The Pennsylvania State University

The Graduate School

The College of the Liberal Arts

WHEN WOMEN ARE CALLED "GIRLS": THE EFFECT OF INFANTILIZING

LABELS ON WOMEN'S SELF-PERCEPTIONS

A Thesis in

Psychology

by

Heather J. MacArthur

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2015

The thesis of Heather J. MacArthur was reviewed and approved* by the following:

Stephanie A. Shields
Professor of Psychology and Women's Studies
Thesis Advisor

Jonathan Cook
Assistant Professor of Psychology

Janet van Hell
Professor of Psychology and Linguistics

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School.

ABSTRACT

Research and anecdotal observations suggest that it is a common occurrence for women to be referred to as "girls" into their thirties and beyond. Although this term is infantilizing and carries with it connotations of immaturity and frivolity, no research has investigated the effects of such language on women's self-perceptions. For my Master's research, I examined this topic by experimentally manipulating the label given to participants in feedback on a work profile test that was ostensibly being used to evaluate them for an advertised leadership position. Through the study outlined in this thesis, I investigated whether being called "girl" rather than "woman" in this feedback influenced women's feelings about their own leadership characteristics (such as strength, confidence, and maturity) as well as their interest in future leadership positions. Results indicated that relative to participants who were called woman, those who were called girl felt less confident that they would obtain the position, reported that the feedback made them feel lower in qualities associated with leadership, and believed that other people who viewed their work profile would also see them as having less of these characteristics. They did not, however, show less interest in future leadership positions. Implications of this research, including the imperative to change current language norms, are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. INTRODUCTION

According to the Meriam-Webster dictionary, the term "girl" is defined as "a female child from birth to adulthood" ("Girl", n.d.). In most U.S. states, adulthood, or the age of majority, is considered to occur at 18 years (Ely & Dulmus, 2010). Therefore, when these two definitions are combined, it becomes clear that girl should technically refer to a female from the time she is born until the day she turns 18, or from birth to age 17. However, when we look at how the popular media defines who is a girl, a different story emerges: numerous examples from movies, books, and television portray women in their late teens, twenties, and even early thirties (and beyond) as girls. For example, the popular television program "Girls" features four college graduates in their mid-twenties, while the title of the prime-time show "New Girl" refers to an elementary school teacher in her early thirties.

Popular media is not the only place where such language can be observed. Indeed, referring to women as girls seems to be commonplace in everyday language and conversation (see "Preliminary Research" below), and it is not difficult to recall instances of its occurrence. For example, in a recent Psychology 100 lecture at Penn State, the instructor conducted an in-class "guys versus girls" demonstration, in which the "girls" were young women in the class between the ages of 18 and 20. What is interesting to note here is that the term "guys" was used as the gender contrast to girls, rather than the semantically more similar term "boys." While girl tends to be a commonly used term for women, the same is not true for the male equivalent, boy. As Lakoff (1973) puts it:

> One seldom hears a man past the age of adolescence referred to as a boy, save in expressions like 'going out with the boys', which are meant to suggest an air of adolescent frivolity and irresponsibility. But women of all ages are 'girls…' (p. 61)

Despite the seeming frequency of this phenomenon, as well as its relative uniqueness to women, little research has investigated the impact that being referred to as a girl may have on women.

The purpose of my Master's research, therefore, is to shed light on the phenomenon of women being referred to as girls and its potential consequences for women, given that girl is an infantilizing term that carries with it child-like and adolescent connotations (see "Preliminary Research" below). In the following sections, I discuss first, research supporting the idea that language influences our perceptions of the world and plays an important role in constructing our social reality, and second, research specifically relating to women being referred to as girls. Next, I discuss the results of my previous work related to the "girls" phenomenon. I then outline the methodology and results of my Master's research, and conclude by discussing the potential implications, as well as the limitations, of this research.

Importance of Language for Constructing Reality

Changing language norms has been a significant goal of the feminist movement for many years (Felski, 1989). Early work, for example, drew attention to the fact that supposedly equivalent terms for men and women were not in fact equal: female terms tended to take on more negative connotations over time. Indeed, this "semantic derogation" of women was said to occur in word pairs such as *bachelor-spinster, master-mistress, governor-governess, and warlock-witch*, in which the feminine word held more negative associations than the masculine word (Shulz, 1975).

The feminist belief that sexist language, such as semantically derogating terms, should be replaced with non-sexist alternatives is based on an assumption that language does not simply reflect reality, but also constitutes it. In other words, it indicates that one believes in the ability of language to influence the way women are perceived. Although the original (strong) version of

this hypothesis, which claims that language *determines* how people think, has been refuted by a significant body of research (e.g., Alford, 2011; Santa & Baker, 1975), there is substantial evidence to suggest that in some circumstances, language influences the way people perceive the world (e.g., Ehrlich, 2001). Indeed, research has shown that language can influence our perceptions and judgments, and may have consequences for how we think about ourselves and others. In some cases, these consequences may be gendered, in that language choices may either reinforce traditional hierarchies or work to subvert them.

One of the most prominent examples of work that shows how gendered language can influence thought comes from research on the use of the so-called masculine generic, or saying "he" when one actually means "he or she." Many studies (e.g., Crawford, 2001; Lambdin, Greer, Jibotian, Wood, & Hamilton, 2003; Romaine, 1999; Weatherall, 2002; Gastil, 1990) have demonstrated that use of the masculine generic leads participants to think of male subjects rather than subjects of both genders. Furthermore, it has been found (e.g., Briere & Lanktree, 1983) that the use of such language influences perception of career choices: when students read a paragraph about career paths in psychology, those who saw a gender neutral version believed psychology was a more appealing career option for women than those who saw a masculine generic version. Similarly, Stout and Dasgupta (2011) showed that when masculine generic language was used in an interview, women reported less belonging, motivation, and expected identification with the job than those who experienced gender-inclusive or gender-neutral language. Importantly, masculine generics have also been shown to affect people's perceptions of themselves. A study by Henley, Gruber, and Lerner (1988), for example, found that one week after sixth grade children had been read stories with either masculine generic, feminine generic, or unbiased language, male students who had been in the masculine pronoun condition reported greater

increases in self-esteem that those in the other conditions, while girls who had been in the neutral pronoun condition reported greater increases in self-esteem than those in the feminine or masculine generic conditions.

Another line of research has looked at the consequences of "asymmetrical gender-marking," which occurs when the gender of one group is not mentioned, but is highlighted for another group. An example of gender marking comes from Penn State athletics: the women's basketball team is known as the "Lady Lions," whereas the men's team is simply referred to as the "Lions." In this case, gender is explicitly marked for the female team. Feminist scholars (e.g., Pratto, Korchmaros, & Hegarty, 2007) have argued that gender marking casts one group (typically men) as the norm, and another (usually women) as deviant, or the effect to be explained. Research has shown that gender marking has significant consequences. For example, participants in a study by Bruckmüller and Abele (2010) perceived a group as more powerful and higher in status when it had been positioned as the norm (e.g., when it was not marked with gender) than when it had been positioned as the effect to be explained (e.g., was gender-marked) in a text comparing two groups. Norm groups were also perceived as more agentic and less communal than the marked groups. Another study (Bruckmüller, Hegarty, & Abele, 2012) showed that participants were quicker to apply gender stereotypes when reading about gender differences in leadership when these differences were framed around a male rather than female linguistic norm. Participants also perceived status differences between the genders to be larger and more legitimate when the male linguistic norm was used. Interestingly, there was no effect of linguistic norm when participants read about gender differences in leisure time preferences, indicating that linguistic framing effects generalize only to contexts in which there are gendered status differences.

Research also shows that more general use of gendered and sexist language affects perceptions. For example, a study by Gaucher, Friesen, and Kay (2011) revealed that when job advertisements were worded so as to include more "masculine" than "feminine" wording, participants believed more men worked within these jobs, and women perceived the jobs as less appealing. This effect was mediated by perceptions of belonging, rather than perceptions of having the skills necessary to perform in the occupation. Such findings show that even small changes in wording can significantly alter people's perceptions, including those related to career choices.

Women Being Referred to as "Girls"

To date, little research has focused specifically on the issue of women being referred to as girls. The literature that is available, however, suggests that it is a common occurrence. For example, in a survey of the spoken commentary accompanying men's and women's televised sports games (basketball and tennis), Messner, Duncan, and Jensen (1993) found that women were commonly referred to as girls by the commentators, whereas men were never referred to as boys. Further evidence that women are considered girls into adulthood comes from Lerner (1976), who noted psychotherapists' frequent use of the word girl to describe female clients in a clinical setting, and from Bebout (1984), who asked participants to name what they believed to be the upper age limit of several labels for males and females. Results of Bebout's study indicated that people believed the term boy could refer *at the oldest* to an older child (28%), young adolescent (39%), or mid-adolescent (33%), while the upper age limit for girl was seen as mid-adolescent (20%), older adolescent (31%), young adult (35%), or young mature (14%). These results make clear that the term girl can be used to refer to women as well as children, and that this use of the term is widely understood.

In her analysis of this phenomenon, Lakoff (1973) suggests that the term girl connotes immaturity, frivolity, innocence, and irresponsibility, and thus when used in everyday language to describe an adult woman, signifies that women are physically and intellectually weak, as well as incapable of taking on responsibility or making important decisions. As the literature described above indicates, language not only reflects, but constitutes reality: even seemingly minor changes in wording can create vastly different connotations and can impact people's perceptions of themselves and others (e.g., Bruckmüller, Hegarty, & Abele, 2012; Gaucher, Friesen, & Kay, 2011; Henley, Gruber, & Lerner, 1988). Therefore, when a woman is referred to as a girl, this labeling cannot simply be seen as neutral, but must be seen as a marker that shapes the associations and judgments that we make about others.

Evidence to support this claim comes from the work of Brannon (1978), who had participants read a story about either a "girl" or a "woman" competing for a high-level executive position, after which they rated her on a number of job-related and personality characteristics. Results indicated that when labeled woman, the target was seen as more mature, tough, dignified, brilliant, and deserving of a high salary than the target in the girl condition. On the other hand, a study conducted by Lipton and Hershaft (1984) showed no differences in the evaluation of artists and artwork when artists were labelled girls versus when they were labelled women. In this study, researchers had participants evaluate an artist and a series of paintings on several affective and cognitive measures. Using the same stimulus materials for each condition, they varied only whether the artist was identified as a "man," "guy," male "person," "woman," "girl," or female "person." Results indicated that there was a main effect of gender, such that participants rated male artists and their paintings more favorably than female artists and their paintings; however, the specific label used to indicate male or female did not affect assessment of the artists. While

this result might seem troubling for the hypothesis that the term girl negatively affects perceptions of the woman who is labelled, it is likely that Lipton and Hershaft did not find significant results for their label manipulation because the negative connotations associated with the term girl (and the positive connotations associated with the term woman) were not relevant to the context being studied. As I describe in the "Preliminary research" section below, the major connotative differences between girl and woman include characteristics such as strength, independence, confidence, maturity, adultness, and dignity, none of which are particularly relevant to the evaluation of artists and artwork. Indeed, a study by Dayhoff (1983) investigating the effects of sexist language on the evaluation of candidates for office revealed that linguistic sexism (which included use of the term girl to describe women candidates) led women to be negatively evaluated only when seeking a "masculine" or "neutral" office, where qualities such as strength and confidence would be perceived as highly relevant. This likely explains why Brannon found a significant difference in the evaluation of a woman versus a girl, while Lipton and Hershaft did not: Brannon situated his study in the context of a masculine and highly competitive domain, where possessing qualities associated with the term girl would be perceived as a major hindrance to performance.

Although Brannon's (1978) study speaks to the ability of a particular label to impact *observer's* perceptions of women, no research has systematically investigated what such labels do to women's perceptions of *themselves*. If being called girl makes others think of a woman as less mature and dignified, might she also start to feel the same about her own abilities? The present research attempts to answer this question by examining the effect of infantilizing labels from the target's perspective. Although much of the research cited in this thesis comes from previous decades, and language norms do change over time (Pinker, 1999), recent pilot tests (see

"Preliminary Research") confirm that girls, when compared with women, are still associated with innocence and childishness, just as Lakoff asserted in 1973.

While infantilizing labels can be seen as problematic for all women, this issue may be especially problematic for women in their late teens to mid-thirties. First, as was made evident from the examples mentioned at the beginning of this paper, this age group is the one that seems most likely to be called girls, and secondly, these women may not have established their adult identities to the same extent that older women have. Indeed, women in their late teens and early twenties are just beginning to transition into independence and adulthood, and women in their late twenties and thirties are continuing to establish their roles and identities in the "adult" world. Being referred to as girl is more likely to be harmful for these women, given its potential to make them feel less adult, mature, and professional. This is particularly concerning when it comes to possible employment consequences, as women who do not feel confident in these qualities may be less likely to take on leadership roles that require responsibility, and may therefore miss out on the greater pay and status that accompany it.

Given the above-mentioned literature on the role that language plays in influencing the way we see the world, as well as the literature on women being referred to as girls, I expected that relative to those who are called woman, participants who were called girl in an experimental setting would: a) report feeling less likely to obtain an advertised leadership position, b) report feeling less mature, adult, and professional (as well as other leadership-associated qualities), c) expect others (e.g., external evaluators) to see less of these qualities in them, and d) be less interested in future leadership opportunities.

In addition to these hypotheses, I anticipated that being referred to as a girl would have an adverse effect on women's cognitive performance: women who are called girls were expected

to perform worse on a test of working memory than those who are properly labelled women. Indeed, a large body of research on stereotype threat shows that when negative stereotypes about women are primed or explicitly mentioned in the context of a stereotype-relevant performance domain, women's performance tends to decline (e.g., Davies, Spencer, Quinn, & Gerhardstein, 2002; Gonzales, Blanton, & Williams, 2002; Nguyen & Ryan, 2008; Schmader, 2002; Spencer, Steele, & Quinn, 1999). More recent work (e.g., Schmader & Johns, 2003; Beilock, Rydell, & McConnell, 2007) has shown that in such situations, women (and other stereotyped groups) begin to worry about confirming the negative stereotypes associated with their group. This worry has a distracting effect, using up cognitive resources and interfering with working memory capacity. Through this threat process, women may face performance deficits in domains like math and science, where stereotypes about women's abilities are plentiful. I do not intend to suggest here that being referred to as a girl activates stereotypes about women's performance in the same way as occurs in stereotype threat; however, such infantilizing language may nevertheless interfere with working memory if it is perceived by women as threatening in the context of a male-stereotyped domain such as leadership. Indeed, leadership was chosen to provide the context and cover story for this study because the qualities that are perceived to differ between girls and women (e.g., confidence and maturity; see "Preliminary Research" below) are highly relevant to leadership fields, and because it should be both familiar and relevant to undergraduate students.

Preliminary Research

Thus far, I have conducted four studies on the phenomenon of women being referred to as girls. In this Preliminary research section, I describe the results of two pilot studies that are

relevant to the present research, as well as one ongoing study being conducted in collaboration with a researcher at the University of Arizona.

Pilot study 1. My first pilot study on the topic of women being referred to as girls was conducted in Fall 2013. For this study, undergraduate participants were recruited from the Psych 100 participant pool, and were asked to respond to a series of online questions assessing which labels they use for males and females of various ages, how often they use different terms, how often they are called certain terms themselves, and how much they enjoy being called each one. The main objective of this survey was to determine which terms are used most frequently, as well as which terms undergraduates most and least prefer to be called by different people.

Results. Several features of Pilot Study 1 results are worth noting (see Table 1). Firstly, there is a marked asymmetry in the terms that are most commonly used to describe women and men, as well as in the terms preferred by each gender. For women, girl was the most frequently used term as well as the preferred term in most contexts. For men, "guy" or "dude" was the most frequently used term, while "man" was preferred in most contexts. Interestingly, qualitative responses to questions asking participants why they liked certain labels indicated that women understood the term "woman" to connote maturity and adultness; however, they often felt unready to accept these positive traits, as they did not feel they had reached the requisite level of maturity yet. In contrast, for men, the term "man" was seen as a sign of respect and maturity, and participants seemed quite ready to accept these associations. In terms of most offensive labels, women universally disliked "chick," while "boy" was universally seen as a sign of disrespect for men.

Table 1
Pilot Study 1 Results

Category	Modal female term	Modal male term
Term participants most frequently use to describe an adult (18+)	Girl	Guy
Term most frequently used by others to describe participants	Girl	Dude
Term most preferred (general)	Girl	Gentleman/Man
Term most preferred from someone older	Young woman	Young man
Term most preferred from someone the same age	Girl	Man
Term most preferred from someone younger	Girl	Man
Term most preferred from a female	Girl	Gentleman
Term most preferred from a male	Girl	Man
Most offensive term (general)	Chick	Boy
Most offensive from someone older	Chick	Dude/Boy
Most offensive from someone the same age	Chick	Boy
Most offensive from someone younger	Chick	Boy
Most offensive from a female	Chick	Boy
Most offensive from a male	Chick	Boy

Pilot study 2. My second pilot study was conducted in the Spring of 2014. The purpose

of this pilot was to determine the connotative differences between the terms girl and woman.

Participants were recruited from the participant pool, and were told that they were taking part in

a study on word associations. They were informed that Qualtrics would randomly assign them a

word or phrase, and instructed that they should respond to questions about that word or phrase

without thinking too much about their answers. Participants were then told that the word they

had been assigned was "girl," "woman," or "young woman." They were asked to list the first 3-5 words that came into their heads when thinking about their assigned word. Following this, they were instructed to indicate how much they associated each of 50 qualities (including both positive and negative traits that were hypothesized to relate to girl and woman) with their word, as well as to indicate how likely a girl, woman, or young woman was to be married, have children, and be of high status.

  Results. Results of the pilot indicated that there were several significant differences between girl associations and woman associations. Unfortunately, young woman had to be left out of the analysis because participants in this condition responded with consistently lower scores than participants in the other two conditions, meaning that almost all traits were found to be significant because young woman was less associated with most traits (even those that were contradictory, like "strong" and "weak") than were the labels in the other two conditions. Given that young woman is not as commonly used as girl or woman, this likely occurred because participants in the young woman condition simply had less strong associations with the phrase.

  When comparing just the associations of woman with the associations of girl, several significant differences were found. The qualities that were significantly (or marginally) more associated with woman than with girl included: secure, strong, mature, intimidating, sophisticated, focused, dignified, adult, financially secure, independent, successful, serious, decisive, entitled, proactive, professional, and confident. The traits that were significantly (or marginally) more associated with girl than with woman were: childlike, naïve, innocent, and playful. Results also indicated that a woman was significantly more likely to be married and have children than a girl.

Ongoing research. The study I am currently working on related to the use of various labels for women is being conducted with Dr. Matthias Mehl at the University of Arizona, using previously collected Electronically Activated Recording (EAR) data. The EAR is a sound sampling device created by Mehl, Pennebaker, Crow, Dabbs, and Price (2001) that uses a microphone and digital recorder to record sound samples from the daily lives of participants. Participants wear the device for up to 7 days, during which ambient sounds from their environment are recorded for 30 second intervals every 12 minutes. Since introducing the EAR in 2001, Mehl has collected a database of sound samples from 228 undergraduate participants, which has been made available to other researchers for use in scientific studies. Using transcripts of the recordings, Dr. Mehl and I are seeking to determine how often different labels are used to describe women and men in everyday conversation. Recently analyzed results indicate that the proportion of "girl" labels for adult women is significantly greater than the proportion of (a) "woman" labels for adult women and (b) "boy" labels for adult men. These findings provide support for my hypothesis that women are frequently referred to as girls in their everyday lives, and consequently, that it is crucial to assess the impact of this language on women.

The Present Study

Given the frequency with which women are called girls in everyday conversation, the association of the word girl with naivety, innocence, and childishness (see "Preliminary Research" above), as well as the potential detriments to women of being referred to by this infantilizing term, the present study experimentally investigated the impact on women of being called girl.

Hypotheses. In this study, I tested five hypotheses. I predicted that relative to women who are properly referred to as women, those who are called girls in feedback on an application for a leadership position would a) report feeling less confident about their likelihood of obtaining the position, b) report that the feedback made them feel less mature and professional (and other qualities associated with leadership), c) report an expectation that others (e.g., external evaluators) will see less of these qualities in them, d) report less interest in future leadership opportunities, and e) perform worse on a test of working memory. I further hypothesized that women's ratings of how the feedback made them feel, as well as their beliefs about how others (e.g., external evaluators) will perceive them, would mediate the relationship between label manipulation (e.g., whether they were referred to as a girl or a woman) and their interest in future leadership opportunities. Finally, I expected that the gender of the labeler (the person who labels the participant either girl or woman) would moderate these effects, such that the impact of being called girl would be greater when coming from a male than from a female labeler.

Gender of labeler as a moderator. While I expected that being referred to as a girl would result in decreased feelings of confidence and maturity for women regardless of who does the labeling, I also hypothesized that the gender of the labeler would moderate participants' ratings. Because men tend to have higher social status than women (e.g., Carli, 1999), they also tend to have more influence over others: people generally believe that men are more competent and knowledgeable than women, and that they have more of a right to act as authorities (Carli, 2001). Because of this advantage in influence, I predicted that when men label women as girls, it will have a greater impact on women's self-perceptions than when they are given the same label by a woman. Some research (e.g., Bianchi, 2014; Rahman, 2012) also suggests that labels may take on different meanings when coming from in-group members versus out-group members,

and this idea is supported by my pilot study data (see "Pilot Study 1" in "Preliminary Research"):

less women chose girl as the most desired label when coming from a man than they did when

coming from a woman.

Figure 1 below depicts the proposed conditional mediation model, with gender of labeler

(B) moderating participants' feelings of leadership and professionalism (C) and others'

perceptions of leadership and professionalism (D), and these two variables (C and D) mediating

the relationship between label manipulation (A) and interest in future leadership opportunities

(E).
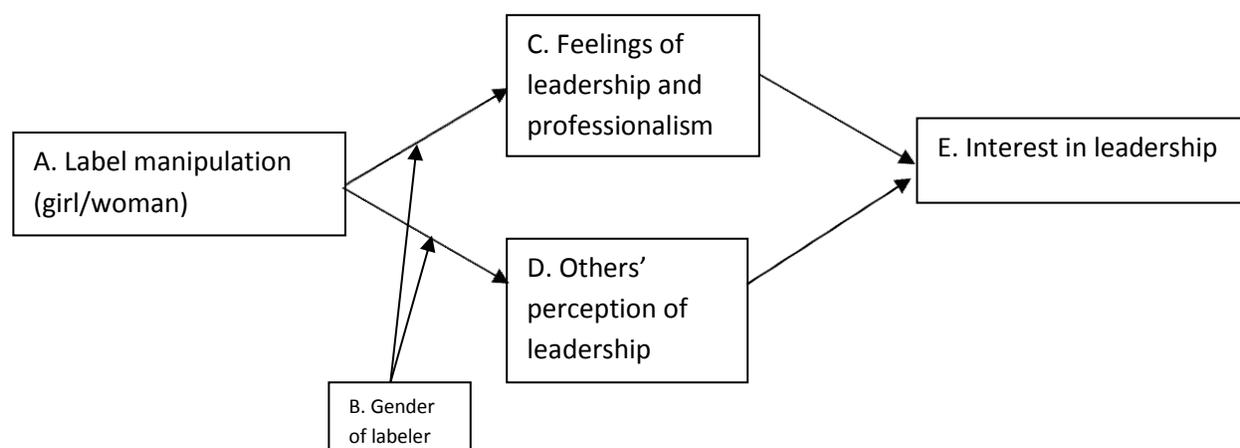
Figure 1. Proposed conditional mediation model.

Global self-esteem as a moderator. Although most people tend to feel badly about

receiving negative feedback and happy about receiving positive feedback, there is some evidence

to suggest that certain individuals may be more susceptible to evaluative feedback than others.

For example, studies (e.g., Blaine & Crocker, 1993) indicate that those with low global self-

esteem have a tendency to interpret a single poor performance as an indicator of overall ability,

and generally have weaker self-serving attributional biases: they see positive and negative

feedback as equally credible (Shrauger, 1975). Although participants in the present study did not

receive negative feedback about their performance per se, self-esteem was included as an exploratory predictor to determine if individuals with low trait self-esteem respond differently to being called girl in an evaluative context than do individuals with high trait self-esteem. Participants with low self-esteem could be particularly vulnerable to the manipulation, as they may be more inclined to judge themselves in accordance with the feedback they receive.

Chapter 2. METHOD

Participants

A total of 438 female undergraduate students from the Penn State introductory

psychology participant pool were recruited for the study in exchange for course credit; however,

178 (40.7%) of these were excluded from analyses for a variety of reasons. Given that the same

word (or concept) can have different meanings in different cultures and languages (e.g., Beaton,

Bombardier, Guillemin, & Ferraz, 2000; Berry, 1980; Brislin, 1970), 51 participants (11.6%)

were excluded for indicating that English was not their first language. Other participants were

excluded for failing attention/manipulation checks, reporting suspicion about the purpose of the

study or the fake nature of the feedback, and/or because of incomplete data or procedural errors

that occurred during the experiment (see Table 2). The decision to exclude participants for these

reasons was made *a priori,* and data collection continued until a pre-determined number of

participants (250) had been reached, taking these exclusions into account.

Most troubling among these exclusions is the fact that nearly 20% of the sample was not

able to correctly identify the gender of the professor who provided them with feedback, even

though this professor's name (Stephen or Susan C. Johnson) was mentioned several times

throughout the study, and an email bearing the professor's name was received directly by

participants. Importantly, a regression analysis indicated that the likelihood of making this error

was not equally distributed across conditions, $b = -.18$, $t (428) = -4.79$, $p < .001$, such that

participants in the male professor conditions were significantly more likely to make an error in

remembering the professor's gender than were participants in the female professor conditions.

This finding is consistent with research showing that memory for expectancy-incongruent

information tends to be better than memory for expectancy-congruent information (Stangor &

Table 2
Study Exclusions

Type of Error	Number of participants
Failed attention check[1]	39 (8.9%)
Correctly guessed hypotheses of study	2 (.46%)
Had suspicions about veracity of feedback[2]	22 (5.02%)
Procedural errors/missing data	11 (2.51%)
Manipulation check- failed to correctly remember professor gender	80 (18.26%)
Manipulation check- failed to correctly remember the word used in their profile (e.g., girl or woman)	25 (5.71%)
Indicated English was not their first language	51 (11.64%)
Total participants excluded	**178 (40.64%)[3]**

McMillan, 1992). Given women's under-representation in leadership fields (Eagly & Karau, 2002), a high-status female professor in the Smeal College of Business may well have represented such an expectancy violation for participants. Also of interest is the fact that a much smaller number of errors were made in recalling the label used (girl or woman) in participants' work profiles. Indeed, it seems that this information was more salient to participants than the gender of the professor providing the feedback, perhaps because it was more self-relevant and relevant to the goal of obtaining the leadership position (and associated prize money). Errors in remembering the label used were not significantly different across conditions. Finally, it is important to note that although a high percentage of participants were excluded based on criteria

---

[1] The attention check was embedded within the Others' Perceptions of Leadership and Professionalism scale and asked participants to "Please choose 'Not at all.'"

[2] The suspicion probe was an open-ended question near the end of the study asking participants if anything about their work profile, or the study in general, seemed unusual.

[3] This number is less than the sum of all the row above because of overlap in exclusions (e.g., people who failed the attention check may have also failed the manipulation checks, etc.).

decided *a priori*, the inclusion of participants from one or all of these categories did not change results for any of the main dependent variables.

   After exclusions, the final sample consisted of 256 female undergraduate participants. While undergraduate students are often thought of as a convenience sample, in this case, the young adult population is a particularly relevant demographic. The mean age of the sample was 18.75 years, and thus within the demographic range for which this research is most salient. The sample was also predominantly White (80.5%), followed by Black or African American (6.6%), Latina/o or Hispanic (6.3%), and Asian (3.9%).

Materials

   Self-esteem. To measure self-esteem, I used the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). The RSES consists of 10 items that measure global self-esteem, on a scale ranging from 1 (*Strongly Agree*) to 4 (*Strongly Disagree*; see Appendix A). Previous studies have reported the reliability of this scale to be in the range of .72 to .90 (Robins, Hendin, & Trzesniewski, 2001). For use in the present study, appropriate RSES items were reverse scored and items were then summed and averaged to create a single self-esteem variable (α= .86), in which higher scores represent greater self-esteem.

   Likelihood of obtaining the position. To measure participants' confidence that they would be chosen for the leadership position, I used a single item ("Now that you have received and read your work profile, how confident do you feel that you will be chosen for the leadership position?"), measured on a 7-point Likert scale ranging from 1 (*Not at all confident*) to 7 (*Extremely confident*). Higher scores on this variable represent greater confidence about being chosen for the leadership position.

Feelings of leadership and professionalism. Participants' feelings of leadership and professionalism were measured using a 7 item scale. The item asked participants to think about the profile they received and to indicate the extent to which it made them feel each of the following: secure, strong, mature, adult, decisive, focused, and professional. Items were measured on a 7-point Likert scale ranging from 1 (*Not at all*) to 7 (*Extremely*). The seven items were summed and averaged to create a single composite score (α= .95), with higher scores indicating greater feelings of leadership and professionalism.

Others' perceptions of leadership and professionalism. To measure the extent to which participants believed that others (i.e., the committee of leaders who were supposedly evaluating their profile) would see qualities of leadership and professionalism in them, I used the same seven items as those in the Feelings of Leadership and Professionalism scale, except framed to focus on the perceptions of the committee ("How confident do you feel that the committee members who are evaluating your profile will perceive you as secure, strong, mature, adult, decisive, focused, professional?"). Once again, this construct was measured on a 7-point Likert scale ranging from 1 (*Not at all*) to 7 (*Extremely*). Scores on this variable were summed and averaged to create a single measure (α= .95), with higher scores indicating greater confidence that committee members would see characteristics of leadership and professionalism in their profile.

Interest in future leadership opportunities. Interest in future leadership opportunities was assessed using three items adapted from Davies, Spencer, and Steele (2005). These items asked participants how willing they would be to participate in future studies, as either a "leader" or a "team-member":

In our lab, we often do studies on the subject of leadership. In these studies, we typically recruit people to take on the roles of either "**leaders**" or "**team-members**." Leaders supervise other students, while team-members work collectively on an activity.

Some of our previous research has shown that the most effective leaders in these studies have the qualities of confidence and maturity, while the most effective team-members are good team players and have excellent communication skills.

How interested would you be in participating in one of our studies as a leader?
How interested would you be in participating in one of our studies as a team-member?

These first two items ("How interested would you be in participating in one of our studies as a leader?" and "How interested would you be in participating in one of our studies as a team-member?") were measured on a 7-point Likert scale ranging from 1 (*Not at all interested*) to 7 (*Extremely interested*). A single score for these items was obtained by calculating a leadership interest difference score, in which interest in the team-member role was subtracted from interest in the leadership role. Higher scores on this variable indicate greater interest in leadership roles relative to team-member roles.

A third item for this measure asked participants to indicate which role they would choose if they were required to participate in future studies as either a leader or a team-member. Responses on this item were multiple choice, with the options being "leader" and "team-member." Given that responses to this question were categorical, the third item was analyzed separately from the first two using a logistic regression.

Working memory. To measure participants' working memory, I used an Operation Span Task (OSpan; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004; Unsworth, Redick, Heitz, Schrock, & Engle, 2005). Specifically, I used Foster, Shipstead, Harrison, Hicks, Redick, and Engle's (2014) automated shortened version of the Operation Span.

The OSpan measured working memory by presenting participants with strings of letters, ranging in length from three to seven, that they were instructed to remember in the correct order.

In between each letter presentation was a distractor math problem that participants were asked to solve, thereby making it more difficult to remember previously presented letters. Throughout the task, participants were reminded that they must answer at least 85% of the math problems correctly; this was to ensure that participants put their full effort into solving the problems, rather than simply focusing on remembering the letters. Participants in the present study completed four blocks of the OSpan: the first three blocks were practice trials, and the fourth consisted of the actual trials on which data analysis was based. In this final block, participants were asked to recall five strings of letters, from which a "partial score" was generated that indicated how many total letters were correctly recalled in the right order (Turner & Engle, 1989). The partial score was used to represent a participant's overall working memory score, with higher scores (maximum 31) indicating greater working memory capacity.

Although the creators of the OSpan task used in this study (Foster et al., 2014) assert that it is important to use more than one (non-practice trial) block of their shortened memory tasks to properly assess working memory capacity, time constraints prevented me from doing so. Asking participants to complete even one shortened block (plus practice blocks) of the OSpan required an allotment of 20 minutes in the experiment, and considering the other measures participants were expected to complete, this was the maximum amount of time that could be reasonably allocated to the working memory task. Because working memory analyses were exploratory and not used to draw firm conclusions about participants' working memory capacity, I decided to proceed using one block of trials, keeping in mind that future research could follow up on any trends in the data with more robust measures.

Exploratory measures. Several exploratory measures were included in the study. These variables were added because they were expected to provide additional information as to how

participants felt about the experience of being called girl, or about potential mechanisms involved in confirmatory or null results.

       Liking of feedback. The extent to which participants liked the feedback they received in their work profiles was measured with a single item: "how much did you like the feedback you received from Dr. Johnson?" This was measured on a Likert scale ranging from 1 (*Extremely disliked*) to 7 (*Liked very much*).

       Perceived accuracy of feedback. The perceived accuracy of the feedback was measured with a single item asking participants how accurate they thought their profile was in describing them. Accuracy was measured on a 6-point Likert scale from 1 (*Very inaccurate*) to 6 (*Very accurate*).

       Actual job interest. The extent to which participants would be interested in applying for the advertised leadership position if it were an actual job on campus was measured with a single item on a 5-point Likert scale. The scale ranged from 1 (*Not at all interested*) to 5 (*Extremely interested*). The item was worded as follows: "If the student leadership was an actual job being advertised on campus and you were looking for work, how interested would you be in applying?"

       Irritation/offensiveness of word used. The extent to which participants felt irritated with and offended by the word used in their profile was measured with a single item: "How irritating/offensive did you find the word used to describe you in your profile (girl, woman, etc.)?" This item was measured on a 5-point Likert scale ranging from 1 (*Not at all irritating/offensive*) to 5 (*Extremely irritating/offensive*).

       Frequency of using "girl." To assess how often participants refer to their own female friends and peers using the term "girl," a single item was used. This item ("how often do you use

the term 'girl' to refer to female friends and peers?") was measured on a 5-point scale ranging from 1 (*Never*) to 5 (*Always*).

Frequency of using "woman." To assess how often participants refer to their own female friends and peers using the term "woman," a single item was used. This item ("how often do you use the term 'woman' to refer to female friends and peers?") was measured on a 5-point scale ranging from 1 (*Never*) to 5 (*Always*).

Procedure

Participants were recruited to the study, titled "Predicting Success in Leadership Positions," through the Penn State SONA system and received course credit for their participation. To take part in the study, participants were asked to come in person to a computer lab in Moore building, with up to six participants scheduled at a time. Upon arrival at the lab, they were greeted by an experimenter (one of four research assistants) who informed them that the study was related to "leadership and how selection procedures are experienced by job applicants" and that in the study, they would be asked to compete for a mock leadership position. They were further informed that by participating in the study, they had a chance to win a $50 Amazon gift card, which would be awarded to the student or students judged to be best suited for the leadership position.

After reading and signing consent forms, participants were seated at a computer and directed to complete a (hard-copy) of the Rosenberg Self-Esteem scale. After completing the scale and placing it in a folder next to their computers, they were asked to begin the computerized survey that was open in front of them. At this point, participants read a description of the leadership position they would be applying for—a student position available on campus (see Appendix B for full wording). Next, participants saw the following text:

To assess your fit for the position just described, we would like you to fill out a work profile test that has been developed in our lab. <u>The test measures characteristics that are diagnostic of future success, and it is a reliable indicator of future performance in leadership positions.</u>

When you have finished answering all of the questions below, your responses will be sent to ***Stephen/Susan C. Johnson***, a professor of leadership in the Smeal College of Business who is collaborating with us on this project. Dr. Johnson will use your responses to generate a work profile for you, and will then send this profile on to a committee of leaders for evaluation.

On the next page, participants saw a profile of the professor who would be generating their profile, complete with information about his/her educational background and a message to participants. This part of the procedure was adapted from London, Downey, Romero-Canyas, Rattan, and Tyson (2012), who used a similar method to assign students fake feedback about their performance on a written essay.

Your evaluator is Dr. Susan/Stephen C. Johnson.

<u>Profile:</u>
Susan/Stephen C. Johnson
Educational Background: Ph.D., Harvard, 1978
Current Position: Full Professor, The Pennsylvania State University

<u>Message To Participants:</u>
Over the course of my career, I have realized how important it is for students to come across as competent leaders. I have found that the work profile questionnaire you are about to fill out is an excellent indicator of students' future leadership performance. I will evaluate your profile based on the personal characteristics that come across in your answers to the questions provided.

I will be forwarding each profile I generate on to a committee of community leaders, and this committee will ultimately decide who gets the $50 gift card for leadership. These leaders will be evaluating the work profiles of all participants in this study to determine which one indicates the best fit for the leadership position.

Good luck!

At this point, participants completed the "work profile questionnaire," which actually consisted of filler questions obtained from a Myers-Briggs personality-type questionnaire ("Jung

Typology Test", n.d.; see Appendix C for a list of filler questions). After completing this questionnaire, participants were given a packet of mazes and sodokus to work on as they waited for the professor (Susan or Stephen Johnson) to ostensibly generate their profiles.

After several minutes, the experimenter, who was working on a laptop in the room with participants, sent an email from one of two accounts bearing the professor's name (Stephen C. Johnson or Susan C. Johnson) to email accounts set up on each of the lab computers. Participants were instructed to open these emails upon receipt. The subject heading of the emails was "Work profile" and the text simply said "Please see attached." Attached to each email was a PDF document, formatted with the Smeal College of Business letterhead (see Appendix D for full wording). Immediately under the letterhead was a brief message to participants from Professor Johnson, which read as follows:

> Dear participant,
> I read through your answers to the work profile test, and generated this profile for you. In my opinion, you are a very strong example of a **Modern Girl/Woman**. Please read your profile carefully so that you will be prepared to answer questions about it in the next portion of the study.
> -Dr. Stephen/Susan C. Johnson

Underneath this message was the participant's work profile, which referred to them either as a "Modern Girl" or a "Modern Woman." To avoid confusion for experimenters, all participants within one study session received the same version of the profile (i.e., were all assigned the same condition). Participants were given a few minutes to read their profiles, and were then asked to open a second survey window on their computers. The opening text read:

Thank you for completing the work profile test! Now that you have seen your profile, it will be sent to the selection committee (made up of community leaders), who will decide which participant(s) get the $50 Amazon gift card.

We would now like to ask you about your experience in applying for the leadership position. **<u>The information you give here will NOT be used to evaluate you-- it is for research purposes only-- so please be honest</u>**.

At this point, participants responded to the primary dependent measures (likelihood of obtaining the position, feelings of leadership and professionalism, others perceptions of leadership and professionalism, and interest in future leadership opportunities). Following this, the experimenter administered the automated OSpan task. Finally, after completing the OSpan, participants were directed to complete a post-survey questionnaire in which they responded to the exploratory dependent measures, were probed for suspicion about the study's design and hypotheses, responded to manipulation checks, and completed a demographic survey (see Appendix E for full post-survey questionnaire wording). Following completion of this final portion of the study, participants were debriefed as to the true nature of the research and dismissed. Typical study sessions lasted approximately 45 minutes from start to finish.

Chapter 3. RESULTS

Before running regression analyses to test my hypotheses, I estimated the correlations among all variables, and these are presented in Table 3. Notable among these is the very high (.91) correlation between feelings of leadership and professionalism and other's perceptions of leadership and professionalism, indicating that these may not be entirely separate constructs. Indeed, the high correlation seems to suggest that participants expected the committee members who were supposedly evaluating their profiles to interpret them in the same way that they themselves did. Table 4 shows means and standard deviations for all dependent variables and self-esteem, separated according to label condition (girl or woman).

Table 3
Correlations Among Measured Variables

	1	2	3	4	5	6
1. Self-Esteem	-					
2. Likelihood of Obtaining the Position	.32**	-				
3. Feelings of Leadership and Professionalism	.27**	.65**	-			
4. Others' Perceptions of Leadership	.24**	.64**	.91**	-		
5. Interest in Future Leadership Opportunities [a]	.19**	.09	-.08	-.01	-	
6. Working Memory	-.05	-.08	-.06	-.11	-.01	-

[a] Includes only the continuous items associated with this variable
** $p < .o1$.

Table 4

Means and Standard Deviations Separated by Label Condition

	Label Condition			
	Girl		Woman	
Variable	M	SD	M	SD
Self-esteem	3.41	.04	3.42	.04
Likelihood of Obtaining Position	4.44	.14	4.92	.11
Feelings of Leadership and Professionalism	4.98	.13	5.77	.08
Others' Perceptions of Leadership	4.79	.13	5.70	.08
Interest in Future Leadership [a]	.67	.16	.63	.14
Working Memory	18.91	.43	17.91	.45

[a] Includes only the continuous items associated with this variable

Note. N = 244

To ensure that there were no effects of the particular experimenter/research assistant on the dependent variables, I conducted a 2 (Label manipulation: girl/woman) x 2 (gender of labeler: female/male) x 3 (Self-esteem: high/medium/low) x 4 (Experimenter: Chelsea/Shelley/Janie/Rob) between subjects ANOVA on each of the five main dependent variables (likelihood of obtaining the position, feelings of leadership, beliefs about others' perceptions of leadership, interest in future leadership opportunities, and working memory). No main effects or interactions of experimenter emerged, so experimenters were collapsed for all subsequent analyses.

To determine the effect of the label experimentally assigned to participants and the potential moderating effects of labeler gender and self-esteem, I conducted a series of hierarchical regressions. Because self-esteem was negatively skewed (skewness statistic of -1.0), with almost all participants scoring between 2.5 and 4.0, I performed a transformation by raising self-esteem scores to the power of three. This transformation improved the normality of the distribution, bringing the skewness statistic to -.08. Before carrying out my analyses, I also mean centered the transformed self-esteem scores, so as to minimize collinearity and make interactions easier to interpret (Tabachnick & Fidell, 2006).

All hierarchical regressions involved the entry of variables in four steps. Because self-esteem was expected to significantly predict scores on most, if not all, of the dependent variables regardless of the label manipulation, and self-esteem as a main effect was not of major theoretical interest, it was entered by itself in Step 1. In Step 2, the main effect variables of label manipulation and gender of labeler were added to the model. In Step 3, all two-way interactions were entered, though only two of these were of theoretical interest (label X self-esteem and label X gender of labeler). The other two-way interaction (gender of labeler X self-esteem) was included for the sake of interpreting the three-way interaction between label manipulation, self-esteem, and gender of labeler that was entered at Step 4. The results of each regression are discussed below.

Likelihood of Obtaining the Position

Standardized and unstandardized coefficients from the hierarchical regression are displayed in Table 5 below. As expected, Step 1 of the regression indicated a significant main effect of self-esteem, $p < .001$. Step 2 revealed a significant main effect of the label manipulation, $p < .01$, such that participants in the girl conditions reported feeling less confident that they would obtain the leadership than those in the woman conditions. No other main effects or interactions were significant.

Table 5

Standardized and Unstandardized Regression Coefficients for Likelihood of Obtaining the Position

	B	SE	β	t	ΔR²
Step 1					.00***
Self-esteem (SE)	.03	.01	.32	5.43***	
Step 2					.02*
Label Manipulation (LM)	.48	.17	.17	2.83**	
Gender of labeler (GL)	-.08	.17	-.03	-.48	
Step 3					.76
LM x SE	-.01	.01	-.07	-.71	
LM x GL	-.27	.34	-.08	-.80	
GL x SE	.00	.01	.02	.21	
Step 4					.38
LM x SE x GL	.02	.02	.11	.88	

Note. N=265.

* p < .05, ** p < .01, *** p < .001

Feelings of Leadership and Professionalism

Table 6 displays the standardized and unstandardized regression coefficients from a hierarchical regression exploring the effect of the label manipulation, gender of labeler, and self-esteem on participants' feelings of leadership and professionalism. As expected, self-esteem was found to be a significant predictor of leadership feelings in Step 1, $p < .001$. Results at Step 2 revealed a significant main effect of label, $p < .001$, such that participants labeled girl reported feeling less of the qualities associated with leadership and professionalism than did participants who were labeled woman. No other significant main effects or interactions were found.

Table 6

Standardized and Unstandardized Regression Coefficients for Feelings of Leadership

	B	SE	β	t	ΔR^2
Step 1					.00***
Self-esteem (SE)	.02	.01	.27	4.42***	
Step 2					.00***
Label Manipulation (LM)	.74	.14	.31	5.27***	
Gender of labeler (GL)	-.07	.14	-.03	-.53	
Step 3					.36
LM x SE	.01	.01	.06	.62	
LM x GL	-.06	.28	-.02	-.21	
GL x SE	-.02	.01	-.13	-1.61	
Step 4					.17
LM x SE x GL	.03	.02	.17	1.38	

Note. N=253.

*** p < .001

Others' Perceptions of Leadership and Professionalism

Presented in Table 7 are standardized and unstandardized regression coefficients for the hierarchical regression examining the effects of the predictors on participants' confidence that those evaluating them would see them as leaders and professionals. Regression statistics in Step 1 revealed that, as anticipated, there was a significant main effect of self-esteem on participants' ratings, $p < .001$. Step 2 also revealed a significant main effect of label, $p < .001$, such that participants who were labeled girl reported feeling less confident that the committee members who were evaluating their profiles would see them as leaders and professionals than those who were labeled woman. No other significant main effects or interactions were revealed.

Table 7

Standardized and Unstandardized Regression Coefficients for Others' Perceptions of Leadership

	B	SE	β	t	ΔR^2
Step 1					.00***
Self-esteem (SE)	.02	.01	.24	3.95***	
Step 2					.00***
Label Manipulation (LM)	.89	.15	.36	6.18***	
Gender of labeler (GL)	-.06	.15	-.02	-.40	
Step 3					.62
LM x SE	.00	.01	.02	.27	
LM x GL	.07	.29	.03	.25	
GL x SE	-.01	.01	-.10	-1.25	
Step 4					.17
LM x SE x GL	.03	.02	.17	1.39	

Note. N=248.

*** $p < .001$

Interest in Future Leadership Opportunities

Because interest in future leadership opportunities had both continuous and categorical items, two analyses were performed. Table 8 below shows standardized and unstandardized beta statistics for the continuous items. Once again, Step 1 shows an expected and significant main effect of self-esteem, $p < .01$. No other significant main effects or interactions emerged.

Table 8

Standardized and Unstandardized Regression Coefficients for Interest in Future Leadership (Continuous)

	B	SE	β	t	ΔR^2
Step 1					.00**
Self-esteem (SE)	.02	.01	.19	3.04**	
Step 2					1.0
Label Manipulation (LM)	.01	.21	.00	.07	
Gender of labeler (GL)	.01	.21	.00	.06	
Step 3					.18
LM x SE	-.01	.02	-.09	-.93	
LM x GL	-.39	.42	-.10	-.92	
GL x SE	.03	.02	.15	1.68	
Step 4					.90
LM x SE x GL	-.00	.03	-.02	-.13	

Note. N=255.

** $p < .01$

Table 9 shows the unstandardized beta statistics for the logistic regression assessing the impact of the predictors on the (categorical) choice made by participants as to whether they wanted to participate in future studies as a leader or a team-member. Consistent with the hierarchical regression for the continuous items, Step 1 indicated a significant main effect of self-esteem, $p < .001$, and no other significant main effects or interactions were found.

Table 9

Standardized and Unstandardized Regression Coefficients for Interest in Future Leadership (Categorical)

	B	SE	Wald	ΔR^2
Step 1				.00***
Self-esteem (SE)	-.05	.02	10.44***	
Step 2				.55
Label Manipulation (LM)	-.52	.44	1.39	
Gender of labeler (GL)	-.12	.42	.08	
Step 3				.61
LM x SE	.01	.02	.19	
LM x GL	.69	.59	1.41	
GL x SE				
Step 4				.66
LM x SE x GL	.02	.04	.19	

Note. N=255.
*** $p < .001$

Working Memory

A hierarchical regression exploring the effect of the predictors on working memory revealed no significant main effects or interactions.

Mediation Analysis

To test the hypothesized mediation model, I analyzed whether feelings of leadership and professionalism and others' perceptions of leadership and professionalism functioned as mediators of the relationship between label manipulation and interest in future leadership opportunities (the continuous items), with gender of labeler moderating the first two pathways

(see Figure 4 below). This model corresponds to Model 7 in the PROCESS macro for SPSS

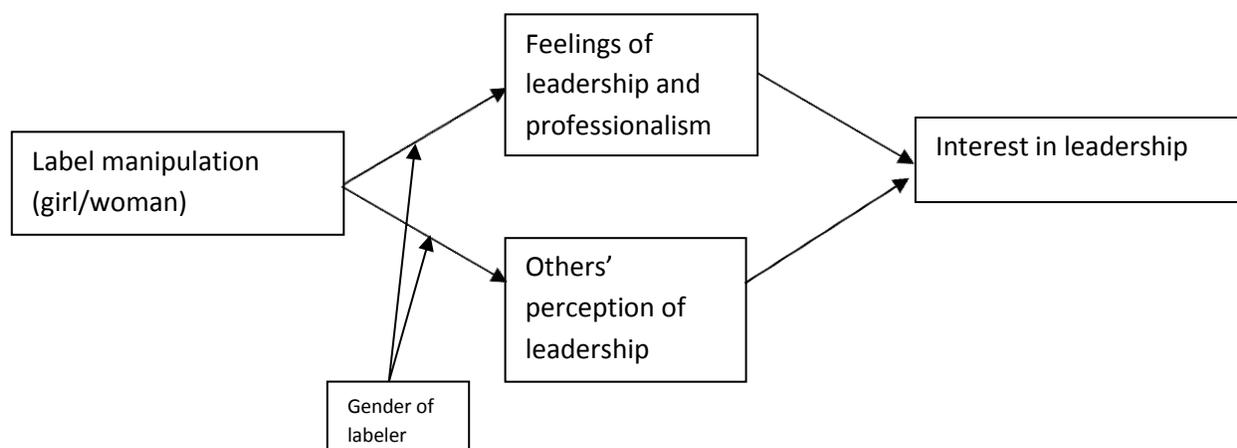(Hayes, 2012), which was used for the mediation analysis.

Figure 2. Proposed conditional mediation model.

Using the PROCESS macro (Hayes, 2012), I entered label manipulation (woman = 1, girl

= 0) as the predictor variable, feelings of leadership and professionalism and others' perceptions

of leadership and professionalism as mediators, gender of labeler (male = 1, female = 0) as the

moderator, and interest in future leadership opportunities (continuous) as the dependent variable.

I selected Model 7 and used 5000 bootstrap samples. Results indicated that the pathway

predicting feelings of leadership and professionalism was significant, $F(3, 240) = 10.01$, $p <$

.001, as was the pathway predicting others' perceptions of leadership and professionalism, $F(3,$

$240) = 12.08$, $p < .001$, with the label manipulation, but not gender of labeler, significantly

predicting both feelings of leadership and professionalism, $b = .73$, $t(240) = 3.50$, $p < .001$, and

others' perceptions of leadership and professionalism, $b = .81$, $t(240) = 3.77$, $p < .001$. The

pathway from both mediators to interest in future leadership opportunities, however, was non-

significant, meaning that feelings of leadership and professionalism and others' perceptions of

leadership and professionalism did not mediate the relationship between the label manipulation

and interest in future leadership opportunities.

Exploratory Analyses

All exploratory analyses were conducted using the same four-step hierarchical structure

as the primary dependent measures. Table 10 shows means and standard deviations for all

exploratory variables, separated according to label condition (girl or woman).

Table 10

Means and Standard Deviations for Exploratory Variables, Separated by Label Condition

| | Label Condition | | | |
| | Girl | | Woman | |
Variable	M	SD	M	SD
Liking of feedback	4.81	.16	5.97	.10
Irritation/offensiveness	2.35	.10	1.29	.06
Perceived accuracy of Feedback	4.63	.20	5.51	.16
Frequency of using "girl"	3.20	.10	3.59	.08
Frequency of using "woman"	2.48	.08	2.20	.07
Actual job interest	3.57	.09	3.67	.08

Note. N=255

Liking of feedback. Step 2 of the hierarchical regression for liking revealed that

participants in the girl conditions liked their profile significantly less than those in the woman

conditions, $b = 1.14$, $t(252) = 6.07$, $p < .001$. No other significant main effects or interactions

were found.

Irritation/offensiveness. Similarly, Step 2 of the regression for irritation/offensiveness

showed that participants in the girl conditions found the word used to describe them significantly

more irritating/offensive than did those in the woman conditions, $b = -1.07$, $t(252) = -9.53$, $p <$

.001. The data revealed no further main effects or interactions.

Perceived accuracy of feedback. Participants in the girl conditions also saw their profile

as significantly less accurate than those in the woman conditions, $b = .86$, $t(252) = 3.43$, $p =$

.001; however, this was qualified by a significant Label X Self-esteem interaction at Step 3, $b =$

.04, $t(250) = 2.44$, $p < .05$. To interpret this interaction, a simple slopes analysis (Aiken & West, 1991) using Hayes' (2012) PROCESS macro for SPSS was performed. As depicted in Figure 2 below, this analysis revealed that participants in the woman conditions perceived their feedback as significantly more accurate when they had high self-esteem (+1 *SD*) than when they had low self-esteem (-1 *SD*), $b = .02$, $t(250) = 1.71$, $p < .05$. As a result of this significant slope, participants with high self-esteem perceived their feedback to be significantly more accurate when they were called woman than when they were called girl, $b = 1.61$, $t(250) = 3.54$, $p < .001$. Participants who were low in self-esteem did not significantly differ in accuracy ratings depending on label condition, nor did those in the girl conditions significantly differ depending on their level of self-esteem. No other significant main effects or interactions were found.
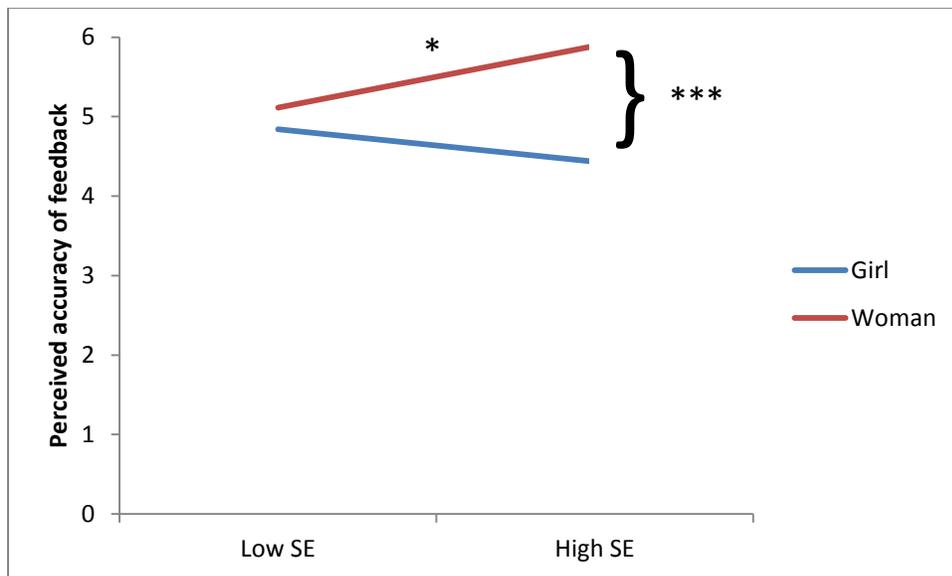

Figure 3. Interaction between label manipulation and self-esteem on perceived accuracy
* $p < .05$, *** $p < .001$.

Frequency of using "girl." Interestingly, participant ratings of how often they use the term girl when referring to female friends and peers was also influenced by the label manipulation. Regression results on Step 2 showed that participants in the girl conditions reported using the term girl significantly less often to refer to female friends and peers than those

in the woman conditions, $b = .40$, $t(252) = 3.07$, $p < .01$; however, this was qualified by a

significant Label X Self-esteem interaction at Step 3, $b = .02$, $t(250) = 2.06$, $p < .05$. To interpret

the interaction, a simple slopes analysis (Aitkin & West, 1991) was conducted using Hayes'

(2012) PROCESS macro for SPSS. As shown in Figure 3 below, participants in the girl

conditions reported using the term girl significantly more often when they had low self-esteem (-

1 SD) than when they had high self-esteem (+1 SD), $b = -.02$, $t(250) = -2.28$, $p < .05$. It was

further revealed that participants with high self-esteem reported using the term girl significantly

more often when they were in the woman conditions than when they were in the girl conditions,

$b = .85$, $t(250) = 3.63$, $p < .001$. Participants with low self-esteem did not differ depending on

condition in reported frequency of using the term girl to describe female friends and peers, nor

did participants in the woman conditions differ depending on self-esteem. One possible

interpretation for this finding is that participants with high self-esteem found the girl label

particularly aversive, leading them to reactively report that they do not use this label to describe

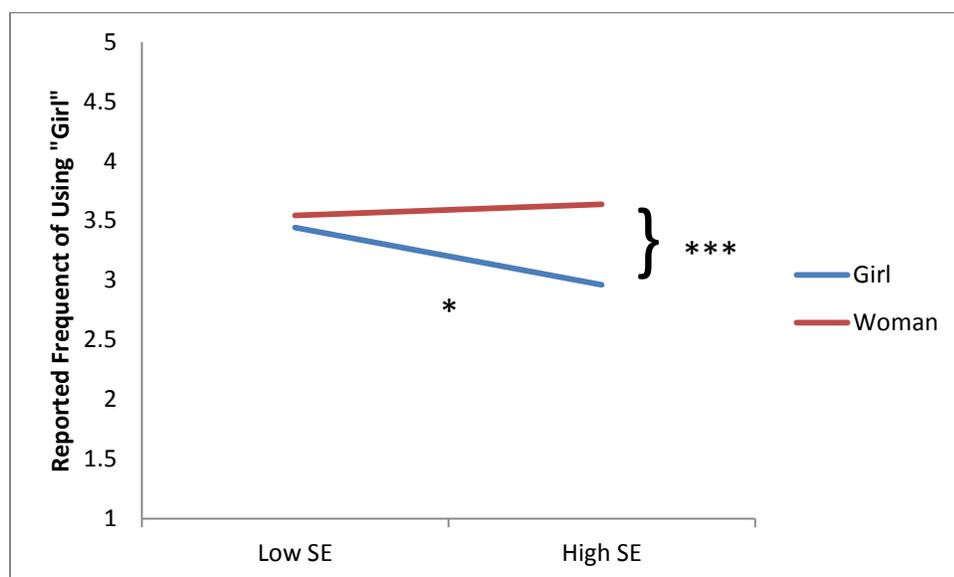others. No other significant main effects or interactions for frequency of using girl were revealed.

Figure 4. Interaction between label manipulation and self-esteem on frequency of using "girl."
*$p < .05$, ***$p < .001$.

Frequency of using "woman." Regression statistics on Step 2 indicated that participants in the girl conditions reported using the term woman significantly more often to refer to female friends and peers than those in the woman conditions, $b = -.27$, $t(252) = -2,52$, $p < .05$. No further significant main effects or interactions were observed for this variable.

Actual job interest. Participant ratings of actual job interest were not significantly affected by the label manipulation; however, a correlation analysis indicated that actual job interest was significantly correlated with interest in future leadership opportunities $r(252) = .27$, $p < .01$. This finding suggests that people came into the study with a certain degree of interest in leadership, and that label manipulations were not strong enough to override these individual differences in interest.

Chapter 4. DISCUSSION

With this Masters project, I sought to examine the effects of infantilizing language on women's self-perceptions and interest in future leadership opportunities. Specifically, I hypothesized that being referred to by the infantilizing label girl would negatively impact women's feelings of leadership and professionalism when applying for a (mock) leadership position, as well as decrease their interest in participating in future studies in a leadership role. Results show support for the hypothesis that infantilizing labels negatively impact women's self-perceptions; however, the label manipulation did not significantly affect behavioral intentions or performance variables.

Main Effect of Label on Participant Perceptions

Regression analyses revealed that the label manipulation significantly influenced the self-perceptions of participants, with those labelled as girls feeling less professional (and other characteristics associated with leadership), less confident that they would obtain the leadership position, and less confident that evaluators reading their profile would see qualities of leadership and professionalism in them. The professor's words made a strong impression on participants, with the label manipulation at times having even larger effects than self-esteem (e.g., for feelings of leadership and others' perceptions of leadership). Importantly, it should be noted that although the effect of the label manipulation was strong, it was also subtle: participants in the girl conditions liked their profiles overall (the mean liking score for participants in the girl conditions was 4.81 on a 7-point Likert scale; this was closest to "liked slightly" on the response scale), and indicated that they felt only mildly irritated and offended by the girl label (the mean irritation score was 2.35 on a 5-point scale; this was closest to "mildly irritating/offensive" on the response scale). Because the girl/woman manipulation was carried out in the context of a profile that was

relatively favorable overall, "girl" did not signal overt derogation for participants, but rather functioned as a subtle indicator of how mature, professional, and confident their evaluators perceived them to be.

The findings of this study speak to the importance of context in determining the effect that language will have on particular individuals. As discussed in the introduction to this paper, Lipton and Hershaft (1984) did not find significant results in their study investigating how paintings were evaluated when artists were labeled girls versus women, nor did I find significant results in a similar study that I conducted in a non-leadership setting. This suggests that the infantilizing label girl has the most impact and is most harmful in contexts where qualities of maturity, leadership, and adultness are most critical, such as in workplace and leadership settings. I do not mean to suggest, however, that being called girl in a non-workplace setting is unproblematic, as it is possible that such labels could have a cumulative effect over time in any setting: the connotations of naivety and innocence may take their toll on women in the long-term.

Gender of Labeler as a Moderator

Perhaps the most surprising finding of the present study was that gender of labeler did not significantly moderate the effect of the label manipulation. The lack of a significant finding does not necessarily indicate, however, that gender of labeler is an unimportant variable in this phenomenon. The fact that nearly 20% of my sample could not correctly recall the gender of the professor who generated their profile suggests that gender was not made salient enough during the experiment. Indeed, perhaps in learning relevant information about the person generating their profile, participants simply saw their evaluator as a high-status professor, and did not internalize information above and beyond those cues that were most important to the situation at hand. Alternatively, perhaps gender is just generally less relevant in computer-mediated

communication, whereas real-life interactions with female and male evaluators would have brought about the hypothesized moderation effect. Another possibility is that the label manipulation was strong enough in this case to have a similar effect on everyone, regardless of labeler gender, but that we would see the moderation effect emerge in a study with a weaker label manipulation.

Self-Esteem as a Moderator

The present study similarly found that self-esteem did not function as a moderator of label effects. Exploratory analyses revealed that while being called girl led high and low self-esteem participants to feel similarly about themselves as a result of their evaluation, those with high self-esteem were less inclined to see this evaluation as accurate. In other words, self-esteem did not affect how participants felt about the feedback they received, but may have made them less likely to internalize it. This interpretation is also supported by the fact that participants with high self-esteem were more responsive than those with low self-esteem to the label manipulation in their reports of how often they themselves use the term girl to refer to female friends and peers. Indeed, this finding seems to suggest that participants with high self-esteem found the girl label more aversive than those with low self-esteem, leading them to deny that they use this label to describe others. As with gender of labeler, it is possible that self-esteem would play a greater moderating role in contexts that are less relevant to leadership, or in experiments with a weaker label manipulation. Because the label manipulation had such a strong effect in this study, it may be that participants were affected equally regardless of self-esteem, but that more subtle manipulations in less relevant contexts would result in it having a stronger moderating influence.

Interest in Future Leadership Opportunities

Although it was hypothesized that the label manipulation would influence participants' interest in participating in future studies as leaders, no significant effects on this variable were found. Indeed, it seems that although the label manipulation was strong enough to affect participants' momentary perceptions and feelings, it was not strong enough to override the leadership-relevant experiences, inclinations, and preferences that participants came into the study with. This interpretation is supported by the fact that interest in actually applying for the student leadership position was significantly correlated with interest in participating in future studies as a leader. Thus, it seems that individual differences in overall leadership interest were strong enough to override the effects of being referred to as a girl, at least in a single instance. It is possible, however, that repeated instances of being called girl could negatively impact self-perceptions to the point of changing women's interest in future leadership opportunities.

Working Memory

Results also indicated that the label manipulation did not significantly influence working memory performance. This lack of effect is likely explained by the fact that participants did not (explicitly) experience their feedback as aversive overall, and thus were not worrying or devoting cognitive resources away from the Operation Span task. Furthermore, it is possible that the working memory task was not the most appropriate measure of performance in this case because it was not directly relevant to the domain being studied (leadership) or the threat experienced by participants (infantilization). Indeed, most stereotype threat research tests women on stereotype-relevant tasks (e.g., math); therefore, future research could investigate whether those who are referred to as girls perform worse on leadership-relevant tasks than those who are called women. There is ample evidence to support this prediction: a number of studies show that decreased

expectancies of success (as was the case in the present study, with participants who were labelled girls reporting less confidence that they would be chosen for the leadership position) results in worse performance on a variety of tasks related to the negative expectancy (e.g., Marshall & Brown, 2004; Baumeister, Hamilton, & Tice, 1985; Crandall & McGhee, 1968). Another possibility is that working memory was not affected by the label manipulation simply because the task used to measure working memory was not reliable (due to using only one block of trials in the analyses). This possibility should be taken into consideration for future research.

Limitations and Directions for Future Research

Given that the sample of the present research was limited to female undergraduate students, this study was not able to address whether or not the results would generalize to an older population of women. Would the same effects occur in women from age 30 to 40, or even older? Future research is needed to examine at what age women stop being referred to as girl, and at what age, if any, the effects of being called girl diminish or change. Because my sample consisted of predominantly White participants, it is also not clear how the results generalize to women from other racial and ethnic backgrounds. It would also be interesting to know who is most likely to be called girl: are there racial, ethnic, or age differences? Is this label ever invoked to reduce the threat posed by high-performing women?

Furthermore, although my study points to some detrimental effects of being called a girl in the workplace, the fact that it took place in a laboratory makes it unclear what the consequences might be in a more realistic workplace setting. Most instances of girl labeling do not occur in work profile questionnaires, meaning future research should investigate more ecologically valid ways of testing these effects. Along the same lines, it would be useful to know

how often women are referred to as girls in various workplace settings, since current data only indicates frequencies in everyday (home) life.

As mentioned previously, one limitation of the present study involved the fact that the gender of the professor giving participants feedback was not sufficiently noticeable to participants. Future research will need to identify a way to increase the salience of gender, perhaps by having participant/evaluator interactions occur face-to-face, rather than through email. Similarly, future studies could examine if and when gender of labeler becomes a significant moderator of labelling effects. Finally, further studies could examine the boundary conditions of these effects: what would happen if the label had been used by a graduate student or peer evaluator instead of a high-status professor? Would the results have changed if participants believed their profile was going to be forwarded to a committee of their peers rather than community leaders?

Implications

Overall, the results of the present study highlight the fact that that common, everyday labels used to describe women can have harmful effects on those who are labeled, particularly in a work/leadership context. Indeed, this study may be the first to show that a commonly used label for a group of people (and one that is even preferred by members of that group in many contexts) can have a detrimental effect on members of that group. Previous research (e.g., Boeckmann & Liew, 2002; Carnaghi & Maass, 2007; Evans & Chapman, 2014; Leets, 2002; Leets & Giles, 1997) has documented the effects of hate speech and overtly derogatory labels on minority group members. The term girl reveals another insidious type of language effect that passes by relatively unnoticed and is deemed "normal," yet has deleterious consequences. Only five participants in this study spontaneously mentioned that girl was a problematic aspect of the

profile they received, revealing that women are referred to as girls so often that this infantilization has become invisible: the label is no longer noticed as inaccurate, inappropriate, or strange. The routine nature of women's infantilization in language, however, should not be taken as a sign of innocuousness: research shows that subtle acts of sexism can have a cumulative effect that builds over time (e.g., Berg, 2006; Swim, Hayes, Cohen, & Ferguson, 2001) and that in some cases, can be even more damaging than more overtly sexist behavior precisely *because* subtle acts are not recognized as sexist and therefore go unchallenged (e.g., Nguyen & Ryan, 2008; Barreto & Ellemers, 2005; Dardenne, Dumont, & Bollier, 2007; Swim, Mallett, & Strangor, 2004).

For these reasons, employers, co-workers, professors, and teachers should be particularly wary about using the term girl to describe women: it could have a detrimental impact on their feelings of leadership, as well as their thoughts about how others in their environment perceive them. Shifting language norms in accordance with information about their potential negative impacts is not only important and necessary, but also achievable. Such changes in "politically correct" terminology have occurred many times in the past, as is evidenced by the fact that "retarded" is no longer a common term used to describe people with mental disabilities (Holbrook, 1995). Because the proposed study focuses on documenting the consequences of the problem rather than testing a potential solution, it is important that further research be undertaken to investigate ways of reducing or eliminating people's use of infantilizing terms for women, particularly in leadership and workplace settings.

Conclusion

For my Masters research, I investigated the effect on women of being called girl, an infantilizing label. In this thesis, I have outlined the importance of language in influencing

perceptions, discussed research indicating that women are frequently referred to as girls in everyday conversation, and outlined the methodology and results of my study, which experimentally manipulated the label given to participants in fake feedback on a leadership profile questionnaire. Results indicated that infantilizing labels significantly impact women's self-perceptions, but not interest in future leadership opportunities. These findings suggest a need to change our current language practices and to avoid the use of infantilizing terms when referring to women.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Alford, D. K. (2011). The demise of the Whorf hypothesis (A major revision in the history of linguistics). *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 4,* 485-489.

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. doi: 10.1002/ejsp.270

Baumeister, R. F., Hamilton, J. C., & Tice, D. M. (1985). Public versus private expectancy of success: Confidence booster or performance pressure?. *Journal of Personality and Social Psychology*, *48*, 1447-1457.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*, 3186-3191. doi: 00007632-200012150-00014

Bebout, L. (1984). Asymmetries in male-female word pairs. *American Speech, 59,* 13-30.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256-276. doi: 10.1037/0096-3445.136.2.256

Berg, S. H. (2006). Everyday sexism and Posttraumatic Stress Disorder in women: A correlational study. *Violence Against Women*, *12*, 970-988. doi: 10.1177/1077801206293082

Berry, J. W. (1980). On cross-cultural comparability. *International Journal of Psychology*, *4*,

207-29. doi: 10.1080/00207596908247261

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics*, *66*, 35-44. doi: 10.1016/j.pragma.2014.02.009

Blaine, B., & Crocker, J. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. In R. F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (55-85). New York, NY: Plenum Press. doi:10.1007/978-1-4684-8956-9_4

Boeckmann, R. J., & Liew, J. (2002). Hate speech: Asian American students' justice judgments and psychological responses. *Journal of Social Issues*, *58*, 363-381. doi: 10.1111/1540-4560.00265

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. doi: 10.1007/BF00290069

Brislin, R. W. (1970). Back translation for cross-cultural research. *Journal of Cross Cultural Psychology*, *1*, 185-216. doi: 10.1177/135910457000100301

Bruckmüller, S., & Abele, A. E. (2010). Comparison focus in intergroup comparisons: Who we compare to whom influences who we see as powerful and agentic. *Personality and Social Psychology Bulletin*, *36*, 1424-1435. doi: 10.1177/0146167210383581

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. doi: 10.1002/ejsp.858

Carli, L. L. (1999). Gender, interpersonal power, and social influence. *Journal of Social Issues,*

55, 81–99. doi: 10.1111/0022-4537.00106

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, *57*, 725–741. doi: 10.1111/0022-4537.00238

Carnaghi, A., & Maass, A. (2007). In-group and out-group perspectives in the use of derogatory group labels gay versus fag. *Journal of Language and Social Psychology*, *26*, 142-156. doi: 10.1177/0261927X07300077

Crandall, V. C., & McGhee, P. E. (1968). Expectancy of reinforcement and academic competence. *Journal of Personality*, *36*, 635-648. doi: 10.1111/1467-6494.ep8933301

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

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. doi: 10.1037/0022-3514.93.5.764

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of personality and social psychology*, *88*, 276. doi: 10.1037/0022-3514.88.2.276

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615-1628. doi: 10.1177/014616702237644

Dayhoff, S. A. (1983). Sexist language and person perception: Evaluation of candidates from newspaper articles. *Sex Roles*, *9*, 527-539. doi: 10.1007/BF00289792

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders.

Psychological Review, *109*, 573-598. doi: 10.1037//0033-295X.109.3.573

Ely, G. E., & Dulmus, C. N. (2010). Disparities in access to reproductive health options for female adolescents. Social work in public health, 25, 341-351. doi: 10.1080/19371910903240779

Ehrlich, S. (2001). *Representing rape: Language and sexual consent*. New York, NY: Routledge.

Evans, C. B., & Chapman, M. V. (2014). Bullied youth: The impact of bullying through lesbian, gay, and bisexual name calling. *American journal of orthopsychiatry*, *84*, 644-652. doi: 10.1037/ort0000031

Felski, R. (1989). *Beyond feminist aesthetics: Feminist literature and social change*. Cambridge, MA: Harvard University Press.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 1-11. doi: 10.3758/s13421-014-0461-7

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, *23*, 629-643. doi: 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. doi: 10.1037/a0022530

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

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and social psychology bulletin*, *28*, 659-670. doi: 10.1177/0146167202288010

Henley, N. M., Gruber, B., & Lerner, L. (1988). *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, October 1984.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Holbrook, M. B. (1995). The three faces of elitism: Postmodernism, political correctness, and popular culture. *Journal of Macromarketing, 15*, 128-165. doi: 10.1177/027614679501500209

Jung Typology Test. (n.d.). Retrieved March 25, 2014, from http://www.humanmetrics.com/cgi-win/jtypes2.asp

Kane, M. J., Hambrick, D. Z., Tuholski, S. W.,Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology*, *133*, 189–217. doi: 10.1037/0096-3445.133.2.189

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

Lambdin, J. R., Greer, K. M., Jibotian, K. S., Wood, K. R., & Hamilton, M. C. (2003). The animal= male hypothesis: Children's and adults' beliefs about the sex of non–sex-specific stuffed animals. *Sex Roles*, *48*, 471-482. doi: 10.1023/A:1023567010708

Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of social issues*, *58*, 341-361. doi: 10.1111/1540-4560.00264

Leets, L., & Giles, H. (1997). Words as weapons—when do they wound? Investigations of

    harmful speech. *Human Communication Research*, *24*, 260-301. doi: 10.1111/j.1468-

    2958.1997.tb00415.x

Lerner, H. E. (1976). Girls, ladies, or women? The unconscious dynamics of language choice.

    Comprehensive Psychiatry, *17*, 295-299. doi:10.1016/0010-440X(76)90003-1

Lipton, J. P., & Hershaft, A. M. (1984). "Girl,""woman,""guy,""man": The effects of sexist

    labeling. *Sex Roles*, *10*, 183-194. doi: 10.1007/BF00287773

London, B., Downey, G., Romero-Canyas, R., Rattan, A., & Tyson, D. (2012). Gender-based

    rejection sensitivity and academic self-silencing in women. *Journal of personality and*

    social psychology, *102*, 961-979. doi: 10.1037/a0026615

Marshall, M. A., & Brown, J. D. (2004). Expectations and realizations: The role of expectancies

    in achievement settings. *Motivation and Emotion*, *28*, 347-361. doi: 10.1007/s11031-004-

    2388-y

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. doi: 10.3758/BF03195410

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. doi:

    10.1177/089124393007001007

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. doi: 10.1037/a0012702

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York, NY: Basic Books.

Pratto, F., Korchmaros, J. D., & Hegarty, P. (2007). When race and gender go without saying. *Social Cognition*, *25*, 221-247. doi: 10.1521/soco.2007.25.2.221

Rahman, J. (2012). The N Word: Its history and use in the African American community. *Journal of English Linguistics*, *40*, 137-171. doi: 10.1177/0075424211414807

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and social psychology bulletin*, *27*, 151-161. doi: 10.1177/0146167201272002

Romaine, S. (1999). *Communicating Gender*. Mahwah, NJ: L. Erlbaum Associates.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Santa, J. L., & Baker, L. (1975). Linguistic influences on visual memory. *Memory & Cognition*, *3*, 445-450. doi: 10.3758/BF03212939

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194-201. doi:10.1006/jesp.2001.1500

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of personality and social psychology*, *85*, 440-452. doi: 10.1037/0022-3514.85.3.440

Schulz, M. (1975). The semantic derogation of women. In B. Thorne & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (64-75). Rowley, MA: Newbury House.

Shrauger, J. S. (1975). Responses to evaluation as a function of initial self-perceptions.

Psychological Bulletin, 82, 581-596. doi: 10.1037/h0076791

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, *35*, 4-28. doi:10.1006/jesp.1998.1373

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*, 42-61. doi: 10.1037/0033-2909.111.1.42

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. doi: 10.1177/0146167211406434

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

Tabachnick, T. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Old Tappan, New Jersey: Allyn & Bacon.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505. doi:10.3758/BF03192720

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

Appendix A

Rosenberg Self-Esteem Scale
(Rosenberg, 1965)

	STATEMENT	Strongly Agree	Agree	Disagree	Strongly Disagree
1.	I feel that I am a person of worth, at least on an equal plane with others.	○	○	○	○
2.	I feel that I have a number of good qualities..	○	○	○	○
3.	All in all, I am inclined to feel that I am a failure.	○	○	○	○
4.	I am able to do things as well as most other people.	○	○	○	○
5.	I feel I do not have much to be proud of.	○	○	○	○
6.	I take a positive attitude toward myself.	○	○	○	○
7.	On the whole, I am satisfied with myself.	○	○	○	○
8.	I wish I could have more respect for myself.	○	○	○	○
9.	I certainly feel useless at times.	○	○	○	○
10.	At times I think I am no good at all.	○	○	○	○

Appendix B

Job Advertisement

<u>Leadership Position Available on Campus</u>

Penn State University is seeking a mature, independent, and capable student to fill a leadership position at the University Park campus. Duties involve actively supervising other students, as well as taking initiative and responsibility for several important projects being undertaken by the Office of Student Affairs. Because the position involves overseeing others, the successful candidate will be assertive, decisive, and confident, and will possess strong leadership skills.

Appendix C

Filler Questions

In responding to the questions below, it is best to go with your gut instinct rather than over-thinking your answers-- the qualities that indicate leadership may surprise you.

Please indicate your agreement with the following statements from 1 (Strongly disagree) to 5 (Strongly agree):

- I am more interested in a general idea than in the details of its realization.

- I tend to be unbiased even if this might endanger my good relations with people.

- Strict observance of the established rules is likely to prevent a good outcome.

- I feel involved when watching TV soaps.

- It is difficult to get me excited.

- I often think about humankind and its destiny.

- I believe the best decision is one that can be easily changed.

- Objective criticism is always useful in any activity.

- I prefer to act immediately rather than speculate.

- I trust reason rather than feelings.

- I am inclined to rely more on improvisation.

- My actions are frequently influenced by emotions.

- I am a person somewhat reserved and distant in communication.

- I often contemplate the complexity of life.

- I think that almost everything can be analyzed.

- I take pleasure in putting things in order.

- I feel at ease in a crowd.

- I easily understand new theoretical principles.

- The process of searching for a solution is more important to me than the solution itself.

- When solving a problem I would rather follow a familiar approach than seek a new one.

- I prefer meeting in small groups over interaction with lots of people.

- When considering a situation I pay more attention to the current situation and less to a possible sequence of events.

- When solving a problem I consider the rational approach to be the best.

- I feel more comfortable sticking to conventional ways.

- I am easily affected by strong emotions.

- I am always looking for opportunities.

- My desk, workbench, etc. is usually neat and orderly.

- As a rule, current preoccupations worry me more than my future plans.

- I am consistent in my habits.

- I willingly involve myself in matters which engage my sympathies.

- I easily perceive various ways in which events could develop.

Appendix D

Work Profile Attachment

Dear participant,

I read through your answers to the work profile test, and generated this profile for you. In my opinion, you are a very strong example of a **Modern Girl**. Please read your profile carefully so that you will be prepared to answer questions about it in the next portion of the study.

-Dr. Stephen C. Johnson

Leadership Profile

Participant **4851A2**
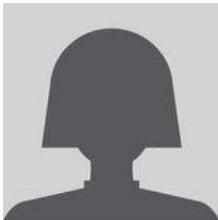
Modern Girl

In the world of work, you are a girl with modern sensibilities. Like many other girls your age, you like to be productive, and always appreciate getting positive feedback from others. As a girl who is interested in the world and the people around her, you also appreciate hearing and incorporating the perspectives of others into your work. While a bad day might get you down, as a modern girl, you continue your efforts, and try to carry on despite it.

Appendix E

Post-Survey Questionnaire

The study is almost complete! We just have a few more questions.

How much did you like the profile you received from Dr. Johnson? (1-7, Extremely disliked-Liked very much)

How accurate do you think it was in describing you? (1-6, Very inaccurate- Very accurate)

If the student leadership position was an actual job being advertised on campus and you were looking for work, how interested would you be in applying? (1-5, Not at all interested-Extremely interested)

Did you hear anything about this study before you participated (from friends, others who participated before you)? If so, what did you hear? (Open-ended)

Do you have any guesses about what the hypothesis of the study might be? If so, please explain. (Open-ended)

Did anything about your work profile, or the study in general, seem unusual? Please explain. (Open-ended)

What was the gender of the professor who generated your profile? (Male, Female, Don't remember)

Which of the following words did your profile use to describe you? (Girl, Woman, Other, Don't remember)

How did you feel about the word used to describe you? (girl, woman, etc.) (Open-ended)

How irritating/offensive did you find it? (1-5, Not at all irritating/offensive- Extremely irritating/offensive)

How often do you use the term "girl" to refer to female friends and peers? (1-5, Never- Always)

How often do you use the term "woman" to refer to female friends and peers? (1-5, Never-Always)

The following questions are used to assess the representativeness of the participants in this study.

What is your gender? (Female, Male, Trans, Other)

Which best describes your race/ethnicity? (Asian, Black or African American, Latina/o or Hispanic, Middle Eastern or Arab, Native American or Alaska Native, Native Hawaiian or Other

Pacific Islander, White or Caucasian (non-Hispanic), I identify with multiple racial/ethnic groups, Other)

What is your age (in years)? (Text-entry)

Are you a native English speaker? (Yes, No)

Are you a U.S. citizen? (Yes, No)