraduate participants, results may not generalize to the broader population. It is possible, for example, that we would find less infantilizing label usage among older individuals. As stated previously, we used an undergraduate sample purposefully because we speculated that infantilization may be most common and damaging for college-aged women. We expect, however, that the overarching masculine bias we observed in our data would be found among participants of all ages, given findings on the pervasiveness of androcentrism (Bailey & LaFrance, 2017; Ganahl et al., 2003; Hamilton, 1991; Lauzen et al., 2008; Len-Rios et al., 2005).

In addition to age, our sample was limited to predominately White participants who likely shared similar educational, socioeconomic, and geographic backgrounds. More broadly, data were collected in an English-speaking, southwestern U.S. cultural context, and therefore, may not generalize to other cultures or languages, or to individuals of other racial-ethnic groups even within southwestern American culture. Indeed, it is likely that the prevalence of gender-biased language differs across racial-ethnic groups, across regions within the United States, and across English-speaking countries and that it may manifest differently (or not at all) in other languages. For these reasons, future research should examine the extent to which these biases occur in other cultural and linguistic contexts.

A further limitation of our data is that it consisted of only short (30-second) sound clips, which made it difficult to assess details beyond the gender and age of targets in each clip. Precautions were taken to minimize this limitation, including having two coders rate each audio clip and having the first author resolve discrepancies. As we noted previously, even when we excluded instances where there were disagreements between coders, analyses revealed the same pattern of results. The nature of our data, however, made it difficult to conduct more fine-grained analyses of label usages. We were not able to ascertain, for example, whether particular labels were more often used to describe someone who was present versus not present, or who was of lower, equal, or higher status than participants. In the future, we hope to replicate our findings using longer segments of conversational data and/or larger datasets sampled from social media contexts, both of which would enable more fine-grained analyses.

Given the nature of our data, many questions remain to be investigated by future research. For example, examining conversational data from workplace contexts would help to uncover how often women are infantilized in professional settings, and data with participants of more varied ages would help to ascertain at what age, if any, women are no longer referred to as *girls*. Recent research has suggested that millennials have experienced a delay in the cultural onset of adulthood (Furstenberg, 2010), which could influence use of infantilizing labels to describe young adults. This recent trend would not account for the findings of the present study, however, because we would not expect usage of infantilizing labels to increase for women and not men with the delayed onset of adulthood. Continued use of *girl* into adulthood may be connected, though, to society's greater emphasis on maintaining a youthful appearance in women (Henss, 1991), and this would be an interesting direction for future research.

Finally, it is also unclear why people use more masculine than feminine labels and whether there are certain contexts in which people might use one more than the other. If masculine-label bias is due in part to the overrepresentation of men in media, for example, we might expect media consumption to be correlated with masculine-label bias and that exposure to women in media might reduce the masculine-label bias. More broadly, future research could examine the individual and situational factors that predict the use of gender-biased language.

Practice Implications

Our data suggest that the linguistic infantilization of women is commonplace in everyday conversation. However, the routine nature of this infantilization should not be taken as a sign of innocuousness: research shows that subtle forms of sexism can have a cumulative effect over time (Berg, 2006; Swim, Hayes, Cohen, & Ferguson, 2001) and, as we discussed previously, can have negative consequences for women (Barreto et al., 2010; Becker & Wright, 2011; Dardenne et al., 2007; Jost & Kay, 2005; Shepherd et al., 2011). Importantly, benevolent sexism/protective paternalism and other subtle forms of bias can be even more damaging than overtly sexist behavior precisely *because* they are not recognized as sexist and therefore go unchallenged (Barreto & Ellemers, 2005; Dardenne et al., 2007; Jones, Peddie, Gilrane, King, & Gray, 2016; Nguyen & Ryan, 2008; Swim, Mallett, & Strangor, 2004). Initial research suggests that infantilizing language may harm women in certain contexts, such as in leadership and workplace settings (Brannon, 1978; MacArthur, 2015). It is unknown, however, how repeated instances of infantilization in everyday conversation affect women. Given the possibility that this type of language is harmful to women, it is important that teachers, professors, clinicians, and others in positions of power be mindful of their own language and how it may include, exclude, or infantilize women.

Similarly, it is important that practitioners of all kinds pay careful attention to their use of androcentric language. We know, for instance, that certain forms of masculine generic language lead people to think of male rather than female exemplars (Bailey & LaFrance, 2017; Crawford, 2001; Gastil, 1990; Sczesny, Formanowicz, & Moser, 2016; Weatherall, 2002) and that newer forms of gender-neutral language may be particularly effective at correcting this bias (Lindqvist, Renström, & Sendén, 2018). Therefore, it is crucial that future research examine whether the modern forms of masculine generic language shown to be pervasive in the present research (e.g., "you guys") lead to similar biases in thinking as what has been found in past research and whether gender-neutral phrases (e.g., "you all;" "y'all") may correct for any bias that may be produced.

Conclusion

Whereas past research has examined the consequences of gender-biased language on thought and behavior, we extend prior work by examining the prevalence of paternalistic and androcentric language in daily life. Our results illuminate the gendered nature of the language landscape that young women and men navigate every day—a landscape in which gender-biased language is used frequently by both female and male speakers. The prevalence with which women are infantilized in language, as well as the masculine-label bias found in the present study, indicate that linguistic sexism is an important topic for further study. In particular, additional research is needed to understand both the causes and cumulative effects of gender bias in everyday language.

References

Alksnis, C., Desmarais, S., & Curtis, J. (2008). Workforce segregation and the gender wage gap: Is "women's" work valued as highly as "men's"? *Journal of Applied Social Psychology*, *38*, 1416-1441. https://doi.org/10.1111/j.1559-1816.2008.00354.x

American Psychological Association. (1994). Guidelines to reduce bias in language. *Publication Manual of the American Psychological Association*, 46-60.

Arnett, J. J. (2007). Emerging adulthood: What is it, and what is it good for? *Child Development Perspectives*, *1*, 68-73. https://doi.org/10.1111/j.1750-8606.2007.00016.x

Bailey, A. H., & LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. *Sex Roles, 76*, 682-693. https://doi.org/10.1007/s11199-016-0648-4

Bakeman, R., McArthur, D., Quera, V., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods, 2*, 357-370. https://doi.org/10.1037/1082-989X.2.4.357

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, *35*, 633-642. https://doi.org/10.1002/ejsp.270

Barreto, M., Ellemers, N., Piebinga, L., & Moya, M. (2010). How nice of us and how dumb of me: The effect of exposure to benevolent sexism on women's task and relational self-descriptions. *Sex Roles*, *62*, 532-544. https://doi.org/10.1007/s11199-009-9699-0

Bebout, L. (1984). Asymmetries in male-female word pairs. *American Speech, 59,* 13-30. https://doi.org/10.2307/454991

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism

undermines and hostile sexism motivates collective action for social change. *Journal of*

Personality and Social Psychology, *101*, 62-77. https://doi.org/10.1037/a0022615

Bem, S. L. (1993). *The lenses of gender*. New Haven, CT: Yale University Press.

Bem, S. L., & Bem, D. J. (1973). Does sex-biased job advertising "aid and abet" sex

discrimination? *Journal of Applied Social Psychology, 3*, 6-18.

https://doi.org/10.1111/j.1559-1816.1973.tb01290.x

Berg, S. H. (2006). Everyday sexism and posttraumatic stress disorder in women: A

correlational study. *Violence Against Women*, *12*, 970-988.

https://doi.org/10.1177/1077801206293082

Bigler, R. S., & Leaper, C. (2015). Gendered language: Psychological principles, evolving

practices, and inclusive policies. *Policy Insights from the Behavioral and Brain Sciences*,

2, 187-194. doi: 0.1177/2372732215600452

Bodine, A. (1975). Androcentrism in prescriptive grammar: singular 'they', sex-indefinite 'he',

and 'he or she'. *Language in Society*, *4*, 129-146.

https://doi.org/10.1017/S0047404500004607

Bohner, G., Ahlborn, K., & Steiner, R. (2010). How sexy are sexist men? Women's perception

of male response profiles in the Ambivalent Sexism Inventory. *Sex Roles*, *62*, 568-582.

https://doi.org/10.1007/s11199-009-9665-x

Bosson, J. K., Vandello, J. A., & Caswell, T. A. (2014). Precarious manhood. In M. K. Ryan &

N. R. Branscombe (Eds.), *The Sage handbook of gender and psychology* (115-130).

London, UK: Sage.

Brannon, R. (1976). The male sex role: Our culture's blueprint for manhood and what it's done

for us lately. In D. David & R. Brannon (Eds.), *The forty-nine percent majority: The male sex role* (pp. 1–48). Reading, MA: Addison-Wesley.

Brannon*,* R. (1978). *The consequences of sexist language*. Paper presented at the American Psychological Association Meetings, Toronto.

Briere, J., & Lanktree, C.B. (1983). Sex role related effects of sex bias in language. *Sex Roles*, *9*, 625-632. https://doi.org/10.1007/BF00290069

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology*, *42*, 210-218. https://doi.org/10.1002/ejsp.858

Burgess, M. C., Stermer, S. P., & Burgess, S. R. (2007). Sex, lies, and video games: The portrayal of male and female characters on video game covers. *Sex roles*, *57*, 419-433. https://doi.org/10.1007/s11199-007-9250-0

Crawford, M. (2001). Gender and language. In R. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 228-244). New York: Wiley.

Crawford, M. (2012). *Transformations: Women, gender, and psychology* (2nd ed.). New York, NY: McGraw-Hill.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: consequences for women's performance. *Journal of Personality and Social Psychology*, *93*, 764-779. https://doi.org/10.1037/0022-3514.93.5.764

Donnelly, K., Twenge, J. M., Clark, M. A., Shaikh, S. K., Beiler-May, A., & Carter, N. T. (2016). Attitudes toward women's work and family roles in the United States, 1976–2013. *Psychology of Women Quarterly*, *40*, 41-54. https://doi.org/10.1177/0361684315590774

Eagly, A. H., & Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and

    men? *Journal of Personality and Social Psychology*, *53*, 451-462.

    http://dx.doi.org/10.1037/0022-3514.53.3.451

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical

    power analysis program for the social, behavioral, and biomedical sciences. *Behavior*

    Research Methods, *39*, 175-191.

Furstenberg, F. F. (2010). On a new schedule: Transitions to adulthood and family change. *The*

    Future of Children, *20*, 67-87.

Ganahl, D. J., Prinsen, T. J., & Netzley, S. B. (2003). A content analysis of prime time

    commercials: A contextual framework of gender representation. *Sex Roles*, *49*, 545-551.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine

    generics. *Sex Roles*, *23*, 629-643. https://doi.org/10.1007/BF00289252

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job

    advertisements exists and sustains gender inequality. *Journal of Personality and Social*

    Psychology, *101*, 109-128. https://doi.org/10.1037/a0022530

Girl. (n.d.). In *Merriam-Webster's online dictionary* (11th ed.). Retrieved from http://www.m-

    w.com/dictionary/girl

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and

    benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491.

    https://doi.org/10.1037//0022-3514.82.6.878

Gooden, A. M., & Gooden, M. A. (2001). Gender representation in notable children's picture

    books: 1995-1999. *Sex Roles, 45*, 89-101. https://doi.org/10.1023/A:1013064418674

Guy. (n.d.). In *Merriam-Webster's online dictionary* (11th ed.). Retrieved from http://www.m-w.com/dictionary/guy

Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People= male, male= people. *Psychology of Women Quarterly*, *15*, 393-402. https://doi.org/10.1111/j.1471-6402.1991.tb00415.x

Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex Roles, 55*, 757-765. https://doi.org/10.1007/s11199-006-9128-6

Halbert, C., & Latimer, M. (1994). "Battling" gendered language: An analysis of the language used by sports commentators in a televised coed tennis competition. *Sociology of Sport Journal*, *11*, 298-308. https://doi.org/10.1123/ssj.11.3.298

Henley, N. M., Gruber, B., & Lerner, L. (1988, October). *Effects of sex-biased language on attitudes and self-esteem*. Paper presented at a meeting of the Southern California Language and Gender Interest Group. Los Angeles, CA.

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology*, *21*, 933-946. https://doi.org/10.1111/j.1559-1816.1991.tb00451.x

Hyde, J. S. (1984). Children's understanding of sexist language. *Developmental Psychology, 20*, 697-706. http://dx.doi.org/10.1037/0012-1649.20.4.697

Jones, K. P., Peddie, C. I., Gilrane, V. L., King, E. B., & Gray, A. L. (2016). Not so subtle: A meta-analytic investigation of the correlates of subtle and overt discrimination. *Journal of Management*, *42*, 1588-1613. https://doi.org/10.1177/0149206313506466

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender

stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498-509. https://doi.org/10.1037/0022-3514.88.3.498

Kimmel, M. (1996). *Manhood in America: A cultural history*. New York, NY: The Free Press.

Lakoff, R. (1973). Language and woman's place. *Language in Society, 2*, 45-79. https://doi.org/10.1017/S0047404500000051

Lauzen, M. M., Dozier, D. M., & Horan, N. (2008). Constructing gender stereotypes through social roles in prime-time television. *Journal of Broadcasting & Electronic Media*, *52*, 200-214. https://doi.org/10.1080/08838150801991971

Len-Ríos, M. E., Rodgers, S., Thorson, E., & Yoon, D. (2005). Representation of women in news and photos: Comparing content to perceptions. *Journal of Communication*, *55*, 152-168. https://doi.org/10.1111/j.1460-2466.2005.tb02664.x

Lerner, H. E. (1976). Girls, ladies, or women? The unconscious dynamics of language choice. *Comprehensive Psychiatry*, *17*, 295-299. https://doi.org/10.1016/0010-440X(76)90003-1

Levanon, A., England, P., & Allison, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950–2000 US census data. *Social Forces*, *88*, 865-891. https://doi.org/10.1353/sof.0.0264

Lindqvist, A., Renström, E. A., & Sendén, M. G. (2018). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. *Sex Roles*, 1-9. https://doi.org/10.1007/s11199-018-0974-9

MacArthur, H. J. (2015). When women are called "girls": The effect of infantilizing

labels on women's self-perceptions (Unpublished master's thesis). The Pennsylvania

State University, University Park, Pennsylvania.

Manson, J. H., & Robbins, M. L. (2017). New evaluation of the Electronically Activated

Recorder (EAR): Obtrusiveness, compliance, and participant self-selection

effects. *Frontiers in Psychology*, *8*, 658. https://doi.org/10.3389/fpsyg.2017.00658

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A method for the naturalistic

observation of daily social behavior. *Current Directions in Psychological Science*, *26*,

184-190. https://doi.org/10.1177/0963721416680611

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*, 862-877. http://dx.doi.org/10.1037/0022-

3514.90.5.862

Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and

participants' compliance with the Electronically Activated Recorder (EAR). *European

Journal of Psychological Assessment*, *23*(4), 248-257. https://doi.org/10.1027/1015-

5759.23.4.248

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The

Electronically Activated Recorder (EAR): A device for sampling naturalistic daily

activities and conversations. *Behavior Research Methods, Instruments, & Computers*, *33*,

517-523. https://doi.org/10.3758/BF03195410

Mehta, C. M., & Strough, J. (2010). Gender segregation and gender-typing in adolescence. *Sex

Roles*, *63*, 251-263. https://doi.org/10.1007/s11199-010-9780-8

Messner, M. A., Duncan, M. C., & Jensen, K. (1993). Separating the men from the girls: The

gendered language of televised sports. *Gender and Society, 7*, 121-137.

https://doi.org/10.1177/089124393007001007

Milburn, S. S., Carney, D. R., & Ramirez, A. M. (2001). Even in modern media, the picture is

still the same: A content analysis of clipart images. *Sex Roles*, *44*, 277-294.

https://doi.org/10.1023/A:1010977515933

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*, 16474-16479. https://doi.org/10.1073/pnas.1211286109

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of

minorities and women? A meta-analysis of experimental evidence. *Journal of Applied*

Psychology, *93*, 1314-1334. https://doi.org/10.1037/a0012702

Parks, J. B., & Robertson, M. A. (1998a). Contemporary arguments against nonsexist language:

Blaubergs (1980) revisited. *Sex Roles, 39*, 445-461.

https://doi.org/10.1023/A:1018827227128

Parks, J. B., & Robertson, M. A. (1998b). Influence of age, gender, and context on attitudes

toward sexist/nonsexist language: Is sport a special case? *Sex Roles, 38*, 477-494.

https://doi.org/10.1023/A:1018766023667

Parks, J. B., & Robertson, M. A. (2000). Development and validation of an instrument to

measure attitudes toward sexist/nonsexist language. *Sex Roles, 42*, 415-438.

https://doi.org/10.1023/A:1007002422225

Prentice, D. A. (1994). Do language reforms change our way of thinking? *Journal of Language*

and Social Psychology, 13, 3-19. https://doi.org/10.1177/0261927X94131001.

Richardson, L. (1993). Gender stereotyping in the English language. In L. Richardson & V.

Taylor (Eds.), *Feminist frontiers III* (pp. 44-50). New York, NY: McGraw-Hill.

Robbins, M. L. (2017). Practical suggestions for legal and ethical concerns with social

environment sampling methods. *Social Psychological and Personality Science*, *8*, 573-

580. https://doi.org/10.1177/1948550617699253

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender

stereotyping and discrimination?. *Frontiers in Psychology*, *7*, 1-11. doi:

10.3389/fpsyg.2016.00025

Shepherd, M., Erchull, M. J., Rosner, A., Taubenberger, L., Queen, E. F., & McKee, J. (2011).

"I'll get that for you": The relationship between benevolent sexism and body self-

perceptions. *Sex Roles*, *64*, 1-8. https://doi.org/10.1007/s11199-010-9859-2

Smith, S. L., Choueiti, M., Pieper, K. (2017). *Inequality in 800 popular films: Examining*

portrayals of gender, race/ethnicity, LGBT, and disability from 2007-2015. Retrieved

from https://annenberg.usc.edu/sites/default/files/Dr_Stacy_L_Smith-

Inequality_in_900_Popular_Films.pdf

Spence, J. T., & Hahn, E. D. (1997). The attitudes toward women scale and attitude change in

college students. *Psychology of Women Quarterly*, *21*, 17-34.

https://doi.org/10.1111/j.1471-6402.1997.tb00098.x

Stout, J. G., & Dasgupta, N. (2011). When he doesn't mean you: Gender-exclusive language as

ostracism for women. *Personality and Social Psychology Bulletin*, *37*, 757-769.

https://doi.org/10.1177/0146167211406434

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned

and modern prejudices. *Journal of Personality and Social Psychology*, *68*, 199-214.

Swim, J. K., Hayes, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday sexism: Evidence

for its incidence, nature, and psychological impact from three daily diary studies. *Journal*

of Social Issues, 57, 31–53. https://doi.org/10.1111/0022-4537.00200

Swim, J. K., Mallett, R., & Stangor, C. (2004). Understanding subtle sexism: Detection and use

of sexist language. *Sex Roles*, *51*, 117-128.

https://doi.org/10.1023/B:SERS.0000037757.73192.06

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex*

Roles, *36*, 305-325.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Male and female pronoun use in US

books reflects women's status, 1900--2008. *Sex Roles, 67*, 488-493.

https://doi.org/10.1007/s11199-012-0194-7

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique

predictive validity of self-ratings and other-ratings of daily behavior. *Journal of*

Personality and Social Psychology, *95*, 1202-1216. https://doi.org/10.1037/a0013314

Weatherall, A. (2002). *Gender, language and discourse.* London: Routledge.

Table 1

Chi-Square Analyses of "girl" versus Other Labels across Speakers' Gender

	Girl label n	Comparison label n	$\chi(1)^2$
Overall			
Girl vs. other feminine labels	167	73	36.82***
Girl vs. *boy*	167	13	131.76***
Female Participants			
Girl vs. other feminine labels	99	42	23.04***
Girl vs. *boy*	99	9	75.00***
Male Participants			
Girl vs. other feminine labels	68	31	13.83***
Girl vs. *boy*	68	4	56.89***

Note. Analyses included only those instances of gendered labels that were targeted toward adult person/s and were consistent with the gender of the target (i.e., the person being called *girl* is actually female).

***$p < .001$

Table 2

Chi-Square Analyses of Masculine versus Feminine Labels across Target Type and Speakers' Gender

	Masculine label *n*	Feminine label *n*	$\chi(1)^2$	*p*
Overall				
All targets	831	283	269.57	<.001
Gender-inconsistent	52	7	34.32	<.001
Gender-consistent	578	250	129.93	<.001
Mixed gender group	16	0	n/a	n/a
Unknown gender	59	2	53.26	<.001
Expression	106	4	94.58	<.001
Non-person	20	20	1.00	.000
Female Participants				
All targets	252	163	19.09	<.001
Gender-inconsistent	31	3	23.06	<.001
Gender-consistent	172	148	1.80	.180
Mixed gender group	7	0	n/a	n/a
Unknown gender	28	0	n/a	n/a
Expression	31	2	25.49	<.001
Non-person	11	10	.05	.827
Male Participants				
All targets	520	118	253.30	<.001
Gender-inconsistent	21	4	11.56	.001
Gender-consistent	406	102	181.92	<.001
Mixed gender group	9	0	n/a	n/a
Unknown gender	31	2	25.49	<.001
Expression	75	2	69.21	<.001
Non-person	9	10	.05	.819