From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2019

Derogatory Gendered Labels: On Twitter and in the Lab

Connor, Rachel

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp014f16c573q

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

DEROGATORY GENDERED LABELS: ON TWITTER AND IN THE LAB

Rachel A. Connor

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF

PSYCHOLOGY

Primary Adviser: Deborah Prentice
Secondary Adviser: Susan Fiske

June 2019

© Copyright by Rachel A. Connor, 2019. All rights reserved.

Abstract

How do overheard derogatory gendered labels (DGLs) affect perceptions of targets and labelers? In this paper, I examined the content, targeting, and evaluative consequences of DGLs in both an experimental and a real-world context. In Part I, I conducted four experiments examining the impact of DGLs on perceptions of men and women varying in emotion stereotypicality. In Part II, using data from Twitter, I investigated DGLs directed at male and female moderators of the 2016 presidential primary debates. In Part I, Studies 1 (N = 243) and 2 (N = 235) showed that, when described with a female-derogatory label, emotion was attributed internally to a female target, which then negatively affected evaluations of her. Further, participants positively evaluated labelers of counterstereotypical targets. In contrast to expectations, in Study 3 (N = 247), gender traditionalists defended a labeled male target, attributing his reaction less to internal factors and providing more positive evaluations of him. Across studies, findings depended on participant gender and gender ideology. Overall, findings support that DGLs are interpreted to fit with pre-existing expectations. In Part II, Study 5 found that people tweeting about Republican debates used DGLs and more abstract language for female stereotype violators. Studies 6 and 7 found that men high in hostile sexism responded to negative tweets, whether or not they contained a DGL, by derogating female targets and rewarding tweet authors. Overall, this study supports that DGLs play a role in stereotype transmission and maintenance. Perceivers who endorse sexism latch onto DGLs that support stereotype-based explanations for women's behavior. Further, internal attributions are a crucial mechanism by which negative labels affect target evaluations.

Keywords: sexism, gender stereotypes, labeling, attribution, social media

Table of Contents

Abstract iii
Acknowledgements vi
Introduction 1
Labeling 2
Labeling and Gender 5
Third-Party Effects of Labels 7
Psychological Mechanisms of Labeling Effects 9
PART I: DEROGATORY LABELS FOR EMOTIONALITY 12
Studies 1-4 Overview 12
STUDY 1 14
Study 1 Method 15
Study 1 Results 18
Study 1 Summary 23
STUDY 2 24
Study 2 Method 25
Study 2 Results 25
Study 2 Summary 28
STUDY 3 29
Study 3 Method 30
Study 3 Results 31
Study 3 Summary 33
STUDY 4 34
Study 4 Method 35
Study 4 Results 36
Study 4 Summary 37
PART I: DISCUSSION AND CONCLUSION 38
PART II: DEROGATORY GENDERED LABELS ON TWITTER 42
Studies 5-7 Overview 43
Creating a Slur Dictionary 45
STUDY 5 50
Labeling Effects on Twitter 50
Study 5 Method 52
Study 5 Results 58

Study 5 Discussion 64
STUDY 6 68
Study 6 Method 68
Study 6 Results 69
Study 6 Summary 70
STUDY 7 71
Study 7 Method 71
Study 7 Results 72
Study 7 Summary 72
GENERAL DISCUSSION 73
Role of Sexism 74
Usage of DGLs 75
Internal Attributions 75
Confirmation Bias – A Better Framework for Understanding DGLs? 76
Future Directions 77
Conclusion 78

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance. First, I am deeply indebted to the formal and informal mentors who have supported me along the way, without you, I wouldn't be here today:

Lara Ault – It was your enthusiasm for social psychological research that first inspired me along this path. Thank you for your guidance and encouraging me to apply to graduate school!

Debbie Prentice – When I applied to graduate school, I never expected to get into Princeton. Thank you for recognizing my potential and giving me this opportunity. Also thank you for always being able to cut to the essence of a problem and see the bigger picture. I am very fortunate to have had you as a mentor!

Susan Fiske – Thank you for taking me on as a secondary advisee and welcoming me in your lab. I am grateful for all of the opportunities you have connected me to and encouraged me to pursue. I feel very privileged for having been able to work with you!

Betsy Paluck – It was a joy to teach with you. Thank you for being a role model and for your willingness to offer support and career advice.

Thank you to the Fiske Lab and Alin Coman for your valuable feedback on my dissertation and other projects! I also must thank the National Science Foundation, the Joint Degree Program in Social Policy, and the Graduate School for financial support.

To my colleagues and friends from the department, especially Cydney, Jeremy, Kyonne, Kaite, and my partner in wine-related crime, Tasha – Thank you for the good times and good conversations. Also for commiserating with me during the toughest times.

My fur babies Zoe, Zartan, and Lexena – Thank you for reminding me to take breaks and loving me unconditionally. Finally, to my long-time partner and now husband, Joshua, who has been with me through every step of this difficult journey, always providing support and encouragement.

Introduction

Gender figured prominently in the 2016 presidential election, not only because candidate Hillary Clinton was the first woman to clinch a major party nomination, but also because of candidate Donald Trump's hypermasculine posturing. Trump engaged in blatantly sexist rhetoric, branding women who challenged him 'bimbos' and 'nasty women' (Diaz, 2016; realDonaldTrump, 2015, 2016). Moreover, he mocked and feminized male opponents by describing them as 'soft weak little babies' and 'pussies' (Santucci & Hopper, 2016; Strauss, 2016). The gendered rhetoric deployed by the Trump campaign positioned femininity and alternative masculinity as deviant and inferior, thus working to reinforce a hegemonic view of masculinity that only 'real men' are fit for positions of power.

Trump's hateful rhetoric conveyed more than his personal prejudice; it also communicated the social acceptability of such prejudice, thus emboldening others to express derogatory, biased language toward stigmatized groups (Anti-Defamation League, 2016; Miller & Werner-Winslow, 2016). Following the election, hate crimes in the U.S. spiked; a number of these incidents involved hateful slurs and could be linked to support for Trump (Miller & Werner-Winslow, 2016). Trump's success normalized hate speech, leading people to view prejudiced speech about groups targeted by Trump as more socially acceptable (Crandall & White, 2016). This demonstrates the power of labels to transmit and maintain culturally-shared beliefs. As Greenberg & Pyszczynski (1985) state, "noxious labels for out-groupers…are not merely symptoms of prejudice but carriers of the disease as well" (p. 70).

The present studies will contribute to the understanding of the effects of derogatory gendered labels on third-party perceptions of female and male targets, in both an experimental and a real-world context. Using an experimental approach, the first set of studies will explore the effects of derogatory gendered labels applied to female and male targets displaying emotions that vary in gender stereotypicality. The second set of studies will investigate the occurrence and spread of derogatory gendered labels on Twitter. I expect that derogatory gendered labels will enhance negative evaluations of women and non-masculine men, particularly among participants who strongly endorse traditional gender ideology.

Labeling

Defining Labels

In his seminal work, 'The Nature of Prejudice,' Allport (1979) argued that labels conjure negative stereotypes and "force [a target] instantly into a rejective category…when these labels are employed we can be almost certain that the speaker *intends* not only to characterize the person's membership, but also to disparage and reject [them]" (p. 181). Ample psychological research has borne out Allport's argument for the importance of labels in prejudice. Labels support the construction of social categories; they demarcate category boundaries, enhancing perceptions of intercategory difference and intracategory similarity (Lupyan, 2012; Waxman & Gelman, 2009; Waxman & Lidz, 2006). Further, they guide inductive reasoning about targets, leading to category-based inferences about underlying properties (Heyman & Gelman, 2000; Keates & Graham, 2008; Waxman, 2010). Finally, an impressive body of research supports that social

category labels evoke stereotypical associations, which influence perceptions of targets (Maass & Arcuri, 1996; Macrae & Bodenhausen, 2000).

However, as Allport notes, in addition to generic group labels (e.g., 'woman'), people use derogatory group labels (e.g., 'bitch'). Unlike generic labels, which are often used simply to denote group membership, derogatory group labels are deployed to degrade individuals and their groups. Although perhaps the quintessential expressions of prejudice, only certain types of derogatory labels have been studied. There has been some interest in ethnic slurs (Greenberg & Pyszczynski, 1985; Kirkland, Greenberg, & Pyszczynski, 1987; Leader, Mullen, & Rice, 2009; Simon & Greenberg, 1996), but much of the most recent work has focused on homophobic slurs (Carnaghi, Maass, & Fasoli, 2011; Carnaghi & Maass, 2008; Fasoli, Maass, & Carnaghi, 2015; Fasoli et al., 2016; Goodman, Schell, Alexander, & Eidelman, 2008; Saucier, Till, Miller, O'Dea, & Andres, 2015). Most pertinently, there are only four studies that I am aware of that focus on derogatory gendered labels (Fasoli, Carnaghi, & Paladino, 2015; James, 1998; Preston & Stanley, 1987; Saucier et al., 2015).[1]

In order to understand the broader phenomenon, it is important that we extend the study of derogatory labels to other types of labels. Derogatory gendered labels (DGLs) differ from racial/ethnic slurs in that they do not disparage an entire group; rather, they target subgroups of men and women on the basis of gender-inappropriate behaviors (Ashwell, 2016). Thus, relative to other types of derogatory group labels, DGLs may differ in function and impact. Further, although several studies show that derogatory

[1] Of course, homophobic slurs can also be considered derogatory gendered labels. The present study will extend its focus beyond male-targeted homophobic slurs (the focus of the majority of research in this area) to include a broader range of gender-linked labels for women and men.

group labels depress evaluations of targets (e.g., Greenberg & Pyszczynski, 1985; Goodman et al., 2008), the mechanisms by which they do so are not well understood. Finally, previous studies have not considered how derogatory group labels affect perceptions of the person using the label, which are likely to affect whether listeners accept or contest it. Accordingly, the present studies consider the effects of derogatory gendered labels (DGLs) on men and women, with the goal of adding to the understanding of how derogatory group labels influence social perception and interaction.

Defining Derogatory Group Labels

Previous research has studied derogatory labels in the form of nouns (e.g., 'nigger,' Greenberg & Pyszczynski, 1985; 'fag' Carnaghi & Maass, 2008). Social category nouns refer to *kinds*, whereas adjectives refer to specific *properties* of kinds (Carnaghi et al., 2008; Rothbart & Taylor, 1992). Carnaghi et al (2008) argue that this distinction, among others, contributes to the greater significance of nouns than adjectives in impression formation. They find that, relative to adjectives, nouns more strongly imply stable, underlying characteristics of categories, and they facilitate more stereotypical inferences.

Still, decades of research on personality (e.g., McCrae & Costa, 2008; John, Angleitner, & Ostendorf, 1988) and social perception (e.g., Jones & Nisbett, 1971; Uleman, Saribay, & Gonzalez, 2008) demonstrate the importance of trait descriptors in thinking about and describing the self and others. People, at least those in western nations, commonly rely on adjectives to describe others (Maass, Karasawa, Politi, & Suga, 2006), and they tend to think of others in dispositional terms (Gilbert & Malone, 1995; Jones & Nisbett, 1971). Further, Semin & Fiedler's Linguistic Category Model

(1991), which categorizes word types according to abstractness (i.e., the degree to which they imply stable, dispositional causes for behavior), does not distinguish between adjectives and nouns, both of which are classified at the highest level of abstractness. A number of gendered insults exist in both nominal and adjectival forms, which are often used in similar ways (e.g., 'she is a bitch', 'she is bitchy'). The same case is true of gendered mental illness labels (e.g., 'she is a hysteric', 'she is hysterical'). From a labeling theory perspective, it makes little sense to distinguish among these uses, both of which attribute deviance to a person on basis of underlying property or characteristic. Thus, in our view, gendered labels include nouns or adjectives that may or may not explicitly refer to gender, but due to their association with gender, evoke gender stereotypes (Mendelberg, 2008; Winter, 2008). Accordingly, in the present studies, I define derogatory gendered labels as nouns *or* adjectives that refer to person properties and convey negative attitudes about a gender group.

Labeling and Gender

Feminist scholars have observed the ways in which a male perspective is encoded in language (e.g., Ehrlich & King, 1994). For example, androcentric language (e.g., the use of masculine generics such as 'he' or 'mankind') treats men as the default category (Bailey & LaFrance, 2017), and women as marked. Further, more negative labels exist for women than for men, which mark women's lower status relative to men (Schur, 1983; Sutton, 2001). However, language does not only disparage women, but upholds a certain form of dominant masculinity that is crucial to maintaining the gender system. This dominant male perspective is reflected in the preponderance of gendered labels that

reinforce gender stereotypes and discourage counterstereotypical behavior in both women and men (e.g., James, 1998).

Gender stereotypes produce certain status-linked expectations for behavior (how men and women *should* and *should not* behave; Burgess & Borgida, 1999). Women are expected to display feminine-typed characteristics such as warmth, sensitivity to others, emotional expressivity, and cooperation; men are expected to display high-status, masculine-typed characteristics such as independence, assertiveness, rationality, and ambitiousness (Prentice & Carranza, 2002; Rudman, Moss-Racusin, Glick, & Phelan, 2012). Gender stereotypes also proscribe or prohibit counterstereotypical displays; they discourage traditionally masculine traits and emotions (e.g., arrogance, dominance, anger) in women and traditionally feminine traits and emotions (e.g., emotionality, weakness, sadness) in men (Brescoll & Uhlmann, 2008; Prentice & Carranza, 2002; Tiedens, 2001).

Female-derogatory labels frequently target women who violate gender expectations, whether regarding sexuality (e.g., 'whore', 'cockteaser', 'dyke'), appearance (e.g., 'dog', 'heifer'), emotionality (e.g., 'crazy', 'hysterical'), or passivity (e.g., 'bitch', 'ballbuster,' James, 1998; Ussher, 1992). Derogatory terms such as 'fag,' 'coward,' and 'sissy' target men who engage in counterstereotypical behavior (e.g., showing weakness or expressing femininity; Pascoe, 2005; Saucier et al., 2015). Accordingly, derogatory gendered labels frequently stigmatize non-stereotype conforming individuals, thus enforcing dominant norms (Schur, 1983).

Importantly, however, whereas both men and women are punished for transgressing gender boundaries, women are also stigmatized for being women (Schur,

1983). Gender stereotypes set up a double bind for women in which they are negatively evaluated for behavior that is perceived as too masculine as well as behavior perceived as too feminine (Connor & Fiske, 2017). Dominant culture places high value on stereotypically male characteristics (e.g., competence, aggressiveness, independence); feminine characteristics (e.g., emotionality, warmth, sensitivity) are comparatively less valued (e.g., Rudman et al., 2012). Stereotypically feminine women, while considered likable, are not always respected (and may be affixed with the labels, 'ditz' or 'bimbo'). Even feminine-typed labels that seem positive (e.g., 'babe', 'honey') can ultimately reinforce women's lower status (Glick & Fiske, 1996). Although men are evaluated harshly when they behave counterstereotypically, they do not face the same penalties for stereotypically masculine behavior. Indeed, it is possible that negative, stereotypical labels for men (e.g., 'dick') could increase men's status, rather than detract from it. In sum, the gender system promotes DGLs that deny societal value to women and non-dominant men (Connell & Messerschmidt, 2005; Schur, 1983).

Third-Party Effects of Labels

Language transmits and maintains bias (Maass & Arcuri, 1996; Van Dijk, 1987). Research has shown that stereotype-related words facilitate the use of stereotypes in subsequent judgments about targets (Devine, 1989; Macrae & Bodenhausen, 2000). Generalized statements about gender (i.e., generics; "boys play with trucks") support the development of essentialist and stereotypic beliefs about gender (Cimpian & Markman, 2011; Cimpian, 2010; Gelman, Taylor, Nguyen, Leaper, & Bigler, 2004; Rhodes, Leslie, & Tworek, 2012). People use more abstract language to describe stereotypical compared to counterstereotypical actions, which leads receivers to infer dispositional causes for

stereotypical behavior (Wigboldus, Semin, & Spears, 2000). For instance, people might use abstract language to describe a woman expressing emotional distress (e.g., "she is emotional"), whereas they might use concrete language to describe the same behavior performed by a male target (e.g., "he is crying"). In other words, stereotypical behavior is often depicted as a characteristic of the target in general, while counstereotypical behavior is characterized as a specific event within a particular context. Language does not only convey the biases and assumptions of the speaker, but it transmits bias to receivers, ultimately reinforcing stereotypes in the wider society (Van Dijk, 1987).

Relative to more subtle forms of linguistic bias, derogatory group labels are particularly "potent communicative devices" (Greenberg, Kirkland, & Pyszczynski, 1988, p. 77). More than just words, derogatory group labels carry symbolic meaning that colors evaluations of individual targets. In a study conducted by Greenberg & Pyszczynski (1985), white participants who overheard a white confederate apply a derogatory ethnic label ('nigger') to a black target rated the target as less skilled in a debate contest (when he lost). Similarly, Goodman et al. (2008) found that participants who overheard a derogatory comment about a gay male leader ('he is so gay!') rated him as lower in leadership ability and displayed more negative nonverbal behavior towards him. Finally, a series of studies conducted by Carnaghi and colleagues found that, compared to generic category labels, exposure to homophobic slurs lead to prejudice and discrimination toward gay men (Carnaghi & Maass, 2008; Fasoli, Maass, et al., 2015; Fasoli et al., 2016). In sum, derogatory group labels can induce in listeners stereotypical thinking and negative affective reactions toward targets.

Psychological Mechanisms of Labeling Effects

Previous research suggests at least three possible mechanisms by which derogatory labels influence impressions of targets: they activate stereotypes, they justify backlash, and they release suppressed prejudice. These mechanisms are not mutually exclusive and likely often operate together. Accordingly, I look for evidence for each.

Although stereotypes often influence impression formation, they are not uniformly applied to judgments across all situations. Whether categorization leads to stereotype activation and application depends on a number of contextual factors, including the presence of category-relevant cues (Blair, 2002; Deaux & Major, 1987). Social category cues facilitate categorization and stereotype activation, leading to stereotypic biases. For example, making gender information salient leads participants to recall lower math SAT scores for a female target, relative to when her ethnic category (Asian) is made salient (Pittinsky, Shih, & Ambady, 2000). Similarly, derogatory gendered labels should cue stereotypical representations, thus producing stereotype-consistent judgments (e.g., Greenberg & Pyszczynski, 1985). Furthermore, these effects may occur independent of actual target behavior, at least when the behavior can be interpreted as consistent with the activated stereotype. In other words, the label affixed to a target—rather the target's behavior itself—activates a set of causal and descriptive beliefs which can result in stigmatized treatment of the target (Link, Cullen, Frank, & Wozniak, 1987; Schur, 1983). Therefore, in the present study, it is possible that a female-derogatory label will lead to more stereotypical ratings of a female target, whether her behavior is consistent with stereotypes or not.

H1. Labels activate stereotypes. The *stereotype activation* hypothesis is that derogatory gendered labels activate stereotypes, which then guide perceptions of the target and attributions for their behavior in line with stereotypes, regardless of whether the target exhibits stereotype-confirming or disconfirming behavior.

Target characteristics also affect the likelihood that stereotypes will be activated and applied to judgments (Deaux & Major, 1987; Fiske & Neuberg, 1990). In the impression formation process, target attributes are evaluated against the initial activated category. When a target fits stereotypical expectations, perceivers rely on stereotypes to guide judgments (Jones & Kaplan, 2003; Maass & Borini, 2005; Stangor & McMillan, 1992). When a target does not readily fit the category (e.g., she behaves in a counterstereotypical manner), perceivers undergo an additional information search. To do this, they activate a subcategory or individuate the target, as a way of resolving initial expectancies with the observed behavior (Fiske & Neuberg, 1990). As a result, impressions of counterstereotypical targets are typically more individuated (i.e., less stereotyped) relative to when targets match expectations.

Thus, people who disconfirm stereotypes are able to avoid being stereotyped, but by violating stereotypes, they are often subjected to social and economic penalties known as backlash. For example, non-normative women (e.g., agentic women) and men (e.g., modest men) encounter social disapproval, sabotage, and discrimination in hiring and promotion (Heilman, Wallen, Fuchs, & Tamkins, 2004; Moss-Racusin, Phelan, & Rudman, 2010; Phelan, Moss-Racusin, & Rudman, 2008; Rudman & Fairchild, 2004; Rudman & Glick, 1999; Rudman & Mescher, 2013). These individuals challenge the existing gender hierarchy, which people are motivated to defend (Jost, Banaji, & Nosek,

2004; Rudman et al., 2012). Backlash thus plays a role in maintaining stereotypes by discouraging counterstereotypical behavior.

Importantly, people are unlikely to enact backlash unless they can find a justification for doing so (Crandall & Eshleman, 2003; Rudman et al., 2012). Perceivers rely on a number of different types of justifications to excuse prejudiced responses, including attributions (Crandall & Eshleman, 2003). By attributing counterstereotypical behavior to a target's disposition, perceivers validate their negative reactions, thereby releasing backlash and avoiding challenges to the status quo. Accordingly, H2 suggests that the effects of the label could depend on the stereotypicality of target, such that it will facilitate internal attributions and depress evaluations for counterstereotypical targets only.

H2. Labels justify backlash. The *punishing deviants* hypothesis is that derogatory labels provide justification to enact backlash (in the form of negative evaluations). Further, since perceivers are motivated to punish people who violate stereotypes, they may also respond positively to people who enforce norms by labeling counterstereotypical targets (*rewarding enforcers* hypothesis).

Finally, the effects of a derogatory label may also be moderated by perceiver characteristics (Brewer, 1988; Deaux & Major, 1987; Fiske & Neuberg, 1990; Kunda & Thagard, 1996). Studies conducted by Link et al. (1987) and Simon and Greenberg (1996) suggest that labeling effects can depend on perceivers' preexisting beliefs. Other research supports this idea, finding that whether stereotypes are automatically activated depends on a perceiver's level of prejudice (Lepore & Brown, 1997; Wittenbrink, Judd, & Park, 1997). The justification-suppression model of prejudice (Crandall & Eshleman,

2003) makes similar predictions. Open expressions of prejudice are no longer viewed as socially acceptable. Suppressing prejudice induces inner conflict, particularly for individuals high in prejudice, resulting in their increased motivation to find justifications for expressing prejudice. It is likely that overhearing a third-party apply a derogatory label would alter perceptions of situational norms; thus licensing prejudiced responses for individuals high in prejudice (Crandall & Eshleman, 2003; Ford & Ferguson, 2004). My final prediction is:

H3. Labels release prejudice. Traditional gender attitudes will moderate the effect of the derogatory gendered label on target judgments, with more prejudiced individuals producing more negative judgments of labeled targets.

PART I: DEROGATORY LABELS FOR EMOTIONALITY

Studies 1-4 Overview

The goal of the first set of studies is to investigate the effects of male- and female-derogatory labels on perceptions of male and female targets expressing negative emotions.

Emotionality, or "ineffectual, misdirected, or trivial emotion" (Shields, 2002, p. 70), has long been associated with femininity; indeed, the belief in gender differences in emotionality is one of the most enduring and widely accepted gender stereotypes (Plant, Hyde, Keltner, & Devine, 2000). Historically, emotionality has been contrasted with a masculine emotion style, characterized by high-minded, controlled passion. The latter emotional style, termed 'manly emotion' by Shields (2002), represents the ideal standard of appropriate emotion for women and men. Normative feminine emotion is incompatible with the dominant norm of emotion. Thus, while men's emotions are evaluated only

against the dominant standard of emotion, women face two sets of emotional expectations (one based on their gender and one based on the dominant standard). These incompatible sets of expectations produce an emotional double bind for women---they are expected to adhere to manly emotion standards while also being sufficiently emotionally expressive. Moreover, gender-emotion stereotypes result in different interpretations of the causes of women's and men's emotions, such that perceivers tend to interpret a woman's emotion as arising from her inherent (female) emotionality, and a man's emotion as a reasonable response to the situation (Barrett & Bliss-Moreau, 2009; Robinson & Johnson, 1997; Shields, 2002).

In addition to differences in emotional style, gender-emotion stereotypes dictate specific display rules for men and women. In accordance with women's lower status in society, women are expected to express more self-directed, powerless emotions, such as sadness and fear (Barrett & Bliss-Moreau, 2009; Hess et al., 2000; Plant et al., 2000; Robinson & Johnson, 1997; Timmers, Fischer, & Manstead, 2003). Conversely, as a high-status group, men are expected to display other-directed, powerful emotions, such as anger and contempt (Hess et al., 2000; Plant et al., 2000; Timmers et al., 2003). These status-linked display rules reinforce gender-status divisions by prohibiting certain emotional displays for men and women. As a low-status group, women are proscribed from powerful emotions, such as anger (Brescoll & Uhlmann, 2008; Lewis, 2000; Salerno & Peter-Hagene, 2015). Men are proscribed from powerless emotions that undermine their status, such as sadness (Brescoll & Uhlmann, 2008; Tiedens, 2001).

Studies 1-4 assess how derogatory gendered labels reinforce gender-emotion stereotypes. In order to evaluate the aforementioned hypotheses, I varied the gender-stereotypicality of the emotional display and the gender stereotypicality of the label.

STUDY 1

Brescoll & Uhlmann (2008) found that, in an interview scenario, participants negatively evaluated a female target expressing anger, because they made more internal attributions for her anger. In contrast, they made more situational attributions for male anger. Importantly, they also found that this pattern for angry but not sad female targets; in contrast, participants evaluated sad male targets as less competent than angry male targets. These authors argue that, in line with backlash theory, counterstereotypical emotional behavior violates expectations, which provokes negative reactions from perceivers and shapes attribution for emotion. Study 1 attempts to extend this research by examining the effects of a female-derogatory label on perceptions of male and female targets expressing anger.

Taking all hypotheses into account, the strongest prediction is that participants high in sexism will derogate labeled angry targets (and internally attribute their anger), when they are female, but not male. That is, people who hold hostile, misogynistic views will be more responsive to the label when applied to a woman, whether because they wish to confirm stereotypic biases, punish gender deviants, and/or justify their prejudice. Sexism should also predict positive evaluations of the labeler when the female target, but not male target, is labeled. Alternatively, in accordance with other studies (e.g., Brescoll & Uhlmann, 2008), the label may similarly affect female targets and their labelers, regardless of sexist ideology.

Study 1 Method

Participants

Five hundred paid participants were recruited through Amazon's Mechanical Turk. Participants who met the following criteria were included in analyses: were willing to participate in the follow-up study (more information below), correctly identified the main characters in the scenario, and correctly responded to an attention check question. Of the total sample, 243 participants met the requirements for inclusion (51% Female, 48% Male, 1% nonbinary/other; M_{age} = 38.73, SD_{age} = 11.96).

Materials and Procedure

This study was completed in two separate sessions. First, participants responded to a HIT (human intelligence task), which described the study's purpose as investigating how people form impressions of others. Participants who consented to participate read the scenario and rated the target and labeler. Two days after this part of the study ended, the experimenter invited participants (via email) to participate in the second part. Sexism scores were measured in the second session to ensure that they were not influenced by the manipulation. Two hundred and forty-three individuals (72% of participants who met the other inclusion criteria) completed both sessions.

Scenario. The scenario was a modified version of that used by Smith (2016). Participants were asked to imagine themselves in a workplace scenario in which they observed a target (Lauren/Brian) angrily confronting a male coworker, and subsequently, overheard a second male coworker gossip about Lauren/Brian. In the label condition, this second male coworker assigned a derogatory label to the target ("She's/he's so

emotional. Sometimes I swear she's/he's bipolar"). In the no label condition, this line was omitted. Participants were randomly assigned to 1 out of 4 possible scenarios (2 [target gender] x 2 [label condition]). After reading the scenario, participants rated the target first and then the labeler. The scenario is provided below:

> *You are a sales representative at Signal Advertising Agency. Two of your coworkers, (Lauren/Brian) and Chris, are assigned to work together on a* project. *You and (Lauren/Brian) are at the copier when you overhear Chris talking with the boss. Chris tells the boss, "The project (Lauren/Brian) and I are working on together isn't finished yet. The issue is, well, (Lauren/Brian) hasn't been pulling (her/his) weight." (Lauren/Brian) is furious and storms over to interrupt the conversation. (She/he) denies Chris's claims and shows the work (she/he) has done, while glaring at Chris. Later, when you go to the break room for some coffee, you hear your coworker, Jordan, and another coworker, talking about (Lauren/Brian). Jordan remarks, "(Lauren/Brian) can be difficult to work with. (She's/he's) so emotional. Sometimes I swear (she's/he's) bipolar."*

These particular labels were selected because they are relevant to emotional behavior, and they represent common categories of DGLs that undermine women and men. Historically, madness has been used to pathologize the feminine experience (Ussher, 1992). Mental illness labels like 'bipolar' reinforce gender by assigning stigma to women who act out devalued female roles, as well as women who dare to challenge them (Chesler, 1972).[2] For men, DGLs that ascribe feminine characteristics, such as

[2] 'Bipolar' is a clinical label used for individuals who have been diagnosed with bipolar disorder (although person-first language is preferred, American Psychiatric Association, 2017; National Alliance on Mental Illness, 2017). By using this label, I do not intend to trivialize the experiences of people with this mental health issue. My interest lies in how certain mental health labels are appropriated for informal use as gendered insults.

'wuss,' are used to enforce masculine prescriptions (James, 1998; Pascoe, 2005; Saucier et al., 2015). These labels target men perceived as insufficiently masculine. Both labels derogate femininity by linking it to inherent madness or impotence, as well as enforce men's and women's compliance with gender norms.

To assess evaluations of targets, perceived likability and competence were measured. These variables are commonly used in backlash research (e.g., Moss-Racusin et al., 2010; Phelan et al., 2008).

Likability. Three items assessed the likability of targets and labelers: She/he is (1 = unlikable; 11 = likable); "I would like to get to know her/him better"; "She/he is the kind of person that I tend to avoid" (r); 1 = strongly disagree; 11 = strongly agree). In Studies 1-3, this scale was reliable for targets (αs = .87-.89), as well as for labelers (αs = .90-.91).

Competence. Participants rated targets and labelers on an 11-point semantic differential scale (incompetent-competent (11); ignorant-knowledgeable (11); Studies 1-3 Target αs = .79-.86; Labeler αs = .78-.83).

Internal attribution. Internal attributions for targets were measured on an eleven-point semantic differential scale using the following single item, "In general, do you think Lauren/Brian is an in control or out of control person?" (1 = in control, 11 = out of control; Brescoll & Uhlmann, 2008).[3]

Ambivalent sexism inventory (follow-up study). Participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996). The 22-item ASI measures

[3] Additional measures were collected in the first part of the study but are not reported here.

hostile sexism (HS) and benevolent sexism (BS).[4] 11-items tap openly hostile attitudes toward women (e.g., "Women are too easily offended."), and 11-items tap subjectively positive but paternalistic attitudes toward women (e.g., "Women should be cherished and protected by men."). Participants responded to each of the 22 items on a 6-point scale (1 = disagree strongly to 6 = agree strongly). Hostile sexism is known to predict negative attitudes toward gender-deviant women (for a review, see Connor, Glick, & Fiske, 2016).

Study 1 Results

Regression analyses were used to assess the effects of label condition manipulations on ratings of male and female targets. Each dependent variable was regressed on mean-centered HS, mean-centered BS, target gender, label condition, and their interactions (Step 1).[5] To evaluate whether participant gender moderated effects, participant gender and its interactions were entered into each model and retained if they accounted for significant variance beyond Step 1. In order to minimize the influence of outlying observations, analyses were conducted using robust regression, which downweights estimates based on the size of their residuals. Specifically, an initial M-S estimator (Maronna & Yohai, 2000), and a final M-estimator with Tukey's bisquare weighting function, were used to estimate regression coefficients (lmrob in the robustbase R package).

If, as predicted, people high in hostile sexism are particularly responsive to the derogatory label and see it as an opportunity to derogate female targets, then three-way interactions between hostile sexism, label condition, and target gender should occur for

[4] Both HS and BS scores were collected and included in analyses; however, because my primary interest was in hostile sexism, I do not report results for benevolent sexism.
[5] Interactions between HS and BS were not included in the model.

target evaluation variables (likability, competence). Further, consistent with Brescoll and Uhlmann (2008), for people high in hostile sexism, internal attributions should statistically account for the relationship between the label and evaluations of the female target. Also, the same three-way interactions should be observed for labeler evaluations, such that people high in hostile sexism provide higher ratings of labelers when they label female targets. Alternatively, if sexism does not moderate the effects of target gender and label condition, we should observe two-way interactions between these variables.

Target Ratings

Likability. For the sake of comprehensibility, detailed results for all variables in this study, including this one, are listed in Table 1. The regression did not yield the expected three-way interaction; however, a four-way interaction between participant gender, HS, target gender, and label condition was observed. To inspect the interaction, analyses were conducted separately for male and female participants. For male participants, there was a two-way interaction between target gender and label, as well as one between target gender and hostile sexism. Male participants rated the female target as less likable when she was labeled relative to when she was not labeled (see Figure 1). In contrast to the hypothesis that labeling effects should be strongest for those high in sexism, hostile sexism negatively predicted likability ratings for the female target in the no label condition only.

The expected 3-way interaction between HS, target gender, and label condition was significant for female participants only. Follow-up analyses showed that the Label x HS interactions, though non-significant, were in the opposite direction for the female and

male target, such that HS negatively predicted likability for the male target when he was not labeled but for the female target when she was labeled.

Thus, the label led to lower evaluations of the female target's likability, but for male participants only. For male participants, hostile sexism also led to lower likability ratings for the non-labeled female target, but it did not moderate the effects of the label. For female participants, hostile sexism did moderate labeling effects, such that high HS women responded in the direction of male participants, rating the female target lower on likability when she was labeled.

Competence. Again, a four-way interaction between HS, target gender, label condition, and participant gender emerged, and data were split by participant gender. For male participants, there were two-way interactions between target gender and label condition and target gender and HS. Surprisingly, male participants rated the male target as more competent when he was labeled than when not labeled. The label did not affect ratings of female target's competence (see Figure 2). In addition, men's HS predicted lower competence ratings for the female target in the no label condition. There were no significant results to report for women. Similar to likability, only men showed significant labeling effects and hostile sexism did not moderate labeling effects. At odds with expectations, the female-derogatory label boosted the male target's competence.

Internal Attribution. For internal attribution, there was a significant 3-way interaction between target gender, label condition, and participant gender. Examining men and women separately, the Target Gender x Label interaction was significant for male participants only. Male participants were more likely to make internal attributions for labeled female targets relative to non-labeled female targets, and they were less likely

to make internal attributions for labeled male targets relative to non-labeled male targets. See Figure 3. Consistent with predictions, the labeled female target's anger was attributed more to her person, although this effect was for male participants only. In contrast, the label led to lower internal attributions for the male target.

Mediation

Target Likability. Due to inconsistent results for female participants, internal attribution was tested as a mediator of the effects of the label on target likability for male participants only. For male participants evaluating male targets, the label increased likability, but not significantly so (b = .68, p = .15; Figure 4a). As significant indirect effects can be observed even when the total effect is not significant (Hayes, 2009), I proceeded to test the indirect effect. A Sobel test indicated a significant indirect effect (Indirect: .63, z = 2.12, p < .05). Thus, for the male target, the label indirectly increased likability through internal attributions. For female targets, the label decreased male participants' likability ratings (b = -1.50, p < .01). The effect of the label on likability was significantly reduced when internal attribution was added to the model (Indirect: -1.03, z = -2.29, p < .05; Figure 4b), suggesting that, for male participants, perceptions of labeled female targets as emotionally out of control explained their lower perceived likability.

Target Competence. Again, mediation analyses were conducted for male participants only. For male participants evaluating male targets, the label increased perceived competence (b = .79, p < .05). When including internal attributions, the effect of label on competence was no longer significant (b = .15, p = .59). See Figure 5a. A Sobel test of the indirect effect through internal attribution provided evidence of mediation (Indirect: .62, z = 2.26, p < .05). Relative to non-labeled male targets, labeled

male targets were seen as more competent because their anger was seen as less internally caused. For female targets, effects were again in the opposite direction. The total effect of the label on female targets' perceived competence was not significant (b = -.59, p = .17). The indirect effect of the label through internal attribution was marginal (Indirect: -.37, z = 1.74, p = .08; Figure 5b).

Labeler Ratings

Likability. Overall, people disliked the labeler when he labeled the target, regardless of gender. However, there was also an interaction between HS, target gender, and label condition. To explore this interaction, separate analyses were conducted for male and female *targets.* The interactions between HS and label, though non-significant, were in opposite directions for the male compared to the female target (Figures 6a & b). HS positively predicted likability of the labeler when the female target was labeled; in contrast, the relationship between HS and likability was in a negative direction for labeled male targets. Not surprisingly, people downgraded evaluations of the labeler when he used a derogatory label, and this was especially true of participants high in hostile sexism evaluating the male target. However, for the female target, this was primarily true of people low in HS. Individuals high in HS did not respond as negatively to the labeler when he referred to her using a derogatory term.

Competence. HS, target gender, and label significantly interacted to predict labeler competence. Separate analyses were conducted for male and female *targets*, and again, interactions between HS and label were non-significant but in opposite directions for male and female targets. The labeler received lower competence ratings when he labeled the male target relative to when he did not, particularly by people high in HS

(Figure 7a). Label condition also negatively predicted labeler competence for female targets, but the relationship between HS and labeler competence marginally differed for labeled and non-labeled female targets. HS positively predicted labeler competence when the female target was labeled but was unrelated to competence when she was not (Figure 7b). Again, people high in HS withheld negative evaluations of the labeler when he labeled the female target; however, they evaluated the labeler as less competent when he labeled the male target.

Study 1 Summary

Study 1 found that a female-derogatory label negatively affected perceptions of an angry female target, such that male participants saw her as less likable. Men also attributed her anger more to her person, which explained the decrease in likability. Unexpectedly, the same label increased men's perceptions of the male target's competence. Furthermore, men attributed the male target's anger less to internal factors, which accounted for more positive evaluations of the male target. There were few significant findings for female participants. However, like men, high HS women reacted more negatively to labeled female targets than similar male targets.

In addition to perceptions of targets, Study 1 also collected perceptions of the labeler. Overall, people evaluated the labeler more negatively when he labeled targets relative to when he did not. However, people high in HS responded differently than people low in HS by withholding negative evaluations of labelers of female targets.

Study 1 provides some support for all three hypotheses. Consistent with the moderation hypothesis, effects of the label on the labeler were dependent on hostile

sexism. The finding that the label negatively affected the female target but positively affected her labeler fits the stereotype activation, as well as both backlash hypotheses.

However, the prediction that participants high in hostile sexism would be particularly responsive to the label was not consistently supported. Rather, male participants, compared to female participants, were more responsive to the label. Similar studies have not found participant gender differences (e.g., Barrett & Bliss-Moreau, 2009; Brescoll & Uhlmann, 2008; Shield & Crowley, 1996), but as dominant group members, men benefit from maintaining their group dominance, and thus endorse hierarchy-enhancing ideologies (Pratto, Sidanius, & Levin, 2006), such as sexism (Glick & Fiske, 1996; Swim, Aikin, Hall, & Hunter, 1995), more than women do. Therefore, compared to women, they might be expected to judge the female target more harshly.

Study 2 repeats the experiment with a stereotypically feminine emotional expression (crying), to determine whether these findings are reproducible.

STUDY 2

Study 2 investigates the effects of the same female-derogatory label; but now, rather than targets reacting with anger at their coworker, they reacted with tears. Unlike anger, crying is stereotypically feminine and violates masculine proscriptions against weakness (Fischer, Eagly, & Oosterwijk, 2013; Lombardo, Cretser, & Roesch, 2001).

As in Study 1, the stereotype activation perspective predicts that the label will activate stereotypes of female emotionality, which should result in negative evaluations of the female target. However, in Study 2, the punishing deviants hypothesis makes a different prediction: since male targets in this study express stereotypically feminine emotion, the label should lead to more negative evaluations of the male (not female)

target. Based on the results of Study 1 and the moderation hypothesis, I expect that, in either case, men high in sexism will be most likely to show the predicted effects. Finally, the rewarding enforcers hypothesis suggests that participants may show their approval of backlash by providing less negative evaluations of labelers who label male targets; this is expected to be primarily true of people who endorse hostile sexism. Thus, Study 2 evaluates the extent to which the effects of DGLs on targets and labelers are consistent with stereotype activation or backlash perspectives.

Study 2 Method

Participants

The total sample consisted of five hundred participants recruited and paid through Amazon's Mechanical Turk. Two hundred thirty-five participants (68% of whom completed the follow-up study) met the requirements for inclusion (50% Female, 50% Male, < 1% nonbinary/other; M_{age} = 38.71, SD_{age} = 12.09).

Study 2 Results

According to the stereotype activation hypothesis, the derogatory label should decrease evaluations of the female target through increased internal attributions, particularly for hostile sexist men. Another possible pattern of results is predicted by the punishing deviants hypothesis: the derogatory label will decrease evaluations of the *male* target through increased internal attributions, particularly for hostile sexist men. Each of these hypotheses corresponds to a four-way interaction between participant gender, hostile sexism, label condition, and target gender. My hypothesis for labeler ratings predicts a three-way interaction between hostile sexism, label condition, and target

gender, such that for participants high in hostile sexism in particular, the label will lead to more positive evaluations of the labeler when he labels a male target.

Target Ratings[6]

Likability. The expected four-way interaction was significant. For male participants, there was a significant three-way interaction between HS, target gender, and label condition. In the control condition, men's hostile sexism negatively predicted likability of the male target, but it positively predicted likability of the female target. In the label condition, however, HS negatively predicted likability ratings for the female target. Consistent with Study 1, and the stereotype activation hypothesis, the label facilitated more negative evaluations of the female target's likability, but only for men high in HS (see Figure 8a). In contrast, hostile sexist men withheld negative judgment of male target when he was labeled (see Figure 8b). No significant effects were found for female participants.

Competence. For target competence, there was also a four-way interaction between HS, target gender, label condition, and participant gender. Three-way interactions between HS, target gender, and label condition were observed for male and female participants. For female participants evaluating female targets, there was a marginal effect of Label x HS, such that women high in HS tended to rate the female target as more competent when she was labeled, relative to those low in HS (Figure 9b). Aside from this interaction, no other effects for female participants approached significance. Results for male participants were in the opposite direction, such that men high in HS more negatively evaluated the female target's competence than men low in

[6] Detailed results for all variables in this study are listed in Table 1.

HS, but only when she was labeled (Figure 9a). No significant effects emerged for male targets.

Internal Attribution. The predicted four-way interaction was not observed, but there was a significant three-way interaction between HS, Target Gender, and Label Condition. Looking at male and female targets separately, HS modified the effect of the label for female targets, such that HS was positively related to internal attribution ratings in the label condition only. However, for male targets, HS was positively associated with internal attribution ratings, regardless of condition. Similar to Study 1, but for high HS individuals only, the label increased internal attributions for the female target. See Figure 10.

Mediation

Study 2 conducted the same mediational analyses as in Study 1. As in Study 1, Study 2 restricted analyses to male participants. However, since HS moderated the effects of the label on target ratings, with men high in HS showing correlations consistent with mediation, estimates were calculated at +1 SD HS.

Target Likability. The label decreased likability of the female target (b = -2.38, p = .001). When including internal attributions in the model, the effect of label on likability was reduced to non-significance (p = .12). Consistent with the stereotype activation hypotheses, the indirect effect was significant (Indirect = -1.02, z = -2.04, $p < .05$), suggesting that the labeled female target was seen as less likable because her sadness was attributed to her person. See Figure 11. For male participants evaluating male targets, the label increased likability ratings, but not significantly so (b = 1.31, p = .11). The test of the indirect effect was not significant (Indirect = .43, p = .40).

Target Competence. The mediation analysis was also conducted on hostile sexist men's competence ratings. The effect of the label on the female target's competence was marginal (b = -1.10, p = .10), but it was reduced with the inclusion of the attribution measure in the model (b = -.54, p = .40; Indirect = -1.20, z = -1.73, p = .08). See Figure 12. Mediation was not present for male targets.

Labeler Ratings

Likability. Results revealed the predicted three-way interaction between HS, label condition, and target gender. In contrast to findings for target ratings, and consistent with the rewarding enforcers hypothesis, HS positively moderated the effect of the label on labeler likability for *male* targets only. In the label condition, HS predicted greater likability of the labeler, and participants high in HS rated the labeler as more likable relative to the control condition (Figure 13).

Competence. HS also moderated the Target Gender x Label interaction for labeler competence. Again, the HS x Label interaction was significant for male targets only. Hostile sexism positively predicted ratings of labeler competence in the label condition (Figure 14). When a counterstereotypical male target was affixed with a female-derogatory label, participants high in HS responded by increasing competence ratings of the labeler.

Study 2 Summary

Study 2 results for female targets mirrored those for Study 1, with the exception that they depended on hostile sexism. The label negatively affected perceptions of a crying female target, such that male participants high in hostile sexism saw her as less likable and competent, and they attributed her crying more to her person, which

explained the lower likability ratings. These results support the stereotype activation hypothesis. The female-derogatory label facilitated negative dispositional attributions about the female target, irrespective of the stereotypicality of her behavior.

Contrary to the punishing deviant hypothesis, I did not find that the label lead to more negative perceptions of the male target. By contrast, results for *labelers* were consistent with the rewarding enforcers prediction: hostile sexism predicted more positive evaluations of the labeler's likability and competence when he labeled the counterstereotypical male relative to when he did not. As expected, these relationships were not present for the female target, whose behavior did not challenge gender norms.

In sum, results for Studies 1 & 2 largely support the stereotype activation account for female targets and the backlash account for labelers of counterstereotypical males. Also consistent with expectations, results were dependent on participant gender and sexism. In Study 3, I extend this research by exploring the effects of male-derogatory labels.

STUDY 3

In Study 3, I shift focus to a male-derogatory label ('pathetic/wuss'). As in Study 2, targets expressed sadness by crying. Male-derogatory labels that ascribe weakness, femininity, or homosexuality are used to police the boundaries of masculinity (Pascoe, 2005; Saucier et al., 2015). Therefore, this label is not expected to reinforce stereotypical judgments of male targets.

Here I predict that, consistent with backlash, the label will lead to more negative evaluations of the counterstereotypical male, particularly for male participants. I also expect that the label will lead to less negative evaluations of the labeler when he labels a

counterstereotypical male. Moreover, traditional ideology about the male role is expected to moderate these effects, such that these responses will be strongest among those who most endorse masculine ideology.

Study 3 Method

Participants

Five hundred paid participants were recruited through Amazon's Mechanical Turk. Out of this sample, 247 participants met the requirements for inclusion (63% Female, 37% Male; M_{age} = 40.34, SD_{age} = 13.34).

Male Role Norms Inventory (follow-up study). Because this study employs a male-derogatory label, attitudes toward men should be more relevant to judgments than attitudes toward women. Thus, the measure of traditional gender ideology was changed to the Masculine Role Norms Inventory (Levant, Hall, & Rankin, 2013). This 21-item scale consists of seven subscales (toughness, homophobia, dominance, avoidance of femininity, repair skills, sexuality, and restrictive emotionality) that measure norms of traditional masculinity (e.g., "Men should be the leader in any group"; "Men should be detached in emotionally charged situations"). Participants responded on a 7-point scale (1 = strongly disagree to 7 = strongly agree). Scores are averaged to create an overall index of masculine ideology (α = .93). People who strongly endorse masculine ideology are expected to be motivated to derogate feminine men when given the opportunity to do so. In fact, rejection of the feminine has been identified as a central component of masculinity prescriptions (Kimmel, 1997). In the follow-up study, 247 individuals (76% of participants who met the other inclusion criteria) completed the MRNI.

Study 3 Results

Similar to Study 3, for target evaluations, I predict a four-way interaction between participant gender, masculine ideology, label condition, and target gender such that, when the male target is labeled, men who endorse masculine ideology will evaluate him more negatively. For labeler evaluations, I predict three-way interactions between masculine ideology, with people who endorse masculine ideology providing more positive evaluations of the labeler when he labels a male target.

Target Ratings

Likability. There were no significant results to report for this measure.

Competence. The regression produced not a four-way interaction, but a significant 3-way interaction between MRNI, Target Gender, and Label. Further inspection revealed that, consistent with expectations, the label affected perceptions of target competence for the male, but not for the female, target. However, opposite to expectations, masculine ideology *positively* predicted the male target's competence when he was labeled. Further, the label decreased perceptions of target competence for people low in masculine ideology, whereas it did not affect perceptions of competence for people high in masculine ideology. See Figure 15. Thus, results are inconsistent with predictions that individuals (and particularly men) high in masculine ideology would provide more negative evaluations of the labeled male target.

Internal Attribution. Results of the regression revealed a three-way interaction between MRNI, Target Gender, and Label. Decomposing the interaction showed a significant two-way interaction between MRNI and Label for the male target, but not for the female target. Again, in contrast to expectations, when male targets were labeled,

people high in MRNI, regardless of gender, attributed the male target's reaction less to his person. Further, for people high in MRNI, the label decreased internal attributions, while for people low in MRNI, it did not significantly affect attributions (Figure 16).

Mediation

Target Likability. Since there were no results for target likability, the mediation model was not tested.

Target Competence. Although the effects of MRNI were inconsistent with predictions, for the sake of consistency with Studies 1 & 2, mediation was assessed for the male target. MRNI moderated the effects of the label on target competence, with people low in masculine ideology responding in a direction opposite to those high in masculine ideology. Figure 17 shows two mediation models, one depicting relationships for those low in masculine ideology (-1 SD); the other depicting relationships for those high in masculine ideology (+1 SD). The indirect effect for those low in masculinity was not significant (Indirect = -.19; z = -1.07, p = .29), but the indirect effect for those high in masculinity was significant (Indirect = .47; z = 2.17, $p < .05$). In sum, opposite to expectations, masculine ideologues reacted to the label by decreasing internal attributions, which in turn, led to increased ratings of the male target's competence.

Labeler Ratings

Likability. The predicted three-way interaction between masculine ideology, target gender, and label condition emerged, though it was marginally significant (p = .08). People thought the labeler was less likable when he used a male-derogatory label for a stereotypical female target. For the male target, the effect of label depended on masculine ideology. Masculine ideology positively predicted likability of the labeler in

the no label condition. Unexpectedly, people high in masculine ideology reacted to the label by decreasing ratings of labeler likability. However, the label did not decrease labeler likability for people low in masculine ideology (see Figure 18).

Competence. As predicted, masculine ideology moderated the effect of Target Gender x Label on labeler competence. For female targets, the label decreased ratings of the labeler's competence, regardless of the participant's level of masculine ideology. For male targets, the label's effects depended on MRNI. At odds with expectations, people scoring 1 SD above the mean for MRNI evaluated the labeler as less competent in the label, relative to the no label, condition. Finally, these individuals also evaluated the labeler as less competent than those low in MRNI in the label condition. See Figure 19.

Study 3 Summary

Contradicting hypotheses, people high in masculine ideology exhibited reactance against the label, in both target and labeler ratings, by rating the male deviant more positively and the labeler more negatively. Moreover, they attributed the labeled male deviant's reaction less to his person, which explained the increase in perceptions of his competence. Further contradicting predictions, these findings did not depend on participant gender.

By contrast, individuals low in masculine ideology responded more in line with predictions: when the male target was affixed with a label, they perceived him as less competent and more personally responsible for his reaction. They also rewarded the labeler of the male target with increased competence ratings. Also consistent with expectations, female targets were relatively unaffected by the label. Study 4 will attempt to address these unexpected results.

STUDY 4

In Study 3, a measure of masculine ideology was chosen in lieu of the Ambivalent Sexism Inventory (Glick & Fiske, 1996, used in Studies 1 & 2) with the reasoning that it would better predict responses toward non-conforming male targets. Masculine ideology refers to the internalization of normative beliefs about masculinity and men's traditional roles (Levant et al., 2013; Levant & Richmond, 2008). It is distinct from attitudes toward women or gender relations (Levant & Richmond, 2008; Thompson, Pleck, & Ferrera, 1992), as well as sexism towards men (e.g., Ambivalence toward Men Inventory, Glick & Fiske, 1999). As such, masculine ideology may be more closely related to positive attitudes toward traditionally masculine men and predict negative attitudes toward non-conforming men only under certain conditions (e.g., Glick, Wilkerson, & Cuffe, 2015).

In order to negatively influence target evaluations, a label must fit the available information about the target (Greenberg & Pyszczynski, 1985). It is possible that this condition was not met. That is, perhaps individuals high in MRNI viewed the male target's behavior as only mildly inappropriate but viewed the label as overly harsh. The shifting standards model (Biernat, 2005) describes how stereotypes function as standards against which members of stereotyped groups are judged. Because men are not expected to display stereotypically feminine emotion, perceivers, particularly strong 'internalizers' of masculine norms, may require more evidence of feminine emotion to confirm that a man is emotional (Kelly & Hutson-Comeaux, 2000). If this is the case, strengthening the intensity of the targets' emotional expression may result in a different pattern of findings.

Glick et al. (2015) recently showed that men's hostile sexism predicted more negative attitudes toward non-conforming men (e.g., stay-at-home dads, feminine men,

feminist men), whereas masculine identification did not. Therefore, Study 4 will also include a measure of ambivalent sexism to allow for the possibility that hostile sexism, rather than masculine ideology, will moderate effects.

Study 4 Method

Participants

Five hundred Amazon Mechanical Turk workers participated in the study. After excluding those who did not meet inclusion criteria, there were 314 participants (48% Male, 52% Female; M_{age} = 36.87, SD_{age} = 10.4).

Revised scenario. The scenario was modified to increase the intensity of targets' emotional expression. Changes are in bold.

> (Lauren/Brian) **bursts into tears** and rushes over to interrupt the conversation. **In an unsteady, cracking voice,** (she/he) denies Chris's claims and shows the work (she/he) has done, while tears stream down (her/his) face. **As (she/he) continues to speak, she begins to sob.**

Masculine ideology and sexism. In addition to changing the scenario, Study 4 measured masculine ideology, as well as ambivalent sexism. Participants completed the Masculine Role Norms Inventory (MRNI; Levant, Hall, & Rankin, 2013) and the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996) in a follow-up study. Of the individuals who met the other inclusion criteria, 76% completed the follow-up study. Coefficient alphas were .95 for the MRNI and .91 & .94, for the BS and HS subscales, respectively. The MRNI showed moderate correlations with the ASI sub-scales (HS = .44, controlling for BS; BS = .44, controlling for HS).

Study 4 Results

Separate regression analyses examined how masculine ideology and ambivalent sexism predicted target and labeler ratings. Predictions were the same as for Study 3, except hostile sexism, rather than masculine ideology, was the predicted moderator. However, despite these predictions, findings for hostile sexism were very similar to those for masculine ideology; thus, for consistency with Studies 1 & 2, I report results for hostile sexism only.

Target Ratings

Likability. Hostile sexism negatively predicted target likability, regardless of target or label condition. Further, a two-way interaction between target gender and label condition emerged. The label increased likability of the male, but not female, target. Figure 20 depicts this interaction.

Competence. Participants rated the female target as more competent than the male target. However, as depicted in Figure 21, this relationship depended on their level of hostile sexism. Relative to participants low in hostile sexism, participants high in hostile sexism rated the female target as less competent.

Internal Attribution. Regardless of label condition or sexism, people ascribed the male target's behavior to internal factors more than they did for female targets.

Mediation

Because label condition did not affect internal attributions, I did not conduct mediation analyses.

Labeler Ratings

Likability. The label decreased labeler likability for both male and female targets. Participants liked the labeler of the female target less, but this depended on HS. Individuals high in HS tended to rate the labeler as more likable than those low in HS, particularly when the target was female (Figure 22).

Competence. Regardless of target gender, the label decreased perceptions of labeler competence. This effect was marginally dependent on hostile sexism ($p = .06$), such that HS positively predicted labeler competence in the control condition, but not in the label condition (Figure 23).

Study 4 Summary

The results of Study 4 did not support hypotheses. Rather, the label *increased* the likability of the counterstereotypical male target, regardless of gender ideology or participant gender. For men and women high in HS, the label decreased ratings of labeler competence for both male and female targets. Further, HS positively predicted likability of labelers of female targets, regardless of label condition. Finally, in contrast to Study 3, traditional gender ideology did not predict more positive evaluations of the male target.

In hindsight, the changes to the manipulation were likely too severe. Ambiguous target information increases the likelihood of stereotyping (Kunda & Sherman-Williams, 1993; Locksley, Borgida, Brekke, & Hepburn, 1980). Compared to Study 3, in Study 4, targets' emotional behavior was more extreme (particularly within the context of the workplace) and less open to interpretation. As a result, participants seemed to assign less weight to the label and more to targets' emotional reactions. This explanation is supported by the relatively few differences observed between the label and no label conditions.

Interestingly, in our statistical models, the Hostile Sexism subscale was virtually indistinguishable from the Male Role Norms Inventory (r = .63), producing very similar patterns of results. This is a surprising finding. Although both measures relate to the internalization of traditional gender role expectancies, the MRNI is conceptualized as a measure of gender role strain, and the Ambivalent Sexism Inventory as a measure of prejudice. However, as Study 3 suggests, these measures may differ in valence and scope, such that the MRNI taps positive attitudes toward traditional masculinity, while HS captures negative attitudes toward gender stereotype violators. In cases where this distinction is not relevant, both measures may produce very similar results.

PART I: DISCUSSION AND CONCLUSION

Across four studies, I tested three hypotheses concerning how DGLs affect evaluations: they activate stereotypes, they justify backlash, and they release prejudice. Overall, results provide some support for all three predictions.

Studies 1 and 2 showed that, for male participants, when female-derogatory labels are applied to female targets, they lower target evaluations and reinforce the view of the female's emotion as a trait-like response. These effects sometimes depend on hostile sexism, with hostile sexism strengthening these effects. These findings are consistent with a stereotype activation explanation.

In Studies 1, 2, and 3, when labelers labeled a counterstereotypical target (sad man, angry woman), some perceivers reacted by rewarding them with more positive evaluations or by withholding negative evaluations. On the other hand, labeling a stereotypical target (as in Studies 1 & 3) often decreased evaluations of the labeler. These results are consistent with participants rewarding enforcers of social norms, suggesting

that evaluations of the labeler are driven by the stereotypicality of the target in addition to the label.

Studies 1 and 2 also supported the releasing prejudice hypothesis. In Study 2, hearing a coworker use a derogatory label to describe a female target, gave men high in hostile sexism license to derogate her (Crandall & Eshleman, 2003; Ford & Ferguson, 2004). Further, in both studies, people high in HS evaluated labelers more positively when they used derogatory labels for female targets.

Results for Study 3 were unexpected, but not entirely inconsistent with Studies 1-2. Rather than derogate the labeled male target, individuals high in masculine ideology defended him. This finding can be interpreted as evidence of DGLs producing negative target evaluations only when they affirm pre-existing beliefs about targets. Further, Study 3 also supports the mediating role of internal attributions: masculine ideologues discounted the label and justified the male target's reaction by attributing his behavior less to internal factors. Gender traditionalists interpreted the female target's behavior was through the lens of the label, but they re-interpreted the label to fit expectations of male targets (Shields & Crowley, 1996). As masculine ideology and hostile sexism were highly correlated in Study 4, it is unclear whether favoritism of men or hostility toward women account for these effects. Results of Study 4, which also examined a male-derogatory label, are less interpretable, but perhaps suggest that when target behavior is sufficiently extreme, labels are unlikely to affect evaluations.

Overall, the male-derogatory label was less damaging to the male target than the female-derogatory label was to the female target. Because the meaning and use of derogatory labels differs by target gender, I did not attempt to equate these labels. As a

result, the female-derogatory label is more directly relevant to emotion than the male-derogatory label and may fit the observed behavior better. It is also possible that, in the context of emotional behavior, DGLs will rarely 'stick' to men as they do women, because emotionality is associated with femininity (Shields, 2002). Findings indicate that for the female target, the female-derogatory label reinforced the idea that she lacked emotional control. However, for the male target, perceivers were less able to reconcile the label with expectations and more likely to discount it (Barrett & Bliss-Moreau, 2009; Brescoll & Uhlmann, 2008; Shields and Crowley, 1996). Perhaps if targets displayed other types of behaviors (e.g., refusing to confront a manipulative coworker or failing to lift a heavy file box), the male-derogatory label would have more negative impact on male targets.

Backlash researchers often find that male and female participants similarly penalize female gender deviants (Parks-Stamm, Heilman, & Hearns, 2008; Phelan et al., 2008; Rudman, 1998). However, in Studies 1 & 2, only men consistently derogated female targets. Conversely, in Studies 3 & 4, participant gender did not moderate any of the observed effects. Studies of derogatory group labels have generally focused on outgroup members' responses to overheard derogatory labels (e.g., Goodman et al., 2008; Greenberg & Pyszczynski, 1987). Less is known about how people respond to such labels when they target their own group. One study found that heterosexuals, compared to gay individuals, more strongly associated derogatory group labels for gay people with negative concepts (Carnaghi & Maass, 2007). So, it is possible that, at least in Studies 1 and 2, male and female participants differently interpreted the label. However, it is also

possible that the men and women in these studies differed on other important ideological variables not measured in this research.

Also of note, ratings of the labeler and target were often inversely related. Perceivers who agreed with the label evaluated labelers more favorably and targets less favorably; those who didn't agree with the label evaluated labelers more negatively and targets more favorability. Traditionally, impression formation studies have examined perceivers' evaluations of singular targets. However, outside of psychological experiments, impression formation rarely occurs in isolation. Perceivers often encounter more than one target at a time, and these target impressions may mutually influence one another. In the future, other researchers might wish to consider more realistic social scenarios in which people develop impressions of multiple targets.

Admittedly, the results of Studies 1-4 are quite nuanced and complex. Perhaps they can best be summarized as follows: People interpret derogatory gendered labels through a motivated lens. When it confirms their biases (e.g., people high in hostile sexism who hear a derogatory label of a woman), it leads to more negative evaluations of the target. When people are especially motivated to punish a gender deviant (e.g., people high in hostile sexism who hear a derogatory label of an angry woman), it leads to less negative evaluations of the labeler. And when it goes against their biases (e.g., high masculine ideologues who hear a derogatory label of a man), it leads to more negative evaluations of the labeler. In other words, derogatory labels feed pre-existing views, but they are rejected if they run contrary to those views.

Having provided experimental evidence on the effects of derogatory gendered labels in Part I, Part II will now move on to an examination of DGLs in a real-world context.

PART II: DEROGATORY GENDERED LABELS ON TWITTER

Part I demonstrated that the effects of derogatory gendered labels are quite nuanced, and they often depend on gender and gender ideology. Studying stereotyping in the laboratory can be challenging. Stereotype activation and application depend on a number of contextual and motivational factors (Blair, 2002; Deaux & Major, 1987; Fiske & Neuberg, 1990). Further, it is difficult to access participants' genuine attitudes with explicit, self-report measures, and when they know they are being observed (Crandall & Eshleman, 2003). Experimental studies strip phenomena of interest of their real-world context, which is necessary for tight control, but can impede understanding of the phenomena and its effects (Fine & Elsbach, 2000). Behavior observed within the artificial constraints of an experiment does not always resemble behavior as it naturally occurs. Multimethod designs that incorporate both quantitative and qualitative data help to mitigate concerns about real-world applicability (Fine & Elsbach, 2000).

Therefore, in complement to Part I, Part II investigates the content, targeting, and evaluative consequences of derogatory gendered labels in a real-world environment, and further, whether gender and gender ideology moderate their impact. Specifically, Part II examines these questions in a dataset of tweets collected from Twitter, a social media platform on which individual users may post 140-character messages (called tweets) for other users to read. Twitter is a useful context for the present study for several reasons. First, Twitter data captures near real-time reactions to real-world events. Further, socially

desirable responding is not a concern, as participants are unaware they are involved in a study. Finally, gender-based harassment is common in online environments like Twitter (Barak, 2005; Citron, 2009).

Given the wealth of data available online (e.g., on social media websites, blogs, etc.), a number of social scientists are increasingly interested in using it in their research. However, psychology has not yet adequately addressed questions of how researchers can best leverage this data. Part II demonstrates a possible approach to exploring behavioral research questions via social media data.

Studies 5-7 Overview

Studies 5-7 investigate derogatory gendered labels used on Twitter within the context of the 2016 presidential primary debates.

Over the past few election cycles, social media platforms have become an increasingly important source of political information. From 2004 to 2008, the number of people who used the internet for campaign news increased threefold from 10% to 33% (Kohut & Parker, 2008). In 2012, around a third of respondents (36%) went online for news about the presidential debates (Kohut, Doherty, Dimock, & Keeter, 2012). By 2016, according to Gottfried, Barthel, Shearer, and Mitchell (2016), close to half of those surveyed (44%) received news about the 2016 election from a social networking site. People's collective contributions via social media outlets have taken on increased cultural importance in shaping narratives about important national events like election debates (Freelon & Karpf, 2014).

Forced to adapt to the changing landscape of media, political candidates increasingly rely on social networking platforms to reach potential voters. Perhaps none

more so than Trump, who used social media, and in particular, Twitter, as a primary means of campaign communication. His tweets, which often contained provocative or "politically incorrect" content directed at rivals (including DGLs), were amplified by his followers, which contributed to coverage by other media sources (Wells et al., 2016; Zhang, Wells, Wang, & Rohe, 2018). Undoubtedly, Trump's vitriolic language toward competitors contributed to tenor of the 2016 presidential race, which some characterized as the dirtiest in recent history (Cummins, 2016). For these reasons, the 2016 election is an attractive context to study the questions of interest.

Studies 5-7 investigate derogatory gendered labels used to describe female and male moderators of the 2016 presidential primary debates. Journalism is a male-dominated profession. In recent reviews of top 20 news sources (including television, print, and online media), the Women's Media Center (2015, 2017) finds that only about a third of bylines and on-camera appearances go to women. Further, men are more likely to report on 'hard' news topics such as politics, business, or criminal justice. Men are also more likely to moderate presidential debates (Presidential Gender Watch, 2016). As of 2012, there had been only three female moderators of general election debates, spurring several high school students to (successfully) petition for a female moderator of the 2012 debates (Dittmar, 2016). Since debate moderating is a traditionally masculine domain, female moderators are expected to draw negative attention on the basis of their gender (Kanter, 1977; Rudman et al., 2012). I chose to examine debate moderators of the presidential primary debates, because: 1) women are better represented in primary than in general election debates; and 2) this permits comparison of female-male moderator pairs.

In contrast to the studies in Part I, Study 5 uses an exploratory, data-driven approach to gain a richer understanding of what labels are used, by whom, and in what context; thus, it examines broad questions rather than specific predictions. Studies 6 & 7 experimentally examine reactions to actual tweets about the debates, in an attempt to replicate the results of Part I. However, in order to identify and categorize derogatory gendered labels in the Twitter dataset, first, I developed a dictionary of slurs and general insults.

CREATING A SLUR DICTIONARY

Researchers typically use one of the following automated approaches to classify large text corpora: unsupervised machine learning, supervised machine learning, or dictionary-based methods (Grimmer & Stewart, 2012). Unsupervised learning techniques are used when categories are not known a priori; they use a bottom-up approach to categorization, sorting text into sets of naturally-occurring groups. In contrast, dictionary and supervised learning methods are used when categories are known. In supervised learning techniques, human raters code a subset of documents into predefined categories. Their ratings are then used to train a statistical model to classify the remaining data. Finally, dictionary methods classify text on the basis of a predefined list of words.

Initially, I intended to classify tweets via supervised learning methods; however, raters performed poorly at identifying DGLs in tweets. Future studies that wish to use supervised methods should consider recruiting gender scholars rather than ordinary crowdsource workers. No validated dictionaries currently exist for the categorization of DGLs. Therefore, I developed a dictionary for the present study. First, MTurk workers generated a number of DGLs, and for comparison purposes, general insults. Then three

coders trained in qualitative research categorized each term according to a set of predefined categories. Their ratings were combined to create a final dictionary.

Participants

DGLs. Participants generating DGLs were a sample of one hundred and fifty MTurk workers. Eight participants were eliminated from the sample due to poor quality responses. The final sample consisted of 142 participants (58.5% Male, 41.5% Female; M_{age} = 34.96, SD_{age} = 10.9).

General insults. Participants generating general insults were a sample of one hundred MTurk workers. Seven participants were eliminated from the sample due to poor quality responses. The final sample was comprised of 93 participants (49.5% Male, 50.5% Female; M_{age} = 34.15, SD_{age} = 10.2).

Procedure

Participants completed a short study that involved listing and defining gender-based slurs. They were informed that the purpose of the study was to create a dictionary of gender-based slurs, which were defined as words that attack or express negative attitudes towards a person on the basis of sex or gender. Instructions emphasized that participants should list slurs they have heard others use, whether or not they personally use them. Participants were asked to list 10 gender-based slurs that target women or female persons and 10 gender-based slurs that target men or male persons, in counterbalanced order. Then participants were instructed to describe the instant image evoked by each of the terms they provided (James, 1998). Participants provided a total of

608 unique slurs. Out of those slurs, 311 were female-derogatory gendered labels, and 370 were male-derogatory gendered labels.[7]

In a subsequent study, one hundred participants generated general (non-gendered) insults using the same method. This group of participants was asked to list general insults, which were defined as words that attack a person's personality, disposition, morality, appearance, or overall worth. Participants provided 307 unique insults.

I combined these two datasets, retaining an indicator for whether a label was a general insult or a gendered label. Then I retained labels listed by at least 3 participants and discarded all others. I also combined similar terms (e.g., bitch and bitchy). This resulted in a total pool of 363 labels, which was submitted to three raters experienced in qualitative research (two male, one female) to code into categories.

The category scheme was based on previous research (James, 1998; Saucier et al., 2015; see Appendix A for coding scheme and instructions). Rather than separate categories by gender (e.g., words that primarily refer to men v. those that primarily refer to women), raters classified slurs according to eight categories (Ethnic, Homophobic, Sexual, Physical, Feminine, Masculine, Incompetence, Other), in which both male-referential and female referential terms could be classified. This category scheme was not meant to be exhaustive of the data, but to include categories of interest in the present study.

Three raters were recruited via Upwork, an online freelancing site. Each rater was experienced in qualitative research and had completed or were currently completing a graduate degree in psychology or a related field. Raters were informed that the coding

[7] Numbers do not add up to 608 because 73 terms were specified as both female- and male-referential.

task was part of a study that aimed to create a dictionary of gender-based slurs for use in future research on gender bias. They were instructed to classify slurs as belonging to one or two of the categories. They were told that they could consult the definitions provided by participants, but that they should also rely on their own understanding of the term in classifying it.

When all three raters placed a given term into a category, the term was assigned to this category. When at least two of the raters placed a given term into a category, it was compared to the experimenter's independent and a priori categorizations. If the experimenter also placed the term in this category, the term was assigned to this category. If she did not, then this term was not assigned to the category ($N = 45$[8]). If raters did not agree on any category for a term, this term was removed ($N = 23$). See Table 2 for reliability and example words by category.

Discussion

Ideally, each category would have the same number of female-derogatory and male-derogatory terms; however, it proved impossible to equate them. The specific labels used to derogate men and women evolve from differing sets of gendered proscriptions. For example, a large number of words for women have sexual connotations, many of which cast women as inappropriately sexual. Contrastingly, words for men in this category are fewer in number and qualitatively different in meaning (e.g., “dog”; “player”). Thus, the gender asymmetry across categories is part of the phenomenon itself.

[8] The majority of these disagreements concerned the physical appearance category ($N = 25$), which raters used less discriminately than the experimenter. This category proved problematic, as discussed more fully later.

The emergence and staying power of certain DGLs are instructive with respect to society's construction of gender-appropriate roles and behavior (James, 1998).

Dictionaries are often used to classify words without consideration of their meaning in context. However, it is important to note that, for some words in particular, categorization depends heavily on contextual meaning. For example, 'bitch' was classified as both Feminine / Weak and Masculine / Aggressive, depending on a target person's gender. The majority of participants described "bitch" in terms of an aggressive woman ("woman that is angry and bitter"; "A mean argumentative attractive girl"; "A female that is bossy, rude, or stubborn"), but a significant proportion of participants also used this label to describe a feminine man ("A guy who doesn't stand up for himself"; "a weak willed man"; "a guy who's not tough"). In addition, the same word can be used in ways that are not derogatory (e.g., "Rachel Maddow is my biggest girl crush of all time because she is the baddest political bitch around on MSNBC"). Thus, while the dictionary can be used to identify potential occurrences of DGLs, if possible, researchers should verify classification accuracy in (at least) a subset of documents.

Unfortunately, the physical appearance category proved problematic in terms of rater consistency. In hindsight, because participants were instructed to describe the instant image evoked by each term, most descriptions focused on physical characteristics. While I assigned slurs to this category only when I felt physicality was essential to the slur's meaning (e.g., 'fat', 'bimbo'), a least one rater treated it as a default category, and he assigned a word to that category whenever participants' descriptions mentioned physical features. Despite its low interrater reliability, I examine this category in analyses; however, it needs to be revised before it can be used in future research.

STUDY 5

Labeling Effects on Twitter

In the previous study, I created a dictionary for DGLs and general insults, with which to classify a corpus of tweets about the 2016 presidential primary debates. Study 5 uses this dictionary to explore the following questions, which align with the predictions made in Part I:

Who gets labeled? Among the first in a traditionally male-dominated field, the female moderators are more visible as representatives of their group and are more likely to be scrutinized, relative to their male counterparts (Kanter, 1977)**.** Moreover, as women in a male-dominated profession, they challenge gender stereotypes. Therefore, in accordance with a backlash perspective, they are expected to receive more derogatory gendered labels.

It is unlikely that male targets will avoid receiving negative tweets. In fact, they may receive a greater number of general insults and personal attacks than female targets, but it is less likely they will be attacked on gendered grounds (Duggan, 2014, 2017). That is, unless men are seen as failing masculine prescriptions within male-typed domains, in which case they are expected to be affixed with DGLs. Thus, Study 5 will assess whether the frequency of DGLs and non-gendered insults differ by target gender. Study 5 also assesses tweet valence, to determine whether this differs by target gender.

In addition, factors other than target gender could influence who is affixed with DGLs. In the 2016 presidential primary elections, there was an unusually large number of Republican candidates (17), but only six Democratic candidates. Debates are zero-sum situations (only one candidate "wins"); the goal of the debate is to "damage and

dominate" opponents (Garcia-Pastor, 2008). In a primary campaign marked by contention (Cummins, 2016), the larger number of GOP competitors may have increased debate hostility that spilled over to moderators. Further, according to Hinck et al. (2017), in the 2016 primary debates, GOP moderators asked more aggressive questions than did Democratic moderators; in GOP debates, questions also increased in hostility from 2012 to 2016. Given that debate moderators contributed to the negative tone of the GOP primary, one might expect that negative, insulting attacks will be more common in GOP debates.

Finally, there is a possibility that DGLs will target certain moderators. For example, Hinck et al. (2017) found that levels of debate hostility varied greatly across the 2016 GOP debates, with the March 3, 2016 debate, moderated by Megyn Kelly and Bret Baier, featuring a particularly high number of direct, personal attacks. In the first Republican debate, Megyn Kelly pointedly asked Trump about his treatment of women:

> "Mr. Trump, one of the things people love about you is you speak your mind and you don't use a politician's filter. However, that is not without its downsides, in particular, when it comes to women. You've called women you don't like "fat pigs, dogs, slobs, and disgusting animals"…Your Twitter account has several disparaging comments about women's looks. You once told a contestant on Celebrity Apprentice it would be a pretty picture to see her on her knees. Does that sound to you like the temperament of a man we should elect as president, and how will you answer the charge from Hillary Clinton, who was likely to be the Democratic nominee, that you are part of the war on women?"

This question marked the beginning of a feud between Trump and Kelly. Trump later implied that Kelly asked aggressive questions in this debate because she was on her period and refused to attend a subsequent debate hosted by Fox News because Kelly was a moderator (Borchers, 2016). On Twitter, Kelly became one of Trump's favorite targets, and he posted tweets that referred to her as "crazy" and "a bimbo" (Lee & Quealy, 2016). Thus, Trump supporters, in particular those who share his views on women, might use more DGLs when tweeting about Megyn Kelly compared to other female moderators.

What labels are used? Labels that ascribe femininity, homosexuality, or weakness are expected to target atypical men (e.g., 'pussy,' 'faggot'). Female targets will likely be subjected to several categories of labels, particularly those that stigmatize gender deviant women as overaggressive (e.g., 'bitch,' 'dyke'), as well as those that degrade women's sexuality (e.g., 'slut,' 'whore'). Furthermore, in keeping with Studies 1-4, I expect that derogatory gendered labels will be accompanied by language that promotes stereotype-based attributions for target behavior.

Who does the labeling? In Studies 1 & 2, participant gender moderated effects of the label, such that men, but not women, decreased evaluations of female targets in response to the label. In Study 5, the gender of Twitter user will be examined as a predictor of the likelihood of using a DGL. Women, compared to men and users who do not provide enough information to be classified according to gender, are less expected to send tweets containing DGLs.

Study 5 Method

Big data offers access to large, rich datasets uncontaminated by experimental reactivity, but this comes at the cost of greater noise in the data (Salganik, 2018). As a

result, Twitter data must undergo substantial preprocessing before they are ready for analysis. First, I searched for tweets containing the names of debate moderators, and then I purged this dataset of duplicate tweets and tweets from bots. Once I cleaned the dataset, I prepared the tweets for textual analysis by stripping them of non-word characters and frequent words (e.g., "debate"; "MSNBC"). A random subset of the total tweets was submitted to raters for annotation.[9] Coders identified the relevance and target of each tweet, and they rated each tweet on valence. Tweets on which coders could not agree were omitted from the dataset. Eighteen percent of tweets were excluded on this basis. Finally, the dictionary I created was used to classify the remaining pool of tweets. Below I describe my steps in collecting, cleaning, and annotating the Study 5 data.

Data Collection

Public tweets containing the name or handle of 12 different debate moderators were collected for the day of and the day after the debate (see Table 3 for a list of the moderators) using Sifter (2018), an online tool through which researchers can purchase access to historical Twitter data. These search terms resulted in a total of 292,691 tweets.

Data Cleaning

Social media bots are automated accounts that mimic human social media behavior. Detection of bots poses a challenge, because today's bots are quite advanced (e.g., capable of participating in complex interactions with other social media users; Ferrara, Varol, Davis, Menczer, & Flammini, 2015). Therefore, it is important that bots

[9] I intended to use these ratings to train a classifier to recognize DGLs. However, raters performed poorly at identifying DGLs in tweets. Thus, analyses were conducted on this random sample.

be identified and removed prior to data analysis, especially given the prevalence of bots on Twitter during the 2016 presidential election.[10] Bots were identified using BotorNot (Davis, Varol, Ferrara, Flammini, & Menczer, 2016), which uses a machine classifier to return a bot-likelihood score for individual Twitter users. Due to Twitter's streaming application programming interface (API) limits (only 180 requests every 15 minutes are permitted), a time-loop was used to retrieve the data. The data were retrieved and collated using the R/Python Interface (rPython; Bellosta, 2014). Following the threshold recommended by Davis et al. (2016), Twitter users scoring .50 or above were removed from the dataset (n = 6867).[11]

Further data cleaning procedures included the removal of exact duplicates, moderators' tweets, non-English tweets, tweets containing no analyzable content (e.g., images), and tweets that did not reference a moderator. Finally, tweets were stripped of special characters, numbers, punctuation, URLs, stopwords, hashtags, mentions, and common words.

Data Annotation

With such a large dataset, it was not feasible to inspect each tweet and verify that it was classified correctly according to target gender or to rate its valence. Therefore, I selected a random sample, stratified by debate, to be annotated by coders on Figure Eight (previously known as CrowdFlower), an online crowdsourcing service. Similar to Mturk, Figure Eight gives researchers access to a pool of workers who complete tasks (e.g., text classification) for small sums of money.

[10] Bessi & Ferrara (2016) estimate that as many as 15% of Twitter users tweeting about the election were bots.

[11] Number of Twitter users removed from the random stratified subsample selected for further analysis. Because of the length of time it took to run this procedure, it was not feasible to run on the entire sample.

Out of the remaining tweets (n = 122,430), approximately 16,000 tweets were coded by 464 paid Crowdflower participants. Coders rated each tweet on several dimensions, four of which were used in this study: relevance to the moderators, target gender, and affect expressed toward the target of the tweet (positive and negative). Coders received detailed annotation guidelines before they begin classifying the data (see Appendix A).

Throughout the task, workers were required to maintain 70% accuracy on a series of test questions provided by the requester. Coders who incorrectly answered these questions received feedback about their error and were informed of the correct response. If their performance dropped below 70%, they were removed from the study. Figure Eight workers were recruited until each tweet was coded by at least 3 workers (*M* number of coders = 5). On average, raters rated 177 tweets.

Tweet relevance. Many of the tweets discussed the political candidates or the debate itself, rather than the moderators. To ensure that selected tweets actually targeted the moderators, participants were asked to indicate whether the tweet was relevant. Relevant was defined as "tweets that express an opinion about, describe, or talk to one or both of the moderators." If participants were unsure of whether the tweet was relevant, they were instructed to select "undecided." Tweets deemed irrelevant (n = 11413) were omitted from the dataset, as were those for which the majority of raters were undecided (n = 39). Figure Eight calculates confidence scores for each tweet, which are based on level of agreement and workers' task accuracy. Across tweets, the average confidence score for relevance was .83 (range is 0 to 1).

Target gender. Participants were also instructed to indicate which moderator(s) each tweet referred to (female moderator only, male moderator only, both moderators, or undecided). The average confidence score for target gender was .81.

Valence. Finally, participants separately rated each tweet according to positive and negative valence. Participants were instructed that they should rate the tweet as positive (or negative) when tweets expressed a positive (negative) opinion or attitude towards the moderator(s) and/or their moderating performance. Tweets were rated on two scales ranging from 0 (not at all) to 3 (strongly). Average reliability across raters was good (Positive $_{Male}$ = .90; Positive $_{Female}$ = .82; Negative $_{Male}$ = .81; Negative $_{Female}$ = .79). For data analysis, these two measures were combined into an overall valence measure, ranging from -6 (more negative than positive) to 6 (more positive than negative), with 0 indicating either neutral or equally positive and negative valence.

We removed all tweets with less than 60% percent agreement (n = 8416). This resulted in a final sample of 6085 tweets.

LIWC Classifications

Linguistic Inquiry and Word Count (LIWC) is a text analysis program developed by Pennebaker, Booth, and Francis (2007) that analyzes and classifies text on a word-by-word basis. LIWC was used to categorized tweets according to the dictionary classifications developed in the previous study. After classification, the author read each tweet, to determine whether the contextual usage of the label matched the dictionary classification; if it did not, the tweet was removed from the category. Tweets containing the word “bitch” were classified to either the Feminine / Weak or Masculine / Aggressive category, based on the contextual meaning of the word. When applied to women, the

usage of "bitch" was more consistent with the latter category (e.g., "MEGAN KELLEY is a 1st class bitch who is only looking for career. Some of Us care. Son in Army, Hubby in AF"); however, when applied to men, its usage was more consistent with the former category (e.g., "Trump is now self-moderating the #GOPDebate, Cavuto is his bottom bitch").

Before analyses, each of these variables was coded into binary categories (0 = absence of label; 1 = presence of label). In addition, tweets were classified according to verbal abstractness.

Internal attribution. The Linguistic Category Model (LCM; Semin & Fiedler,1991) describes four word classes that vary in abstractness (ranging from descriptive action verb, the most concrete, to adjectives, the most abstract). The more abstract a word, the more it conveys dispositional meaning. Thus, the LCM provides a way of scoring tweets based on the extent to which they include dispositional language. Words within each tweet will be classified using the General Inquirer's list of LCM words (Stone, 2002). Then a mean abstractness score will be computed for each tweet following the procedure detailed in (Semin & Fiedler, 1992).

Other Variables

User gender. User gender was determined via a combination of methods. First, using the gender package in R (Mullen, Blevins, & Schmidt, 2015), the real first names of Twitter users were compared to a dataset containing user profiles from social media sites. This information was used to return a gender prediction and probability score. Second, user profiles that contained gender indicators (e.g., "*mother* of two"; "proud *dad*") were classified according to the appropriate category. Third, for the remaining

individuals, user profile images were downloaded via the Twitter API and classified by the author. Users that could not be classified by any of these methods were assigned to an unknown gender category.

Analytic Strategy

Individuals can tweet an unlimited number of times about each debate, which introduces dependence into the data. However, individuals can also tweet only once about a single debate. Thus, tweets were nested both within persons and within debate. Multilevel models account for dependence between observations; they also handle unbalanced data better than other repeated-measures approaches (Cohen, Cohen, West, & Aiken, 2003). Given the structure of the data, data were initially analyzed in a cross-classified multilevel model, with tweets nested within individuals and debate. However, in most cases, models failed to converge with the inclusion of a random effect for debate; hence, this term was included only when it accounted for a significant portion of outcome variance.

Tweets were the unit of analysis. Level 2 variables included political party of debate (0 = Democratic; 1 = Republican), labeler gender (two dummies with women set as the reference group), and target gender (0 = Male or both Male and Female; 1 = Female). Dependent variables were the presence or absence of DGLs and general insults, valence, and verbal abstractness. Results are reported for fixed effects only.

Study 5 Results

Derogatory Gendered Labels

First, I examined whether the likelihood of using gendered derogatory labels differed by political party of debate, labeler gender, or target gender. Tweets containing

DGLs made up a small proportion of the overall set of tweets (2.3%). When they were used, DGLs were used more frequently for female, relative to male, targets, but only in GOP debates (Female Target x GOP: $b = 1.82$, $p < .05$, OR = 3.10). However, further inspection suggested that this finding was due to a disproportionate amount of DGLs directed at Megyn Kelly (see Table 4). Male labelers were marginally less likely than female labelers to use DGLs ($b = -1.12$, $p = .07$) in Democratic debates, whereas they did not differ in GOP debates (Male Labeler x GOP: $b = 1.82$, $p < .05$). Finally, a marginally significant three-way interaction between target gender, political party, and labeler gender was observed ($b = -2.14$, $p = .06$). Two-way interactions between target gender and labeler gender were in opposite directions for GOP ($b = -.65$, $p = .35$) and Democratic debates ($b = 1.39$, $p = .10$), but neither were significant. However, the pattern of results suggests that when tweeting about Democratic debates, male, relative to female labelers, tended to use DGLs less for male targets. See Figure 25 for a depiction of the specific DGLs used in the dataset.

General Derogatory Labels

In the interest of comparison with DGLs, analyses predicting the occurrence of general insults were also conducted. Only one significant relationship emerged: general insults were 3.7 times more likely in GOP than Democratic debates ($b = 1.31$, $p < .05$).

Specific Labels Used[12]

[12] The dictionary developed for this study also included an Ethnic category; however, only one tweet (directed at Carl Quintanilla) contained words for this category.

Sexual. Target gender and political party perfectly predicted the usage of sexual labels, with only female moderators of GOP debates receiving these labels; thus, the binary logistic GLMM model was inappropriate for analyses. Megyn Kelly was the main target of sexual labels (n = 25). Sexual labels were also used twice for Becky Quick, another GOP moderator. The most frequently used terms were "whore" (n = 10) and "bimbo" (n = 8), and male labelers provided the majority of labels (n = 15). Sexual labels were occasionally accompanied by references to sexual behavior (e.g., "megyn kelly is whore who slept her way to the top with blood hound brit hume"), but most were not. The majority of sexual labels were seemingly applied to women perceived as engaging in behavior inappropriate for a woman (e.g., "YOUR KIDDING, MEGYN CRAWLEY BIMBO KELLY PROFESSIONAL"). In all cases, these terms were deployed in ways that undercut target credibility, and in many, they implied the target was disgusting or dirty.

Homophobic. Few tweets (n = 9) contained a homophobic label, so no analyses were conducted. Two homophobic labels were used for male targets, and seven were used for female targets. Men and women were similarly likely to use these labels, and label occurrence did not differ by political party. "Gay" and "dike/dyke" were the most frequently used homophobic labels (n = 2), and they were used to infer a target's sexual orientation based on stereotypes (e.g., "David Muir is so hot (debate moderator). Proving again what they say about all the good ones. ...gay :("), or to insult a target's femininity (e.g., "Megyn Kelly is a dike").

Incompetent. Incompetence labels were more common in GOP debates, relative to Democratic debates (b = .74, p < .05, OR = 2.10). However, in contrast to

expectations, incompetence labels did not differ by target gender or labeler gender. The highest occurrence of these terms was in Debate 5, with Becky Quick and Carl Quintanilla (see Table 5), and the most frequently used terms were "stupid" and "idiot." In contrast to the two previous types of labels, incompetence labels often described questions asked in the debate, rather than the targets themselves (e.g., "Gotta give it to @tedcruz for calling @carlquintanilla out for asking stupid ass questions. Com'On Carl! Not very professional"). In addition, incompetence labels were more evenly spread across debates than most other types of labels.

Feminine / Weak. Feminine labels were more frequently used for male and both male and female targets than for female targets alone (b = -1.40, p < .05, OR = .25; n_{Male} = 18; n_{Both} = 4; n_{Female} = 13); their usage did not differ by political party or labeler gender. "Bitch" was the most frequently used feminine DGL (n = 12), and like other terms in this category, when used for men, it implied that they were insufficiently assertive or masculine (e.g. "LOL @carlquintanilla just got bitch slapped by the audience! #CNBCGOPDEBATE"). When used for women, words in this category were often group labels used in a derogatory sense; for example, the diminutive "girl" (e.g., @megynkelly U should truly be ashamed! U have stooped to a new low. Does yur deceit make u feel like a big girl?) or "woman" in place of a moderator's name ("@BeckyQuick you should be relegated to THE KITCHEN. FIX ME A SAMMICH WOMAN!!").

Masculine / Aggressive. Masculine labels appeared more frequently in GOP, relative to Democratic debates, and particularly, in Debates 5 & 6 (b = 1.59, p < .01, OR = 4.90; see Table 6). "Bitch" was the most frequently used term (n = 24), followed by "witch" (n = 10); in this category, these labels were used exclusively for women. In

contrast, “asshole” (n = 6) and “douchebag” (n = 5) referred to male targets. Although female-referential and male-referential labels in this category possess similar meanings (e.g., mistreats others), labels used for women also implied overaggressiveness or overambitiousness (e.g., “@BeckyQuick I must say you were one snarky bitch the other night. Hope thats not you all the time. FYI bad timing to pull that crap. Cheers!”; “Megan Kelly is a pompous, headline-seeking, self-aggrandizing bitch”). In addition, several tweets insulted female moderators’ femininity by using gender-inappropriate category labels (e.g., “Is Maddow the 14 year old boy they allowed to ask questions?”).

Physical appearance. Despite the low rater reliability for this category, I examined whether physical slurs differed by the variables of interest. A two-way interaction between female target and political party (b = 2.61, p < .01) indicated that female moderators received more physical labels than male moderators in GOP debates (b = 1.54, p < .01; n_{Male} = 10; n_{Both} = 2; n_{Female} = 41). The majority of these labels were gendered (unlike those targeting men), and they targeted Megyn Kelly (see Table 7). The most frequently used terms were “witch” and “ugly.” Men were labeled “fat” more often than female targets. This category also included a few relatively positive but objectifying labels (e.g., “babe”) that referred primarily to women. Due to poor reliability, results for this category must be interpreted with caution; however, they are consistent with results for other categories.

Valence

I conducted LMM analyses for the remaining variables, since they were collected on a continuous scale. Overall, tweets in GOP debates were more negative than tweets in Democratic debates (b = -.98, p < .001). Political party of debate modified the effect of

target gender on tweet valence (b = .56, p < .01). Further, there was a three-way interaction (Female Target x GOP x Other-gender Labeler: b = .70, p < .05). In Democratic debates, tweets about female targets were more negative than those about male or both targets (b = -.28, p < .05). Also, male (b = -.47, p < .001) and gender-undetermined labelers (b = -.36, p = .05) provided more negative tweets than female labelers.

In GOP debates, gender-undetermined (b = -.84, p < .001) posted more negative tweets about male or both targets, relative to female targets. Figures 26 a and b depicts these results.

The above analyses examined predictors of valence in the total corpus of tweets. Additional analyses were conducted to assess whether tweets containing DGLs and general insults differed in valence from other tweets. Because there was a large difference in the number of tweets across these groups, I drew a random sample of tweets without DGLs or general insults for comparison. This analysis showed that tweets containing DGLs (b = -1.24, p < .05) and general insults (b = -1.66, p < .001) were significantly more negative than other tweets, but they did not differ in valence (p = .41).

Internal Attribution

Finally, tweets about GOP debates contained more abstract language than Democratic debates (b = .05, p < .05, OR = 1.05). However, this depended on target gender (Female Target x GOP: b = .06, p < .05). For female targets, there was a larger increase in abstract language in GOP relative to Democratic debates (b = .11, p < .001, OR = 1.12). Within GOP debates, tweets about female targets were more abstract than those about male targets (p < .05).

Again, additional analyses were conducted to assess whether tweets containing DGLs and general insults differed from other tweets. Tweets containing DGLs were more abstract than tweets without DGLs (b = .45, p < .01) or general insults (b = .51, p < .01). However, the relationship between tweet type and abstractness was modified by target gender (Female x DGL: -.55, p < .05; Female x General Insult: -.42, p < .05). Interestingly, tweets targeting men (or both moderators) were more verbally abstract than those targeting women when they contained DGLs or general insults (p's < .05). Across tweets, those with DGLs tended to be more abstract than those without, with the exception that for female targets, non-insulting tweets were similarly abstract. See Figure 27.

Study 5 Discussion

Study 5 conducted an exploratory study examining factors that influence the likelihood of using DGLs. The discussion is organized by the questions posed in the introduction to this study:

Who gets labeled?

As expected, target gender affected the likelihood of DGLs, such that women, relative to men, were more frequent targets. However, this was true of GOP debates only. The prediction that male moderators would receive more general insults than female moderators was not supported. However, in GOP debates, tweets about men were more negative than those about women.

Results highly depended on the political party of the debate. Study 5 found that general insults, negative tweets, and incompetence labels occurred more frequently in GOP debates. Also, as mentioned above, in Republican debates, DGLs were used more

frequently for female, relative to male, targets. These findings speak to a greater level of debate hostility, which is consistent with Hinck et al.'s findings (2017) of less civility in the 2016 GOP, relative to Democratic, debates.

Also consistent with Hinck et al. (2017), the occurrence of negative language varied widely across debates. For instance, incompetence labels occurred most frequently in the much-criticized CNBC Republican debate. After this debate, the candidates complained that the moderators asked non-substantive "gotcha" questions, and the RNC cancelled a later debate scheduled with NBC (Byers, 2016; Zezima, 2015). Carl Quintanilla, a moderator of this debate, received more DGLs than other male targets, the majority of which were Masculine/Aggressive or Feminine/Weak labels. He was also the only moderator who received a derogatory ethnic label. Unlike the other GOP moderators, who appear to be white, Carl Quintanilla appears to be multi-ethnic. These findings are consistent with the idea that, in addition to women, DGLs also stigmatize subordinate group men (Connell & Messerschmidt, 2005; Schur, 1983). However, it is impossible to determine whether Quintanilla received more negative labels because of his ethnicity or his debate performance. It would be interesting to know whether the third moderator, John Harwood, who is white, was similarly labeled. However, I did not collect this data.

By publicly challenging Trump on his treatment of women, Megyn became a target of harsh backlash. Overall, Megyn Kelly was the most frequent target of DGLs, and in particular, Sexual labels. A closer examination reveals that the majority of these tweets either mentioned Trump or were sent by Trump supporters[13] (54% compared to

[13] Trump supporters were defined as those who exhibited support for Trump in their user bios, usernames, or posts (e.g., #MAGA2016).

33% of tweets about GOP debates that contained DGLs). Interestingly, some of these tweets contained the same labels previously used by Trump (e.g., "bimbo"). Undoubtedly, some of Trump's supporters share his sexist views; it is most likely these individuals who authored these tweets. These results highlight the role of political elites in the discursive reproduction of sexism (van Dijk, 1987).

Across variables, it is likely that sexism accounts to some degree for the differences observed between GOP and Democratic debates. Since I examined the presidential primary, rather than general election, debates, it is likely that the majority of people were tweeting about debates hosted by their political party. As conservatism is correlated with sexism (Duckitt & Sibley, 2009; Sibley, Wilson, & Duckitt, 2007), it may be acting as its proxy in this study. However, this is impossible to confirm.

What labels are used?

In line with expectations, Sexual DGLs targeted women only and were most often used by male tweeters. The most frequently used Sexual labels cast women as inappropriately sexual. Although, on the basis of backlash theory, I expected female moderators to receive more Masculine/Aggressive DGLs, there was no difference between male and female targets. Arguably, though, the specific labels used implied different meanings depending on target gender. When used for a woman, Masculine/Aggressive labels seemed to imply that she was behaving overly aggressive for a woman. When used for a man, these terms, while still derogatory, did not carry quite the same meaning. As expected, men, relative to women, were more frequently attached with Feminine/Weak labels. Taken together, these results support the assertion that DGLs are deployed to disparage people perceived as violating gender norms.

In addition, tweets containing DGLs implied dispositional attributions to a greater degree than did other tweets. Notably, verbal abstractness is distinct from valence; thus, highly abstract tweets not containing DGLs or general insults could be positive. However, it is reasonable to assume that tweets containing DGLs and general insults were relatively negative, Thus, overall, this finding support the prediction that tweets containing DGLs would be accompanied by more dispositional language.

Who does the labeling?

Although I predicted that men and gender-nonspecified individuals would be more likely than women to use DGLs, this prediction was largely unsupported. One main exception was the use of Sexual DGLs, which were provided primarily by men. Beyond this, in GOP debates, gender-nonspecified tweeters posted more tweets that were negative towards male targets. Accordingly, this study does not shed much light on user factors that predict the likelihood of using DGLs.

Finally, it is important to note that DGLs made up a small proportion of the total tweets. Because I collected data retrospectively, it is likely that some of the most offensive tweets had already been deleted. Indeed, even in this dataset, a number of posts were unavailable for viewing because the user had been suspended, or the original tweet or account deleted. However, even if DGLs occur infrequently, the other studies reported here indicate they are impactful. Moreover, studying DGLs is of particular import in the current sociopolitical environment, in which they (and other prejudiced speech) are being used by politicians in power to shape public discourse.

STUDY 6

Study 5 examined how DGLs were deployed in political conversations on Twitter. Study 6 examines whether the findings of Part I can be replicated in a different context, with actual tweets about the 2016 election. In Part I, participant gender and sexist ideology influenced responses to DGLs, with men high in hostile sexism most likely to derogate a labeled female target. Thus, Study 6 collects men's ratings of tweet targets and authors of tweets which either contain a DGL or not. I expect that, for men high in hostile sexism, the DGL will decrease evaluations of the (female) target and increase evaluations of the author.

Study 6 Method

Participants

Studies 1, 2, and 4 collected ambivalent sexism scores. Rather than collect these scores again, I sent an invitation to participate in Study 6 to 382 men who participated in these initial studies. Prospective participants were told that they qualified for a study on first impressions. 44.7 % (N = 171) completed the study. Participants' prior HS and BS scores were collated and combined with the data collected in Study 6.

Stimuli

Study 6 participants were asked to rate three types of tweets (DGL, Negative, and Positive) in a repeated measures design. Six tweets, one for each female moderator, were selected for each condition. In selecting tweets with DGLs, I avoided tweets with strong negative labels, to minimize participant suspicion (see Figure 28 for example tweets). DGL tweets did not differ from Negative tweets or Positive tweets on word count, F (2, 15) = .81, p = .46. Nor did they differ from Negative or Positive tweets on verbal

abstractness, F (2, 15) = .49, p = .62. Positive tweets were significantly more positively-valenced than Negative and DGL tweets (F (2, 15) = 5.08, p < .05), and significantly less negatively-valenced than Negative and DGL tweets (F (2, 15) = 8.64, p < .01), which did not differ from each other. See Table 8 for means.

Participants rated all tweets on target and labeler (i.e., tweeter) likability,[14] on the same 11-point semantic differential scale used in Studies 1-4.

Study 6 Results

Multilevel analyses, with tweets nested within participants and targets, were conducted on target and labeler likability. Results are reported for fixed effects only. Predictions are that participants high in hostile sexism will provide more negative evaluations of targets and more positive evaluations of labelers when tweets contain a DGL, relative to other conditions. Thus, I expect two-way interactions between HS and tweet type in both regressions.

Target Rating

Likability. As expected, there was a two-way interaction between hostile sexism and the dummy indicator for positive tweets (b = .27, p < .001), such that HS negatively predicted target likability when tweets contained a DGL (b = -.38, p < .001) but not when they were positive. However, in contrast to expectations, HS negatively predicted target likability in the negative condition as well (see Figure 29).

Labeler Rating

Likability. In line with expectations, there was a significant HS x Positive interaction (b = -.46, p < .001); HS positively predicted labeler likability in the DGL (b =

[14] To minimize the length of the study, I did not collect competence ratings.

.35, $p < .001$), but not positive, condition. The HS x Negative interaction was marginal ($b = -.13$, $p = .07$), indicating that the relationship between HS and labeler likability was less strongly positive in the negative, relative to the DGL, condition ($b = .21$, $p < .01$; see Figure 30).

Study 6 Summary

I predicted that people high in hostile sexism would be particularly approving of the labeler and disparaging of the target in the derogatory label condition. Results were consistent with this pattern; however, they also held for the negative condition. Hostile sexism was unrelated to target and labeler likability when tweets were positive towards female targets, but regardless of the type of negative tweet, participants high in hostile sexism derogated the female target (and rewarded the labeler).

In Part I, I hypothesized that DGLs affect target perceptions by facilitating stereotype-based impressions of targets, justifying backlash, and releasing prejudice. Although I did not compare DGLs to other negative labels, on the basis of pilot studies, I assumed that these mechanisms were unique (or at least particularly important) to DGLs. However, it is possible that people who harbor hostile attitudes toward women will seize on any opportunity to justify expressing that hostility. If this is true, then individuals high in hostile sexism should derogate female targets in response to any negative tweet.

In Study 6, I equated both types of negative tweets on verbal abstractness. However, the results of Part I suggest that DGLs impact evaluations *because* they imply internal attributions. Rather than negativity per se, sexists may be responding to negative dispositional language. Therefore, I conducted the study again, this time varying the extent to which tweets conveyed dispositional meanings.

STUDY 7

In Study 6, both negative tweets with and without a DGL similarly negatively affected target and labeler evaluations. However, each set of tweets was matched on verbal abstractness. As discussed before, more verbally abstract language implies stronger dispositional meanings. For example, "Amy is aggressive" conveys information about Amy's disposition, whereas "Amy hit John" conveys information about a concrete action performed in a particular situation (Semin & Fiedler, 1991). By varying the verbal abstractness of tweets, Study 7 will manipulate the mediator tested in Part I. If individuals high in hostile sexism are responding to language that implies negative attributional meanings, rather than DGLs per se, then HS should predict likability when tweets contain a DGL or are highly abstract, but not when they are less abstract.

Study 7 Method

Participants

A total of 300 men were recruited via Turk Prime's panel feature. Other demographic information was not collected.

Stimuli

Study 7 had the same design as Study 6, except that positive tweets were replaced with negative tweets with low verbal abstractness (see Figure 31 for an example tweet). DGL tweets did not differ from Negative-High Abstract or Negative-Low Abstract tweets on word count, F (2, 15) = 1.48, p = .25, positive valence (F (2, 15) = 1.07, p = .37), or negative valence (F (2, 15) = .19, p = .83). Negative-Low Abstract tweets were significantly less abstract compared to DGL and Negative-High Abstract tweets, F (2, 15) = 9.95, p < .01. See Table 9 for means.

One day after initial data collection, participants who completed the study were invited to participate in a follow-up study on first impressions. As before, sexism scores were collected in this follow-up study. 63% completed the study and passed an attention check, resulting in a final sample of 190.

Study 7 Results

Target Rating

Likability. In contrast to expectations, the HS x Negative-Low Abstractness interaction was not significant. Rather, there were main effects of condition and hostile sexism. Target likability was higher in the negative, low abstractness condition, relative to the DGL (b = .20, p < .01) and negative, high abstractness conditions (b = .15, p < .05). Regardless of condition, hostile sexism negatively predicted target likability (b = -.33, p < .01). In addition, participants low in HS rated the labeler as least likable in the DGL condition. Figure 32 displays these results.

Labeler Rating

Likability. The regression yielded the predicted two-way interactions between HS and both negative conditions (HS x Negative-Low Abstractness: b = -.29, p <.001; HS x Negative-High Abstractness: b = -.14, p < .05). HS positively predicted labeler likability in all conditions, but this relationship was most pronounced in the DGL condition (see Figure 33).

Study 7 Summary

Study 7 found that abstractness negatively affected the likability of female targets, regardless of participants' levels of sexism. The more verbally abstract the tweet, the less likable targets were rated. In contrast, across tweets, hostile sexism predicted labeler

likability. This finding seems to suggest that individuals high in hostile sexism provide more positive ratings of people who disparage women, regardless of the genderedness or abstractness of the message. However, the difference between people low in HS and high in HS was most pronounced when the tweet contained a DGL, suggesting that sexism influenced responses to this type of tweet.

As discussed previously, a possible explanation for why Part I results did not replicate in this new context, is that the DGLs used in this study did not pertain to emotional behavior. A core feature of gender-emotion stereotypes is that they carry attributional information; inherent emotionality is deeply associated with femininity (Shields, 2002). It is possible that other types of DGLs do not affect perceptions of women the same way. Future research should examine a variety of DGLs to assess whether this is the case.

GENERAL DISCUSSION

This dissertation employed two methodological approaches to examine the content, targeting, and evaluative consequences of derogatory gendered labels. Although, results differed across the seven studies presented in this dissertation, several broad conclusions can be drawn: First, sexism shapes responses to DGLs. Individuals who endorse sexist views are more likely than those who do not to interpret DGLs in ways that are harmful to perceptions of women. Moreover, they are more likely to reward others when they use DGLs. Second, consistent with prior theorizing (e.g., (Schur, 1983), DGLs are used to enforce gender norms; however, though labels are used for men, they appear to be 'stickier' for women. Finally, results were generally supportive of a labeling theory perspective that, to the extent that labels are seen as ascribing essential character,

they indelibly mark individuals. Thereby, the present studies contribute to our understanding of how derogatory gendered labels affect social perceptions.

A second contribution of this dissertation is that it presents an example of how one might combine experimental methods with social media analysis. Combining methodological approaches can yield important insights into phenomena not available through the use of a single method (Fine & Elsbach, 2000). Together the studies conducted in Part I and II contribute to a more complete understanding of DGLs.

Role of Sexism

Across studies, ideological factors were consistently important. Results suggest that, for people predisposed to view women negatively, DGLs "stick" to and harm female, more than male, targets. Thus, overheard DGLs most powerfully shape perceptions for people seeking to confirm their sexist views. Furthermore, labels do not only affect targets but also labelers; sexist individuals respond positively to individuals who use DGLs. This finding is consistent with the argument that, in the 2016 election, sexism contributed to support for Trump (Setzler & Yanus, 2018). Sexist individuals are also more frequent users of DGLs themselves, particularly when elites signal the social acceptability of such language (Crandall & White, 2016; van Dijk, 1987).

A feedback loop may exist in which individuals high in sexism seek out exposure to derogatory sexist language, which in turn, feeds or amplifies pre-existing beliefs (Slater, 2007), dynamics which may be only heightened on social media. Through social media outlets, everyday individuals can produce information that reaches a wide number of people. One no longer needs to be the most prominent or strident voice to be heard; for better or worse, the voice of a relative few can disproportionately impact public

discourse. In addition to amplifying misogynistic extremism, advances in online technology have made it easier to engage in biased information search (Sunstein, 2007). Individuals high in sexism are likely to seek out sexist content, and as the present research suggests, are also likely to be more impacted by it (Slater, 2007). It is not difficult to imagine how this has contributed to, in certain corners of the internet, the development of insular echo chambers for sexist sentiment.

Usage of DGLs

Comparing DGLs to other negative labels, there were differences in meaning. General insults seemed to fall into three primary categories: a) intelligence (e.g., "dumb", "moron"); b) appearance (e.g., "fat"; "ugly"); and c) mistreats others (e.g., "jerk"; "scumbag"). In contrast, the majority of DGLs centered on enforcing gender prescriptions. For example, some DGLs also fell under the last category, but primarily male-referential terms were closer in meaning to general terms (e.g., "asshole"; "bastard"), while primarily female-referential terms ascribed gender-proscribed traits (e.g., "bitch"; "ballbuster"). Moreover, DGLs tended to be used to disparage targets perceived as violating gender norms. These findings highlight the stereotype-maintaining function of derogatory gendered labels.

Internal Attributions

Interestingly, DGLs appeared to be 'stickier' for women, whereas men were often viewed positively in spite of the labels applied to them. The present studies underscore internal attributions as a crucial mechanism underlying this difference. To the extent that DGLs implied dispositional meanings, they negatively affected perceptions of women. Contrastingly, discounting internal causes of DGLs lead to more positive perceptions of

male targets. Prejudice influences language use and interpretation; sexist individuals may both communicate and interpret DGLs in abstract ways that align with stereotypical expectations (Maass & Arcuri, 1996). Further, by interpreting the negative label as evidence of underlying personality characteristics, sexist individuals can justify misogynistic responses (Crandall & Eshleman, 2003).

Notably, findings differed across the two methodological approaches. Of particular interest are the discrepancies between Studies 1-3 and Studies 6-7. Studies 1-3 found that internal attributions mediated the relationship between the DGL and target evaluations. In contrast, Studies 6-7 found that negative, abstract tweets, regardless of whether or not they contained a DGL, decreased target likability. The latter studies raise questions about whether negative dispositional language, rather than DGLs per se, accounts for the effects on perceptions of targets. Differences across these studies, including whether target behavior is observed and whether this behavior is relevant to gender-emotion stereotypes, limit the ability to draw firm conclusions. It is also worth noting the difficulty in identifying underlying mechanisms in big data (Shriffin, 2016). However, even if any sufficiently abstract, negative language produces the target-labeler patterns observed, Study 5 suggests that, at least in the context of Twitter, message containing DGLs tend to be couched in more abstract terms than other types of messages.

Confirmation Bias – A Better Framework for Understanding DGLs?

It could be reasonably argued that motivated confirmation bias (Nickerson, 1998) explains many of the findings in this dissertation. Perceivers who were motivated to defend the gender hierarchy selectively interpreted DGLs to be consistent with their pre-existing beliefs. That is, they attributed labeled women's behavior to inherent femininity,

whereas they attributed labeled men's behavior to situational factors (Barrett & Bliss-Moreau, 2009; Brescoll & Uhlmann, 2008; Shields and Crowley, 1996). Thus, stereotypic expectancies function as "hypotheses about the likely disposition of a person" (Darley & Gross, 1983), and DGLs are taken as corroborating evidence when they match expectancies but dismissed as puzzling exceptions when they do not. In line with H3, justification-suppression theory (Crandall & Eshleman, 2003) offers an equally plausible and not necessarily incompatible explanation for the findings: DGLs are selectively interpreted to justify the expression of suppressed hostility toward women. The precise motivation for derogating female targets, whether to defend closely-held beliefs or to relieve tension caused by prejudice suppression, in this case, leads to the same outcome.

In understanding DGLs, it is important to distinguish between the ways in which they are used (for stereotype violators) and how they affect target perceptions (reinforce stereotype-based impressions). A cognitive bias perspective does not address how DGLs are deployed to enforce gender norms, and in doing so, selectively punish certain types of women (Ashwell, 2016). DGLs recast stereotype violators in narrow, stereotypic terms in order to demarcate and enforce gender boundaries (Ussher, 1992). Accordingly, backlash theory (Rudman et al., 2012) remains a useful framework for considering DGLs and, more broadly, derogatory groups labels. Future work might consider how derogatory labels divide other groups into good-compliant and bad-threatening subgroups.

Future Directions

The results of this dissertation suggest a variety of possible future directions. First, future research should determine whether, as suggested by Study 7, highly abstract negative language that is not gendered also leads to more negative target evaluations.

Second, another possible explanation suggested by Studies 6 & 7 is that people high in trait hostility, rather than hostile sexism per se, respond positively to negative language. In order to assess these possibilities, follow-up studies should be conducted that vary the language used and target gender. Third, in Study 5, data were collected retrospectively. A drawback of collecting historical Twitter data is that it may not capture the most offensive comments, which could have already been deleted. In addition, in the time since the tweet was first published, accounts may be suspended or deleted, making them inaccessible. Future studies of biased language on Twitter should be aware of this limitation and if possible, capture tweets as they occur. Finally, a valuable feature of Twitter data is that it is possible to use retweets to assess the spread of a message. When Twitter users retweet a message, they are reposting a message written by another person to be seen by their followers. Another interesting extension of this research would be to identify influential spreaders of DGLs and to examine how quickly and broadly DGLs spread through online social networks relative to other types of messages.

Conclusion

Relative to subordinate social groups, dominant groups have more power to shape discourse in line with ingroup interests and perspectives. Their particular access to means of public discourse grants them control over what messages are disseminated and amplified (van Dijk, 1993). Dominant discourse operates to naturalize and perpetuate intergroup inequality, by cultivating negative images of dominated groups and positive images of dominant groups (Hollander & Abelson, 2014; Ruscher, 1998; Van Dijk, 1987, 1993). This dissertation suggests that derogatory gendered labels are not simply general negative terms, but they derive from, and uphold, stereotypes used to justify the

subordinate status of women (James, 1998; Jimenez, 1997; Link & Phelan, 2001). Symbolically, such labels position women as less than, as well as 'other' than men (Link & Phelan, 2001; Schur, 1983). Importantly, those most invested in maintaining male dominance engage in motivated construal of DGLs in ways that perpetuate gender stereotypes. A quotation from Ussher (1992) summarizes the present study's findings: "Thus the labels applied to women, labels which so cleverly place the problem within her as a person" function "as means of both dismissing behavior and controlling what is 'normal.' "

References

Allport, G. W. (1979). The nature of prejudice. Reading, MA: Perseus Books.

American Psychiatric Association. (2017). Words matter: Reporting on mental health conditions. Retrieved from https://www.psychiatry.org/newsroom/reporting-on-mental-health-conditions

Anti-Defamation League. (2016). Anti-Semitic targeting of journalists during the 2016 presidential campaign: A report from ADL's Task Force on Harassment and Journalism. New York: ADL.

Ashwell, L. (2016). Gendered slurs. *Social Theory and Practice, 42*(2), 228-239.

Bailey, A. H., & LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. *Sex Roles, 76*(11-12), 682–693.

Barak, A. (2005). Sexual harassment on the Internet. *Social Science Computer Review, 23*(1), 77–92.

Barrett, L. F., & Bliss-Moreau, E. (2009). She's emotional. He's having a bad day: Attributional explanations for emotion stereotypes. *Emotion, 9*(5), 649–658.

Bellosta, C. J. G. (2014). rPython: Package allowing R to call Python. R package version 0.0-5.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242–261.

Borcher, C. (2016, January). The long, strange history of the Donald Trump-Megyn Kelly feud. Washington Post. Retrieved from https://www.washingtonpost.com/news/the-fix/wp/2016/01/27/the-long-strange-history-of-the-donald-trump-megyn-kelly-feud/?noredirect=on&utm_term=.837612d6a7f2

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science, 19*(3), 268–275.

Brewer, M. B. (1988). Advances in social cognition: A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). New York, New York: Psychology Press.

Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law, 5*(3), 665–692.

Byers, D. (January, 2016). NBC replaced by CNN for GOP's Super Tuesday debate. CNN Business. Retrieved from https://money.cnn.com/2016/01/18/media/gop-debate-nbc-news/index.html

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*(2), 142-156.

Carnaghi, & Maass, A. (2008). Gay or fag? On the consequences of derogatory labels. In Y. Kashima, K. Fielder, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to stereotype formation, maintenance, and transformation* (pp. 117–134). Mahwah, NJ: Erlbaum.

Carnaghi, Maass, & Fasoli, F. (2011). Enhancing masculinity by slandering homosexuals: The role of homophobic epithets in heterosexual gender identity. *Personality and Social Psychology Bulletin, 37*(12), 1655–1665.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008).

Nomina sunt omina: On the inductive potential of nouns and adjectives in person perception. *Journal of Personality and Social Psychology, 94*(5), 839–859.

Chesler, P. (1972). Women and madness. New York, New York: Avon Books.

Cimpian, A. (2010). The impact of generic language about ability on children's achievement motivation. *Developmental Psychology, 46*(5), 1333–1340.

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development, 82*(2), 471–492.

Citron, D. K. (2009). Law's expressive value in combating cyber gender harassment. *Michigan Law Review*, 373–415.

Cohen, Cohen, West, & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). New York, NY: Taylor & Francis.

Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society, 19*(6), 829–859.

Connor, R. A., & Fiske, S. T. (2017). Warmth and competence: A feminist look at power and negotiation. In C. B. Travis & J. W. White (Eds.), *APA handbook of the psychology of women.* Washington, D.C.: American Psychological Association.

Connor, Glick, & Fiske, S. T. (2016). Ambivalent Sexism in the 21st Century. In F. C. & Barlow Sibley (Ed.), *Cambridge handbook of the psychology of prejudice.* Cambridge, UK: Cambridge University Press.

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin, 129*(3), 414-446.

Crandall, & White, M. H. II. (2016, November). Trump and the social psychology of

prejudice. *Undark.* Retrieved from https://undark.org/article/trump-social-psychology-prejudice-unleashed/

Cummins, J. (2016). This is the dirtiest presidential race since ’72. *Politico.* Retrieved from https://www.politico.com/magazine/story/2016/02/2016-elections-nastiest-presidential-election-since-1972-213644

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 273–274). International World Wide Web Conferences Steering Committee.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review, 94*(3), 369–389.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Diaz, D. (2016, October). Trump calls Clinton “a nasty woman”. CNN. Retrieved from http://www.cnn.com/2016/10/19/politics/donald-trump-hillary-clinton-nasty-woman/index.html

Dittmar, K. (2016, September 25). Before there was a woman nominee, women moderators provided gender diversity on the presidential debate stage. Presidential Gender Watch. Retrieved from http://presidentialgenderwatch.org/woman-nominee-women-moderators-provided-gender-diversity-presidential-debate-stage/

Duckitt, J., & Sibley, C. G. (2009). A dual-process motivational model of ideology, politics, and prejudice. *Psychological Inquiry, 20*, 98–109.

Duggan. (2017, July 11). Online Harrassment 2017. Pew Research Center. Retrieved

from http://www.pewinternet.org/2017/07/11/online-harassment-2017/

Duggan, M. (2014, October). Online Harassment. Pew Research Center. Retrieved from http://www.pewinternet.org/2014/10/22/online-harassment/

Ehrlich, S., & King, R. (1994). Feminist meanings and the (de) politicization of the lexicon. *Language in Society, 23*(1), 59–76.

Fasoli, Carnaghi, & Paladino, M. P. (2015). Social acceptability of sexist derogatory and sexist objectifying slurs across contexts. *Language Sciences, 52*, 98–107.

Fasoli, Maass, & Carnaghi, A. (2015). Labelling and discrimination: Do homophobic epithets undermine fair distribution of resources? *British Journal of Social Psychology, 54*(2), 383–393.

Fasoli, Paladino, Carnaghi, Jetten, Bastian, & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology, 46*(2), 237–248.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104.

Fine, G. A., & Elsbach, K. D. (2000). Ethnography and experiment in social psychological theory building: Tactics for integrating qualitative field data with quantitative lab data. *Journal of Experimental Social Psychology, 36*(1), 51-76.

Fischer, A. H., Eagly, A. H., & Oosterwijk, S. (2013). The meaning of tears: Which sex seems emotional depends on the social context. *European Journal of Social Psychology, 43*(6), 505–515.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and

motivation on attention and interpretation. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Freelon, D., & Karpf, D. (2015). Of big birds and bayonets: Hybrid Twitter interactivity in the 2012 presidential debates. *Information, Communication & Society, 18*(4), 390-406.

Ford, T. E., & Ferguson, M. A. (2004). Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review, 8*(1), 79–94.

García-Pastor, M. D. (2008). Political campaign debates as zero-sum games: Impoliteness and power in candidates' exchanges. In D. Bousfield & M. A. Locher (Eds.), *Impoliteness in language: Studies on its interplay with power in theory and practice* (pp. 101-124). Berlin: Mouton de Gruyter.

Gelman, S. A., Taylor, M. G., Nguyen, S. P., Leaper, C., & Bigler, R. S. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in child Development, 69*(1), i-142.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21–38.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491–512.

Glick, P., & Fiske, S. T. (1999). The Ambivalence toward Men Inventory: Differentiating

hostile and benevolent beliefs about men. *Psychology of Women Quarterly, 23*(3), 519–536.

Glick, P., Wilkerson, M., & Cuffe, M. (2015). Masculine identity, ambivalent sexism, and attitudes toward gender subtypes. *Social Psychology*, *46*(4), 210–217.

Goodman, J. A., Schell, J., Alexander, M. G., & Eidelman, S. (2008). The impact of a derogatory remark on prejudice toward a gay male leader. *Journal of Applied Social Psychology, 38*(2), 542–555.

Gottfried, J., Barthel, M., Shearer, E., & Mitchell, A. (2016). The 2016 presidential campaign – a news event that's hard to miss. Washington, D. C.: Pew Research Center.

Greenberg, Kirkland, & Pyszczynski, T. (1988). Some theoretical notions and preliminary research concerning derogatory ethnic labels. In G. Smitherman-Donaldson & T. A. van Dijk (Eds.), *Discourse and discrimination.* Detroit, Michigan: Wayne State University Press.

Greenberg, J., & Pyszczynski, T. (1985). The effect of an overheard ethnic slur on evaluations of the target: How to spread a social disease. *Journal of Experimental Social Psychology*, *21*(1), 61–72.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267-297.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408–420.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for

success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*(3), 416–427.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition & Emotion, 14*(5), 609–642.

Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of novel predicates to make inductive inferences about people. *Cognitive Development, 15*(3), 263–280.

Hinck, S. S., Hinck, E. A., Dailey, W. O. Hinck, R. S., Hansen, E., & Madsen, B. (2018). A tale of two parties: Comparing the intensity of face threats in the Democratic and Republican primary debates of 2012 and 2016. *Argumentation and Advocacy*, *54*(1-2), 104-121.

Hollander, J. A., & Abelson, M. J. (2014). Language and talk. In J. McLeod, E. Lawler, & M. Schwalbe (Eds.), *Handbook of the social psychology of inequality* (pp. 181–206). Dordrecht, Netherlands: Springer.

James, D. (1998). Gender-linked derogatory terms and their use by women and men. *American Speech, 73*(4), 399–420.

Jimenez, M. A. (1997). Gender and psychiatry: Psychiatric conceptions of mental disorders in women, 1960-1994. *Affilia, 12*(2), 154–175.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2*(3), 171–203.

Jones, C. S., & Kaplan, M. F. (2003). The effects of racially stereotypical crimes on juror

decision-making and information-processing strategies. *Basic and Applied Social Psychology, 25*(1), 1–13.

Jones, & Nisbett, R. E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & W. Bernard (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Morristown, NJ: General Learning Press.

Jost, Banaji, & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*(6), 881–919.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology, 82*(5), 965-990.

Keates, J., & Graham, S. A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science, 19*(12), 1287–1293.

Kelly, J. R., & Hutson-Comeaux, S. L. (2000). The appropriateness of emotional expression in women and men: The double-bind of emotion. *Journal of Social Behavior and Personality, 15*(4), 515–528.

Kimmel, M. S. (1997). Masculinity as homophobia: Fear, shame, and silence in the construction of gender identity. In M. Gergen & S. Davis. (Eds.), *Toward a new psychology of gender* (pp. 223–227). New York, New York: Routledge.

Kirkland, S. L., Greenberg, J., & Pyszczynski, T. (1987). Further evidence of the deleterious effects of overheard derogatory ethnic labels: Derogation beyond the target. *Personality and Social Psychology Bulletin, 13*(2), 216–227.

Kohut, A., Doherty, C., Dimock, M., & Keeter, S. (2012). One-in-ten 'dual-screened' the

presidential debate: TV trumps digital sources for debate coverage. Washington, D. C.: Pew Research Center.

Kohut, A., & Parker, A. (2008). Internet now major source of campaign news: Continuing partisan divide in cable TV news audiences. Washington, D. C.: Pew Research Center.

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin, 19*(1), 90–99.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*(2), 284–308.

Leader, T., Mullen, B., & Rice, D. (2009). Complexity and valence in ethnophaulisms and exclusion of ethnic out-groups: What puts the " hate" into hate speech? *Journal of Personality and Social Psychology, 96*(1), 170–182.

Lee, J. C., & Quealy, K. (2016, October). The 282 people, places and things Donald Trump has insulted on Twitter: A complete list. New York Times. Retrieved from http://www.nytimes.com/interactive/2016/01/28/upshot/donald-trump-twitter-insults.html

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*(2), 275-287.

Levant, R. F., Hall, R. J., & Rankin, T. J. (2013). Male Role Norms Inventory–Short Form (MRNI-SF): Development, confirmatory factor analytic investigation of structure, and measurement invariance across gender. *Journal of Counseling Psychology, 60*(2), 228–238.

Levant, R. F., & Richmond, K. (2008). A review of research on masculinity ideologies using the Male Role Norms Inventory. *The Journal of Men's Studies, 15*(2), 130–146.

Lewis, K. M. (2000). When leaders display emotion: How followers respond to negative emotional expression of male and female leaders. *Journal of Organizational Behavior,* 221–234.

Link, B. G., Cullen, F. T., Frank, J., & Wozniak, J. F. (1987). The social rejection of former mental patients: Understanding why labels matter. *American Journal of Sociology, 92*(6), 1461–1500.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27*(1), 363–385.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology, 39*(5), 821-831.

Lombardo, W. K., Cretser, G. A., & Roesch, S. C. (2001). For crying out loud—The differences persist into the'90s. *Sex Roles, 45*(7), 529–547.

Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. In B. H. Ross (Ed.), *Psychology of learning and motivation-Advances in research and theory* (Vol. 57, pp. 255–297). San Diego, CA: Academic Press.

Maass, A., Arcuri, L. (1996) Language and stereotyping. In N. Macrae, M. Hewstone, & C. Stangor (Eds.) *The foundations of stereotypes and stereotyping* (pp. 193-226). New York: Guilford.

Maass, A., & Borini, C. (2005). Converting verbs into adjectives: Asymmetrical memory distortions for stereotypic and counterstereotypic information. *Group Processes & Intergroup Relations, 8*(3), 271–290.

Maass, Karasawa, Politi, & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic analysis of person representation. *Journal of Personality and Social Psychology, 90*(5), 734–750.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*(1), 93–120.

Maronna, R. A., & Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference, 89*(1), 197–214.

McCrae, & Costa, P. T. Jr. (2008). A five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 139–153). New York, New York: Guilford Press.

Mendelberg, T. (2008). Racial priming revived. *Perspectives on Politics, 6*(1), 109–123.

Miller, C., & Werner-Winslow, A. (2016). Ten days after: Harassment and intimidation in the aftermath of the election. Montgomery, AL: Southern Poverty Law Center.

Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity, 11*(2), 140–151.

Mullen, Blevins, & Schmidt, B. (2015). gender: Predict gender from names using historical data. Retrieved from https://cran.r-project.org/web/packages/gender/gender.pdf

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

Parks-Stamm, E. J., Heilman, M. E., & Hearns, K. A. (2008). Motivated to penalize:

Women's strategic rejection of successful women. *Personality and Social Psychology Bulletin, 34*(2), 237-247.

Pascoe, C. J. (2005). "Dude, you're a fag": Adolescent masculinity and the fag discourse. *Sexualities, 8*(3), 329–346.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC [computer software]. Austin, TX: LIWC.net.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly, 32*(4), 406–413.

Pittinsky, T. L., Shih, M., & Ambady, N. (2000). Will a category cue affect you? Category cues, positive stereotypes and reviewer recall for applicants. *Social Psychology of Education, 4*(1), 53–65.

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly, 24*(1), 81–92.

Postmes, T., & Spears, R. (2002). Behavior online: Does anonymous computer communication reduce gender inequality?. *Personality and Social Psychology Bulletin, 28*(8), 1073-1083.

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology, 17*(1), 271-320.

Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*(4), 269–281.

Presidential Gender Watch. (2016). Finding gender in election 2016: Lessons from presidential gender watch. Retrieved from http://www.cawp.rutgers.edu/sites/default/files/resources/presidential-gender-gap_report_final.pdf

Preston, K., & Stanley, K. (1987). "What's the worst thing...?" Gender-directed insults. *Sex Roles, 17*(3), 209–219.

realDonaldTrump. (2015, August 7). @timjcam: @megynkelly @FrankLuntz @realDonaldTrump Fox viewers give low marks to bimbo @MegynKelly will consider other programs! [Tweet]. Retrieved from https://twitter.com/realDonaldTrump/status/629553612839124992

realDonaldTrump. (2016, January 27). I refuse to call Megyn Kelly a bimbo, because that would not be politically correct. Instead I will only call her a lightweight reporter! [Tweet]. Retrieved from https://twitter.com/realDonaldTrump/status/692312112115380224

Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, 109*(34), 13526–13531.

Robinson, M. D., & Johnson, J. T. (1997). Is it emotion or is it stress? Gender stereotypes and the perception of subjective experience. *Sex Roles, 36*(3-4), 235–258.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). London, UK: Sage.

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

Social Psychology, *74*(3), 629-645.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology, 87*(2), 157–176.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010.

Rudman, L. A., & Mescher, K. (2013). Penalizing men who request a family leave: Is flexibility stigma a femininity stigma? *Journal of Social Issues, 69*(2), 322–340.

Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012). Reactions to vanguards: Advances In backlash theory. In P. G. Devine & E. A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 45, pp. 167–227). New York, New York: Elsevier.

Ruscher, J. B. (1998). Prejudice and stereotyping in everyday communication. *Advances in Experimental Social Psychology*, *30*, 241–307.

Salerno, J. M., & Peter-Hagene, L. C. (2015). One angry woman: Anger expression increases influence for men, but decreases influence for women, during group deliberation. *Law and Human Behavior, 39*(6), 581–592.

Salganik, M. J. (2017). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Santucci, J., & Hopper, J. (2016). Donald Trump Calls Ted Cruz a “weak little baby” ahead of Nevada caucuses. ABC News. Retrieved from http://abcnews.go.com/Politics/donald-trump-calls-ted-cruz-weak-baby-ahead/story?id=37147435

Saucier, Till, Miller, O'Dea, & Andres, E. (2015). Slurs against masculinity: Masculine honor beliefs and men's reactions to slurs. *Language Sciences, 52,* 108–120.

Schur, E. M. (1983). Labeling women deviant: Gender, stigma, and social control. Philadelphia, PA: Temple University Press.

Semin, G. R., & Fiedler, K. (1992). Language, interaction, and social cognition. Newbury Park, CA: Sage.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European Review of Social Psychology*, *2*(1), 1–30.

Setzler, M., & Yanus, A. B. (2018). Why did women vote for Donald Trump?. *Political Science & Politics, 51*(3), 523-527.

Shields, S. A. (2002). Speaking from the heart: Gender and the social meaning of emotion. Cambridge, UK: Cambridge University Press.

Shields, & Crowley, J. C. (1996). Appropriating questionnaires and rating scales for a feminist psychology: A multi-method approach to gender and emotion. In S. Wilkinson (Ed.), Feminist social psychologies (pp. 218–232). Bristol, PA: Open University Press.

Shiffrin, R. M. (2016). Drawing causal inference from big data. *Proceedings of the National Academy of Sciences*, *113*(27), 7308-7309.

Slater, M. D. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, *17*(3), 281-303.

Sibley, C. G., Wilson, M., & Duckitt, J. (2007). Antecedents of men's hostile and

benevolent sexism: The dual roles of Social Dominance Orientation and Right-Wing Authoritarianism. *Personality and Social Psychology Bulletin, 33*, 160–172.

Sifter (2018). Retrieved from https://sifter.texifter.com/

Simon, L., & Greenberg, J. (1996). Further progress in understanding the effects of derogatory ethnic labels: The role of preexisting attitudes toward the targeted group. *Personality and Social Psychology Bulletin, 22*(12), 1195–1204.

Smith, J. S. (2016). Too much emotion: Intersections of gender, status, and group membership in perception of emotional appropriateness. Presented at the 1st Gender SPSP Preconference, San Diego, CA.

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*(1), 42–61.

Stone, P. (2002). The General Inquirer: A computer approach to content analysis. Retrieved from http://www.wjh.harvard.edu/ inquirer/

Strauss, D. (2016). Donald Trump calls Ted Cruz the P word. Politico. Retrieved from http://www.politico.com/story/2016/02/donald-trump-calls-ted-cruz-the-p-word-218984

Sunstein, C. (2003). R. 2007. Republic. com 2.0. Princeton, NJ: Princeton University Press.

Sutton, L. A. (2001). Bitches and skankly hobags: The place of women in contemporary slang. In K. Hall & M. Bucholtz (Eds.), *Gender articulated: Language and the socially constructed self* (pp. 279–296). New York: Routledge.

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

fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199-214.

Thompson, E. H., Pleck, J. H., & Ferrera, D. L. (1992). Men and masculinities: Scales for masculinity ideology and masculinity-related constructs. *Sex Roles, 27*(11), 573–607.

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *Journal of Personality and Social Psychology, 80*(1), 86–94.

Timmers, M., Fischer, A., & Manstead, A. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. *Cognition & Emotion, 17*(1), 41–63.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59*, 329–360.

Ussher, J. (1992). Women's madness: Misogyny or mental illness? Amherst, MA: University of Maschusetts Press.

Van Dijk, T. A. (1987). Communicating racism: Ethnic prejudice in thought and talk. Newbury Park, CA: Sage.

Van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, *4*(2), 249–283.

Waxman, & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13*(6), 258–263.

Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial

and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4), 593–610.

Waxman, & Lidz, J. L. (2006). Early word learning. In R. D. & Siegler Kuhn (Eds.), Handbook of child psychology. Hoboken, NJ: Wiley.

Wells, C., Shah, D. V., Pevehouse, J. C., Yang, J., Pelled, A., Boehm, F., ... & Schmidt, J. L. (2016). How Trump drove coverage to the nomination: Hybrid media campaigning. *Political Communication*, *33*(4), 669-676.

Wigboldus, D. H., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, *78*(1), 5–18.

Winter, N. (2008). Dangerous frames: How ideas about race and gender shape public opinion. Chicago, IL: University of Chicago Press.

Wittenbrink, Judd, & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262-274.

Women's Media Center. (2015). The status of U. S. women in the media 2015. Retrieved from http://www.womensmediacenter.com/reports/2015-statistics

Women's Media Center. (2017). The status of U. S. women in the media 2017. Retrieved from http://www.womensmediacenter.com/reports/the-status-of-women-in-u.s.-media-2017

Zezima, K. (October, 2015). The debate over CNBC's moderators. Washington Post.

Retrieved from https://www.washingtonpost.com/politics/how-cnbc-actually-messed-up-wednesdays-gop-debate/2015/10/29/b21203bc-7e52-11e5-afce-2afd1d3eb896_story.html?utm_term=.bf616d434d5a

Zhang, Y., Wells, C., Wang, S., & Rohe, K. (2018). Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election. *New Media & Society, 20*(9), 3161-3182.

Table 1

Results of regression analyses for Studies 1-4.

	Variables				**Interactions**										
	Target Gender (TG)	Label (L)	Participant Gender (PG)	Gender Ideology (GI)	2-ways						3-ways				4-way
					TG*GI	TG*L	TG*PG	L*GI	L*PG	PG*GI	TG*GI*PG	TG*L*PG	TG*L*GI	L*PG*GI	
Study 1															
Target															
Likability	2.10***				-1.41*	-2.24**					2.49**	2.53*			-2.52*
Competence	1.52**	.84*			-1.21**	-1.44*					2.07**				-1.85*
Internal Attribution	-2.18**					2.87**		-1.12*	1.82*			-3.96**			
Labeler															
Likability		-1.01*											1.02†		
Competence		-.76**		.40†									.84*		
Study 2															
Target															
Likability				-1.23**	1.75**								-2.02*		3.10**
Competence											-1.58*		-1.46*		2.92**
Internal Attribution	-1.04**			.61*	-.88*								1.59**		
Labeler															
Likability								.95**					-1.20*		
Competence								.63**					-.76*		
Study 3															
Target															
Likability															
Competence								.96**					-1.09*		
Internal Attribution								-.99**					.99*		
Labeler															
Likability		-1.72***		.55*				-.82*					.88†		
Competence		-.85**						-.86**					.94*		

	Variables				Interactions										
	Target Gender (TG)	Label (L)	Participant Gender (PG)	Gender Ideology (GI)	2-ways						3-ways				4-way
					TG*GI	TG*L	TG*PG	L*GI	L*PG	PG*GI	TG*GI*PG	TG*L*PG	TG*L*GI	L*PG*GI	
Study 4															
Target															
Likability	1.81***			-.53*		-1.19*									
Competence	.85**				-.62*										
Internal Attribution	-.89**														
Labeler															
Likability	-1.41***	-2.73***			.82**										
Competence		-1.15***		.54**											

Note. Unstandardized coefficients are reported. † = $p = .08$; * = $p < .05$; ** = $p < .01$; *** = $p < .001$

Table 2

Descriptive information on dictionary categories

	Example Words	Krippendorff's alpha	Size of Category
Sexual	Whore, cuck, cheater	.72	63
Ethnic	Hood rat, ape, beaner	.72	23
Homophobic	Dyke, fag, homo	.76	30
Incompetent	Airhead, jackass, dipshit	.52	49
Feminine/Weak	Priss, cuck, cry baby	.67	57
Masculine/Aggressive	Bitch, dick, jerk	.55	79
Physical Appearance	Titsy, neckbeard, ugly	.36	88

Table 3

List of moderators included in Study 5.

	Gender of Moderator		
Political Party & Date of Debate	**Male**	**Female**	**# Tweets**
Democratic			
12/19/2015	David Muir	Martha Raddatz	12205
1/17/2016	Lester Holt	Andrea Mitchell	32344
2/4/2016	Chuck Todd	Rachel Maddow	48834
Republican			
10/28/2015	Carl Quintanilla	Becky Quick	28123
1/14/2016	Neil Cavuto	Maria Bartiromo	23885
3/3/2016	Bret Baier	Megyn Kelly	147300

Table 4

Derogatory gendered labels by target gender and debate.

Debate	Male/Both Targets	Female Target
Dem - Martha / David, Dec 19	4	0
Dem - Andrea / Lester, January 17	2	4
Dem - Rachel / Chuck, Feb 4	4	0
GOP - Maria / Neil, Jan 14	7	10
GOP - Becky / Carl, Oct 28	25	21
GOP - Megyn / Bret, March 3	2	63

Table 5

Incompetence labels by debate.

Debate	
Dem - Martha / David, Dec 19	4
Dem - Andrea / Lester, January 17	20
Dem - Rachel / Chuck, Feb 4	7
GOP - Maria / Neil, Jan 14	22
GOP - Becky / Carl, Oct 28	53
GOP - Megyn / Bret, March 3	17

Table 6

Masculine labels by debate.

Debate	
Dem - Martha / David, Dec 19	1
Dem - Andrea / Lester, January 17	7
Dem - Rachel / Chuck, Feb 4	2
GOP - Maria / Neil, Jan 14	8
GOP - Becky / Carl, Oct 28	35
GOP - Megyn / Bret, March 3	40

Table 7

Physical appearance labels by target gender and debate.

Debate	Male/Both Targets	Female Target
Dem - Martha / David, Dec 19	1	0
Dem - Andrea / Lester, January 17	0	1
Dem - Rachel / Chuck, Feb 4	2	4
GOP - Maria / Neil, Jan 14	6	2
GOP - Becky / Carl, Oct 28	3	2
GOP - Megyn / Bret, March 3	0	32

Table 8

Study 6. Tweet type by word count, valence, and verbal abstractness.

	Tweet Type			p-values	
Variable	**DGL**	**Negative**	**Positive**	**DGL-Neg**	**DGL-Pos**
Word count	17.17	15.33	14.33	.23	.43
Positive valence	.83	.83	2.17	1.00	.02*
Negative valence	1.83	1.83	0	1.00	.003**
Verbal abstractness	2.51	2.76	2.56	.36	.83

Table 9

Study 7. Tweet type by word count, valence, and verbal abstractness.

	Tweet Type			p-values	
Variable	**DGL**	**Negative/High Abstract**	**Negative/Low Abstract**	**DGL-Neg HA**	**DGL-Neg LA**
Word count	17.17	15.33	14.00	.18	.14
Positive valence	.83	.83	1.33	1.00	.22
Negative valence	1.83	1.83	1.50	1.00	.60
Verbal abstractness	2.51	2.76	1.73	.31	.006**

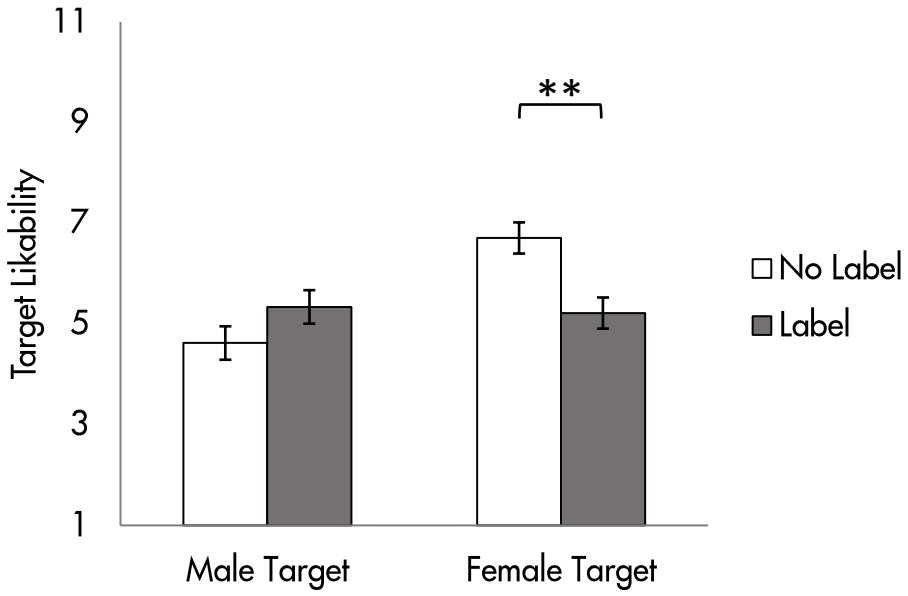


Figure 1. Study 1. Male participants only. Target likability by label condition and target gender. Difference between the male and female target in the no label condition is also significant at p = .001.

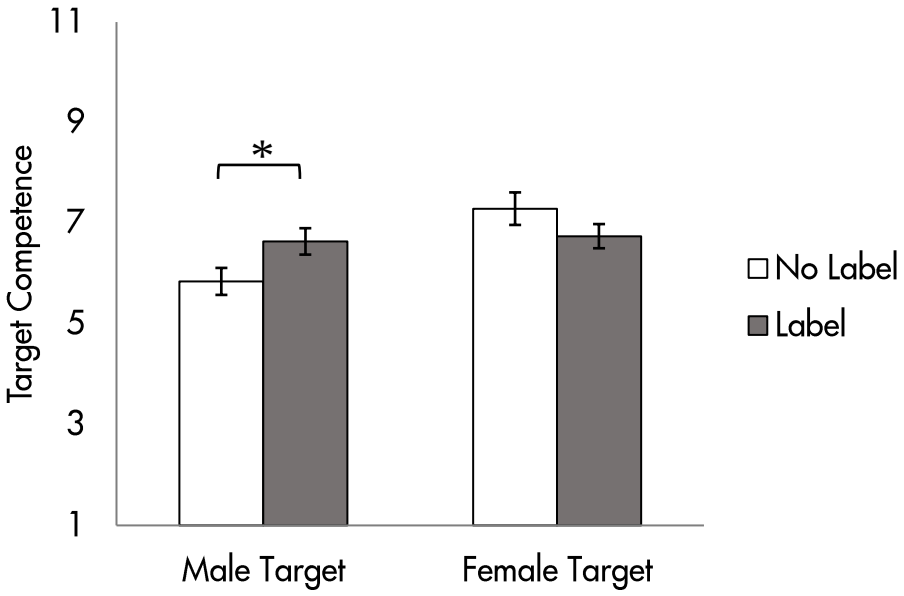


Figure 2. Study 1. Male participants only. Target competence by label condition and target gender. Difference between the male and female target in the no label condition is also significant at $p < .001$.

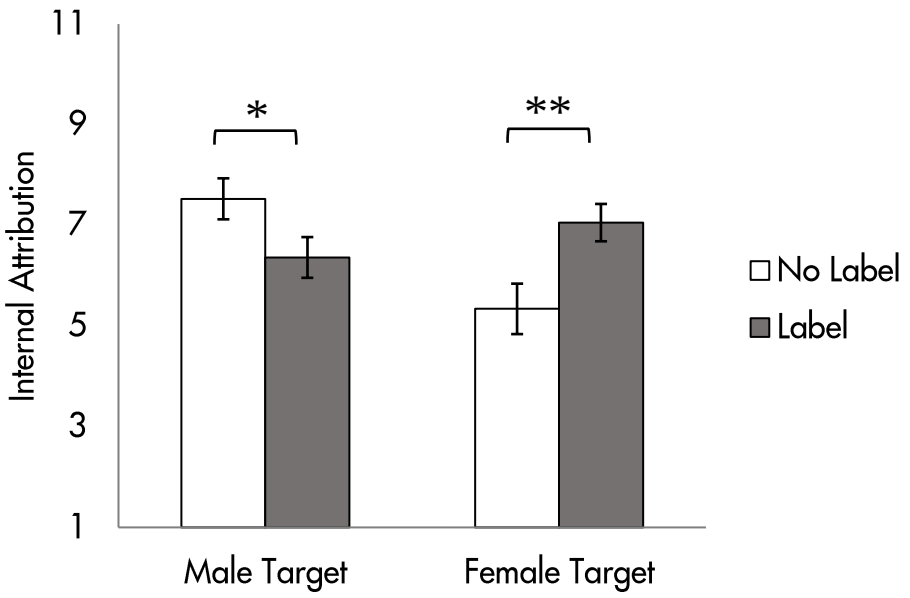


Figure 3. Study 1. Internal attribution by label condition and target gender. Difference between the male and female target in the no label condition is also significant at $p < .01$.

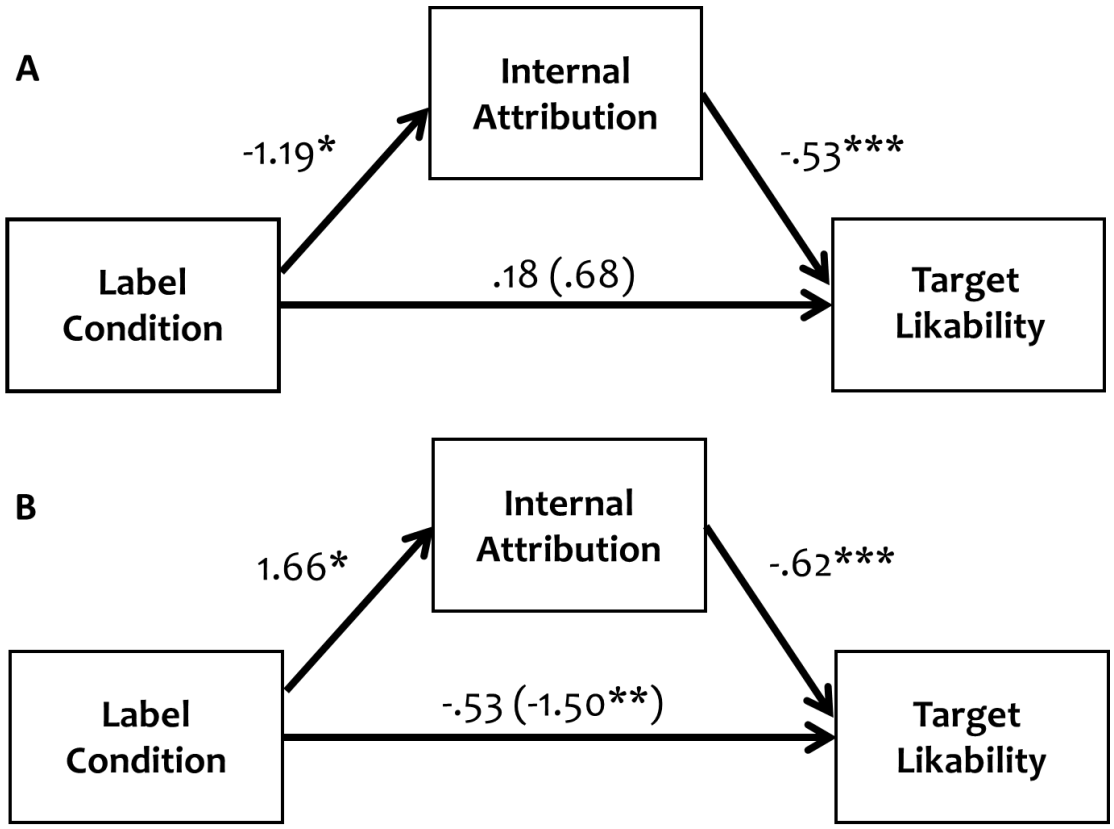


Figure 4. Study 1. Male participants only. Internal attribution as a mediator of the relationship between label condition and target likability for male targets (A) and female targets (B).

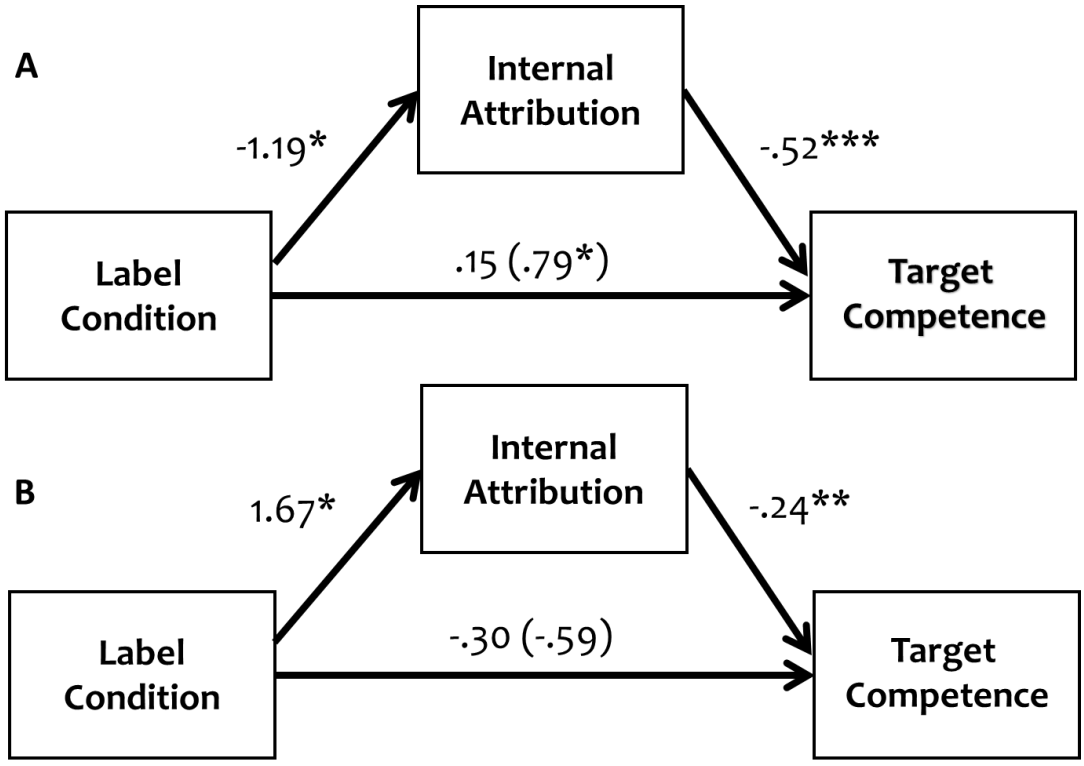


Figure 5. Study 1. Male participants only. Internal attribution as a mediator of the relationship between label condition and target competence for male targets (A) and female targets (B).

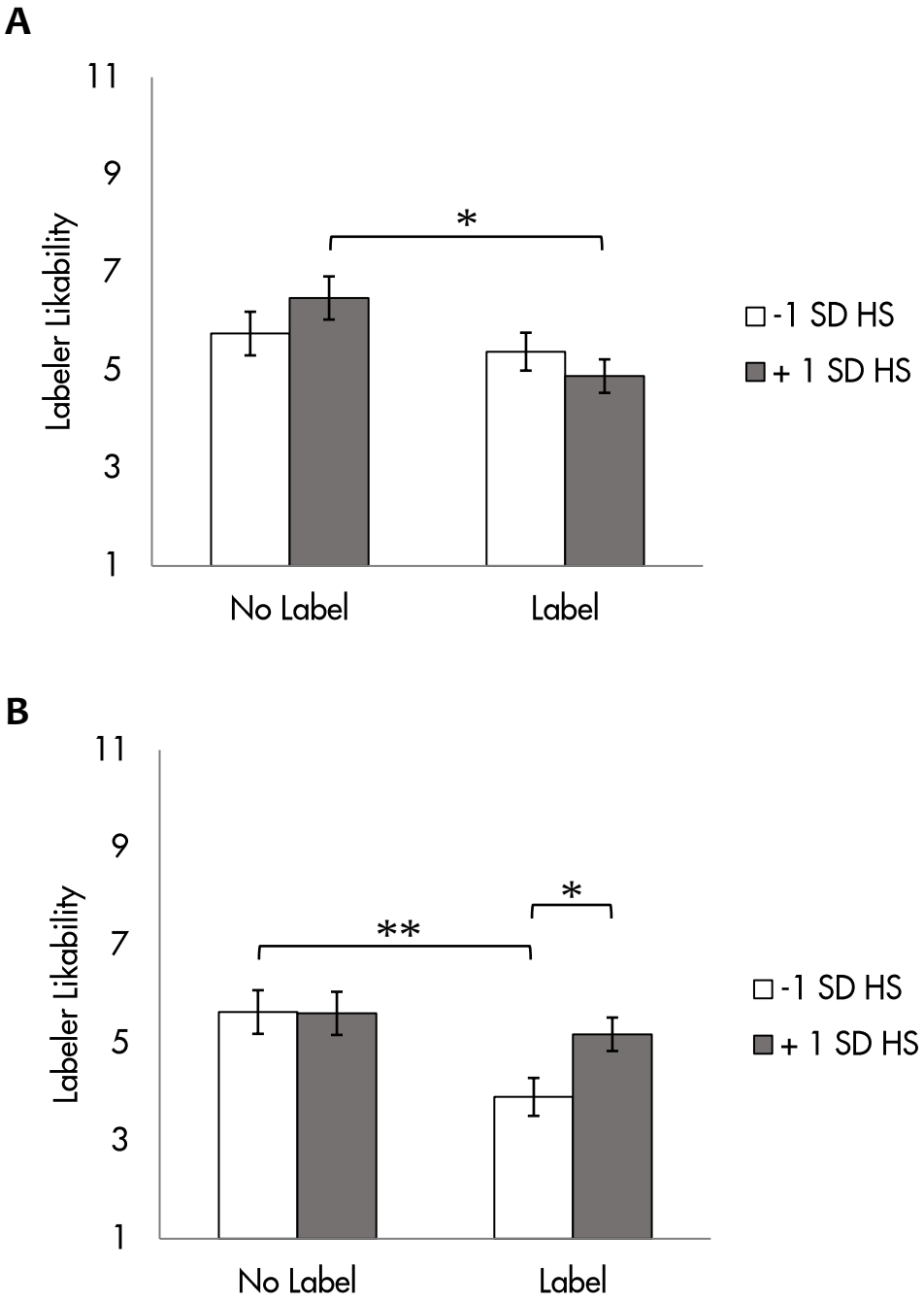


Figure 6. Study 1. Labeler likability by label condition and hostile sexism for male targets (A) and female targets (B).

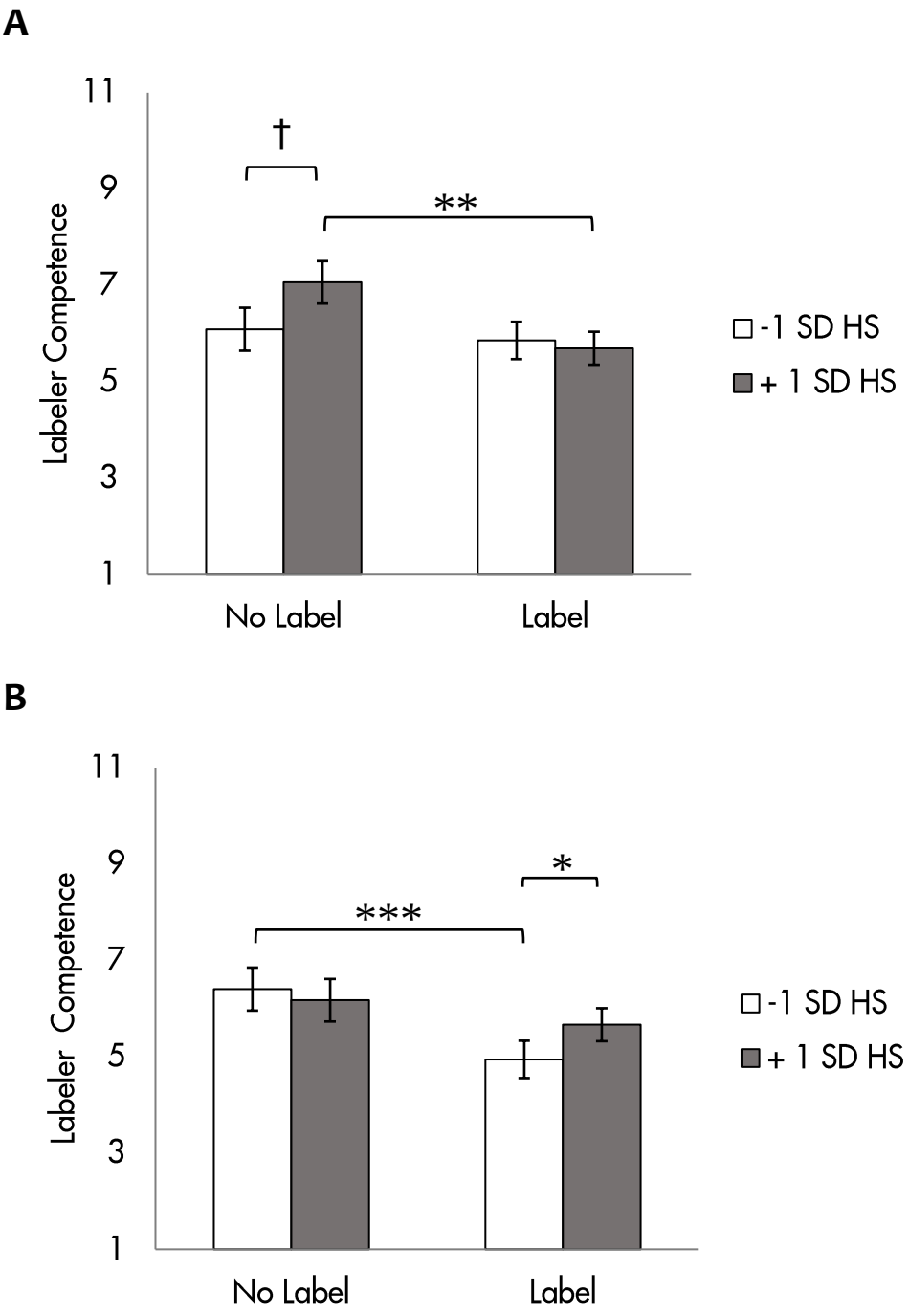


Figure 7. Study 1. Labeler competence by label condition and hostile sexism for male targets (A) and female targets (B).

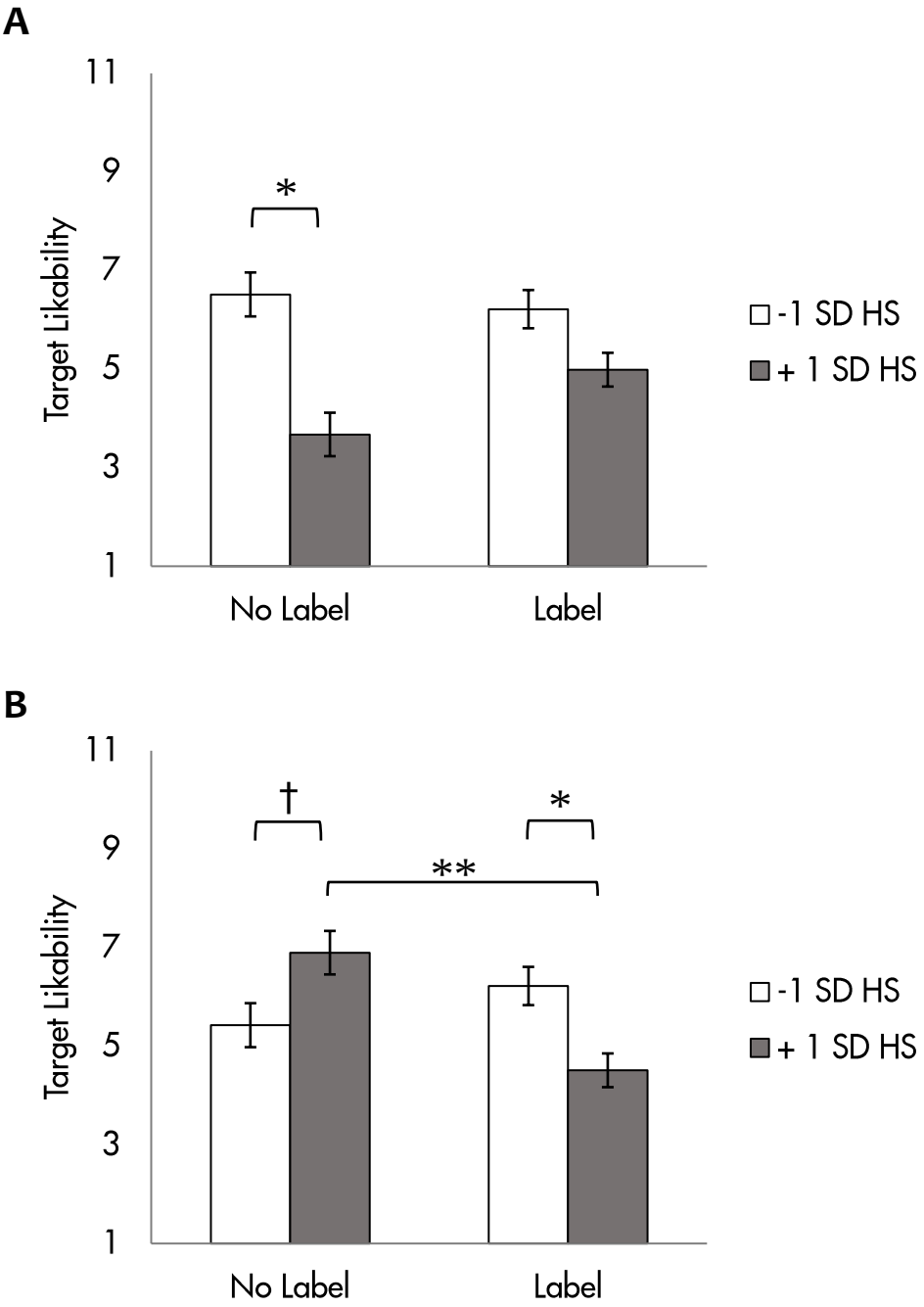


Figure 8. Study 2. Target likability by label condition and hostile sexism for male targets (A) and female targets (B).

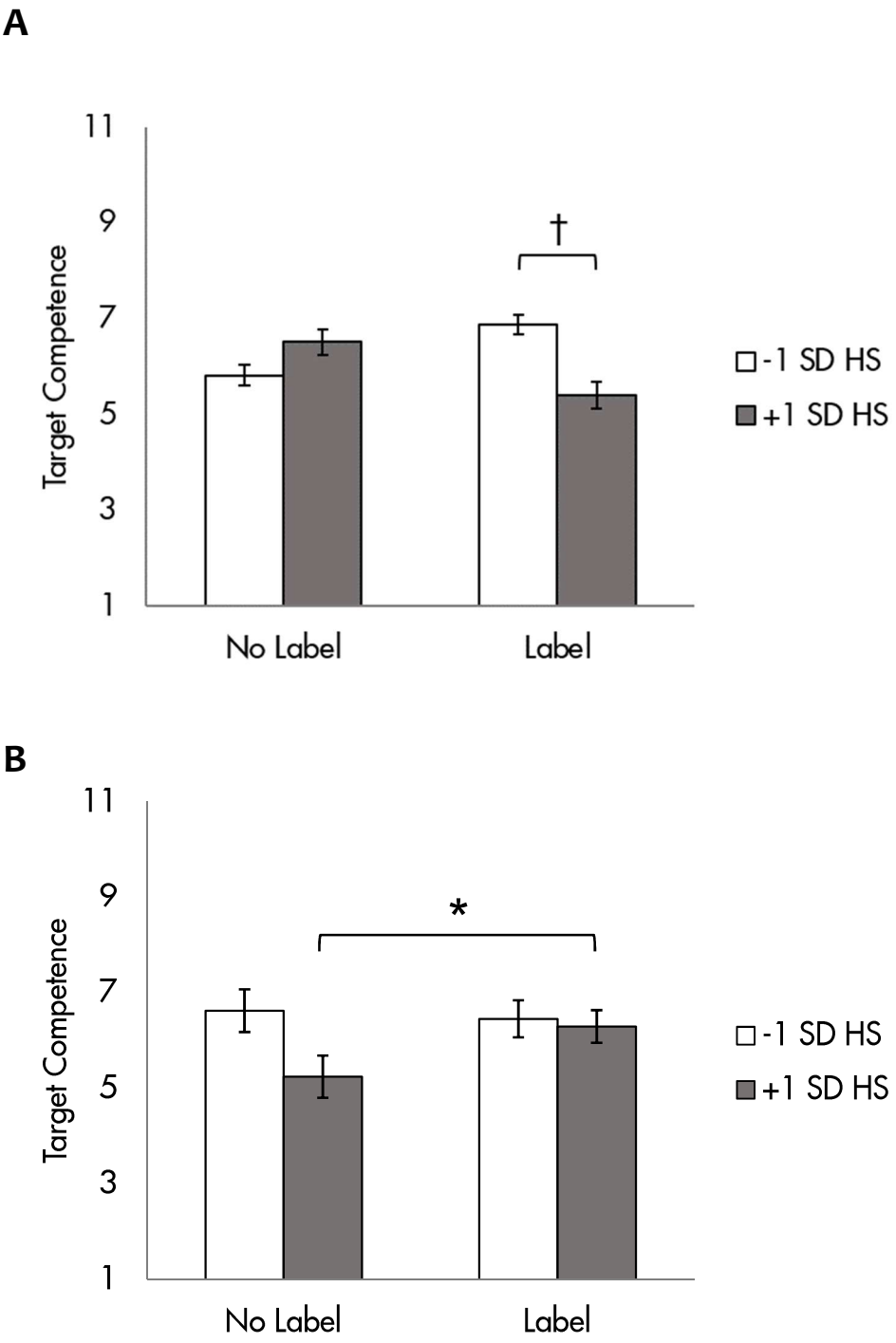


Figure 9. Study 2. Target competence by label condition and hostile sexism for male participants (A) and female participants (B) evaluating female targets.

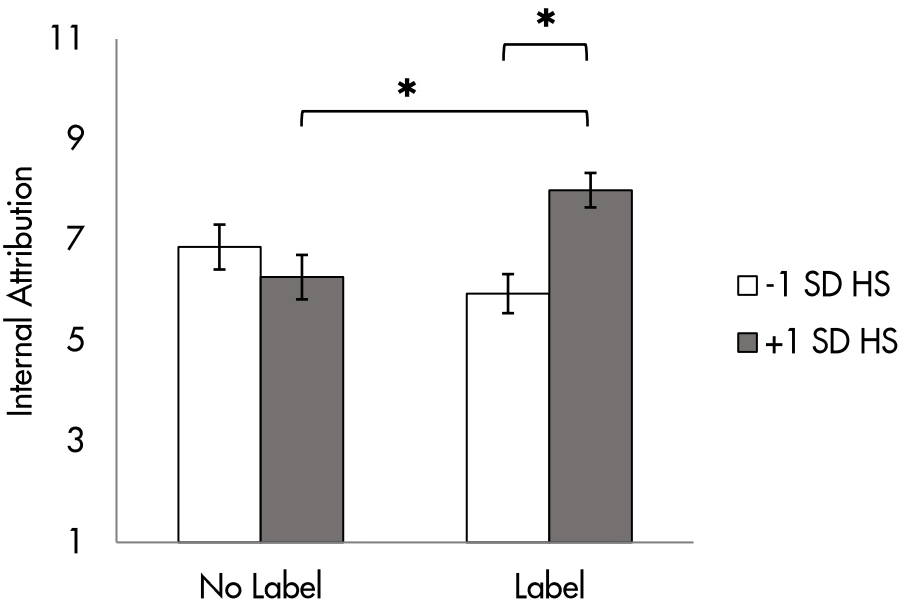


Figure 10. Study 2. Internal attribution by label condition and hostile sexism for the female target (all participants).

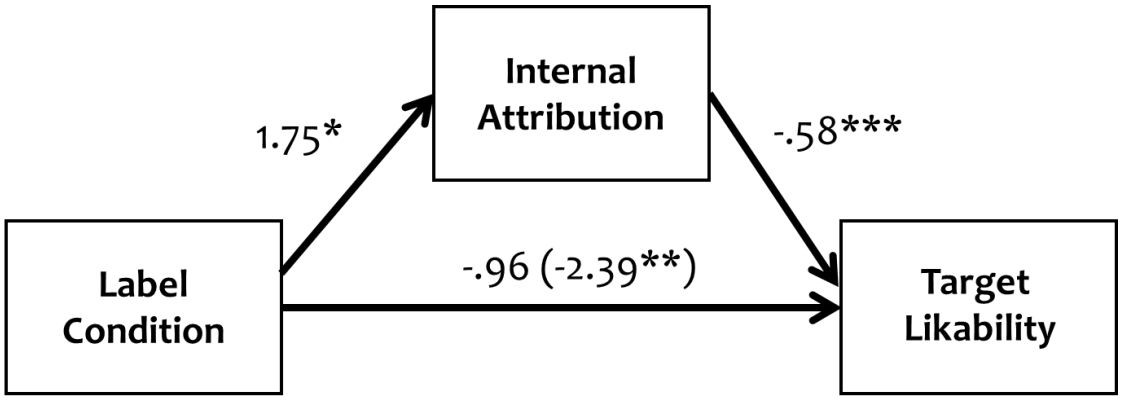


Figure 11. Study 2. Male participants only. Estimates calculated at 1 SD above the mean of hostile sexism. Internal attribution as a mediator of the relationship between label condition and target likability for female targets.

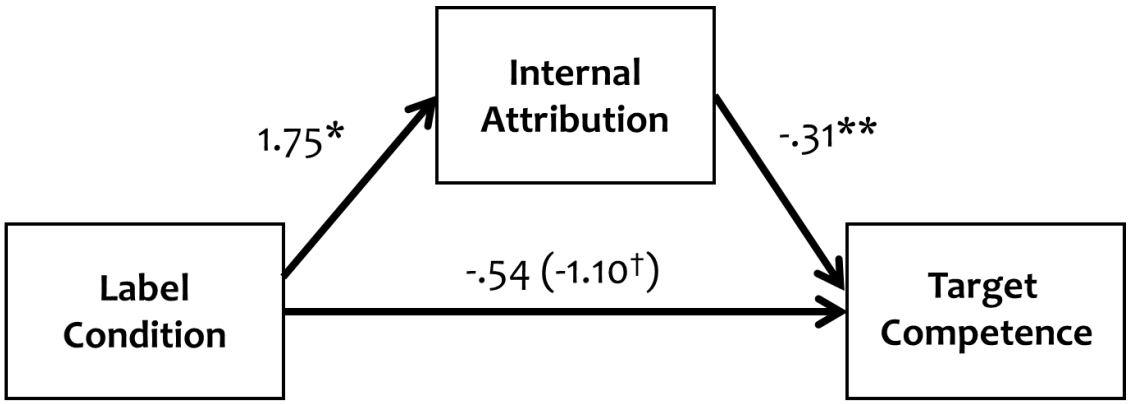


Figure 12. Study 2. Male participants only. Estimates calculated at 1 SD above the mean of hostile sexism. Internal attribution as a mediator of the relationship between label condition and target competence for female targets.

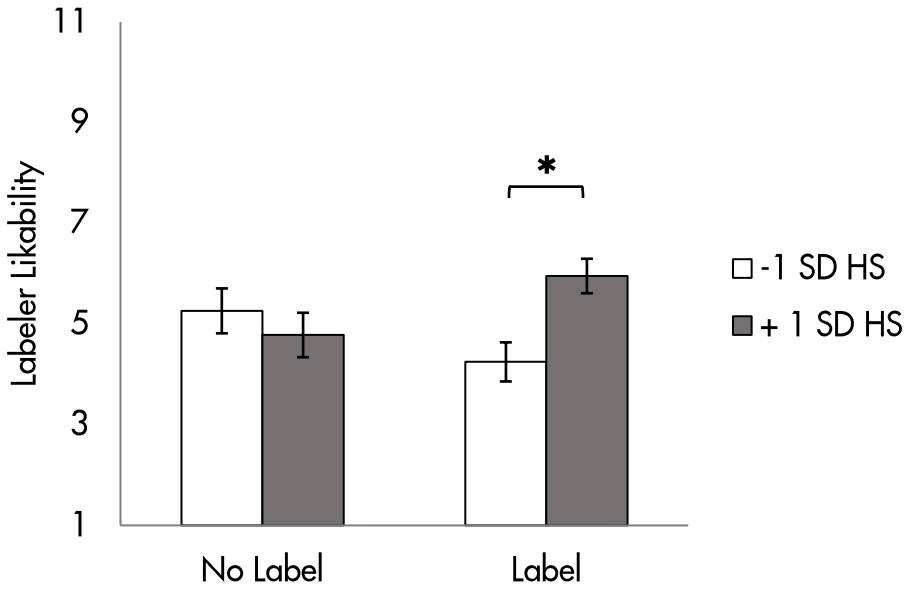


Figure 13. Study 2. Estimates calculated at 1 SD above the mean of hostile sexism. Labeler likability by label condition and hostile sexism for the male target.

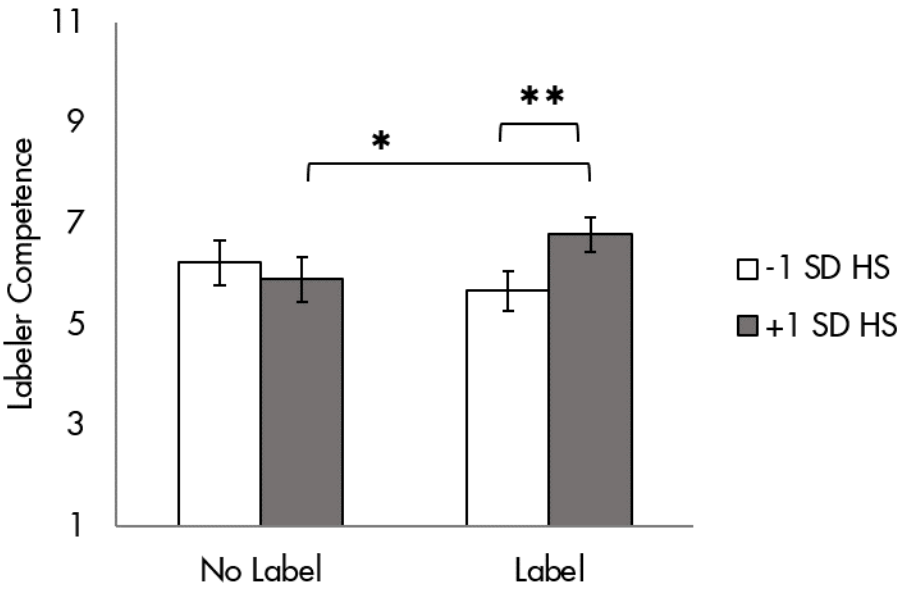


Figure 14. Study 2. Estimates calculated at 1 SD above the mean of hostile sexism. Labeler competence by label condition and hostile sexism for the male target.

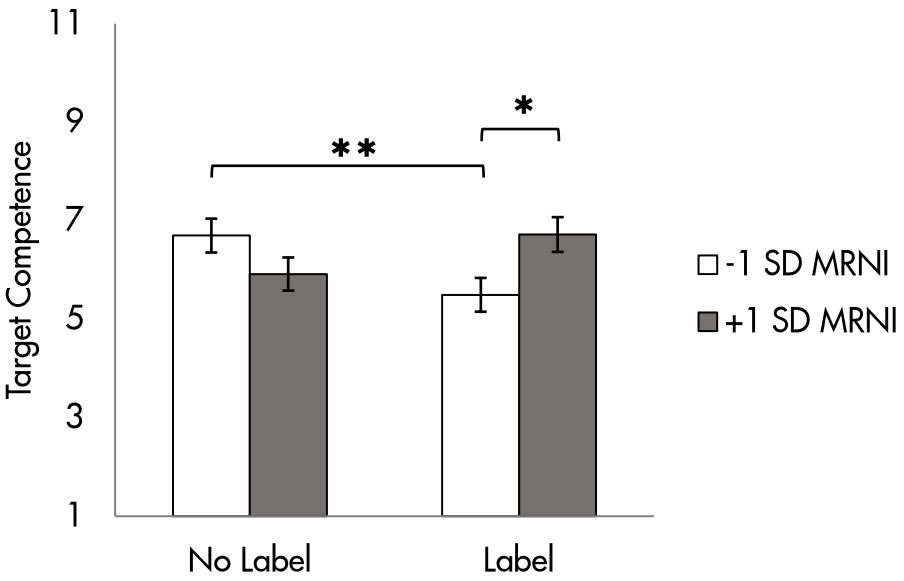


Figure 15. Study 3. Target competence by label condition and masculine ideology for the male target.

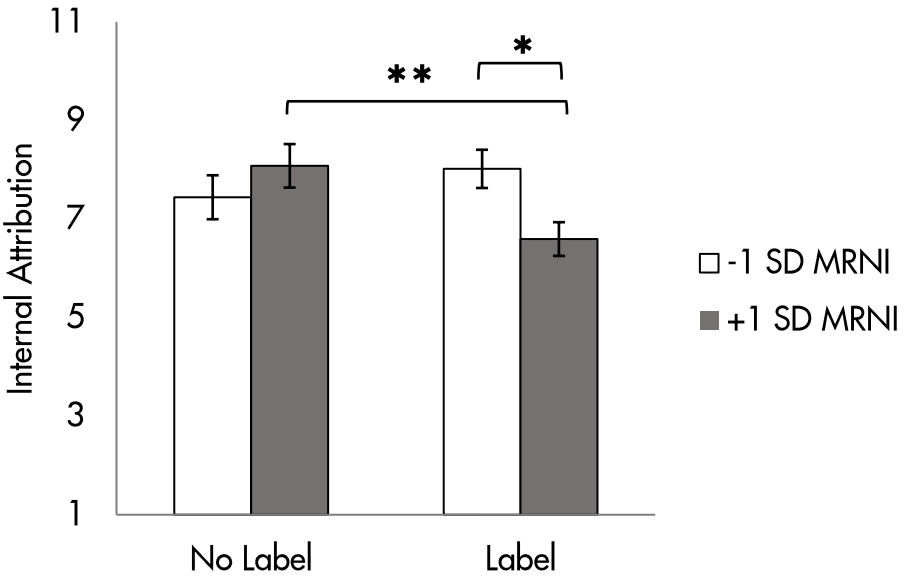


Figure 16. Study 3. Internal attribution by label condition and masculine ideology for the male target.

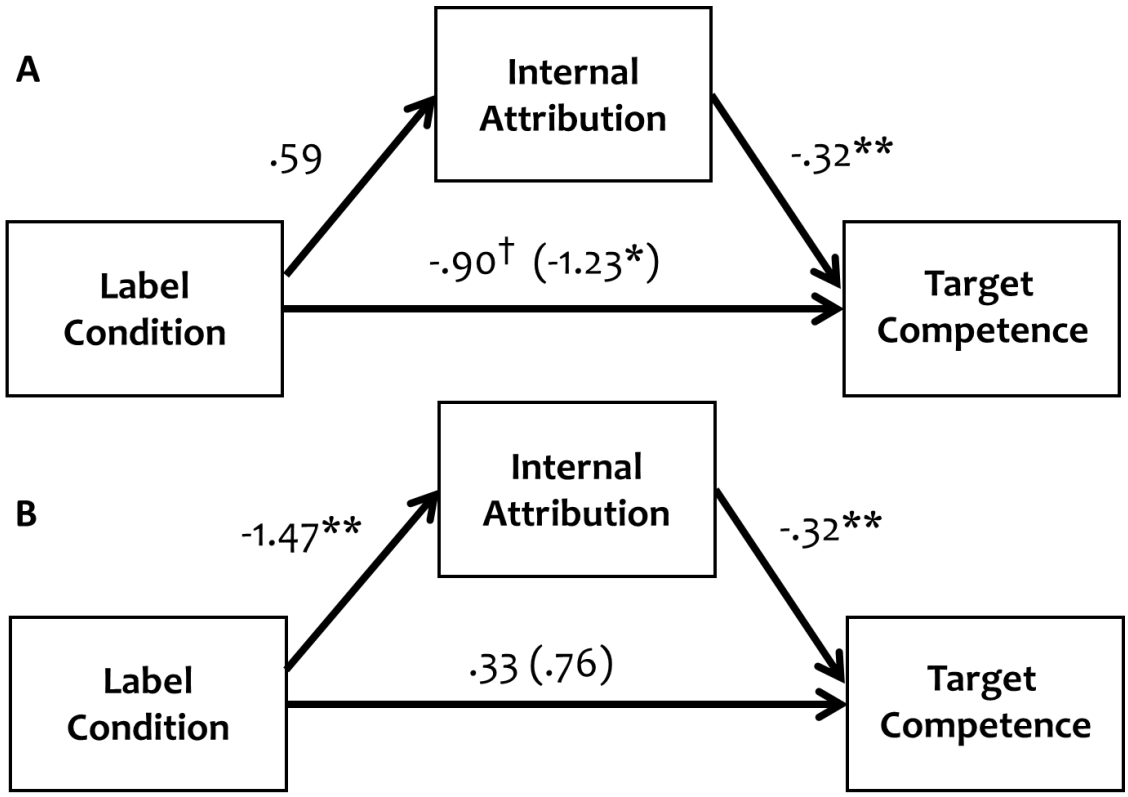


Figure 17. Study 3. Male target only. Internal attribution as a mediator of the relationship between label condition and labeler likability. Estimates calculated at 1 SD below (A) and 1 SD above the mean of MRNI (B).

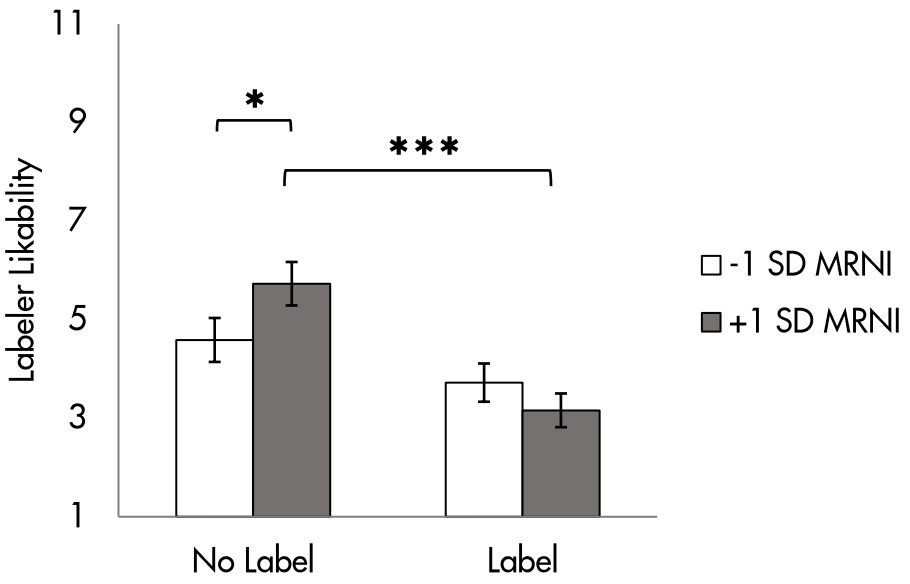


Figure 18. Study 3. Labeler likability by label condition and masculine ideology for the male target.

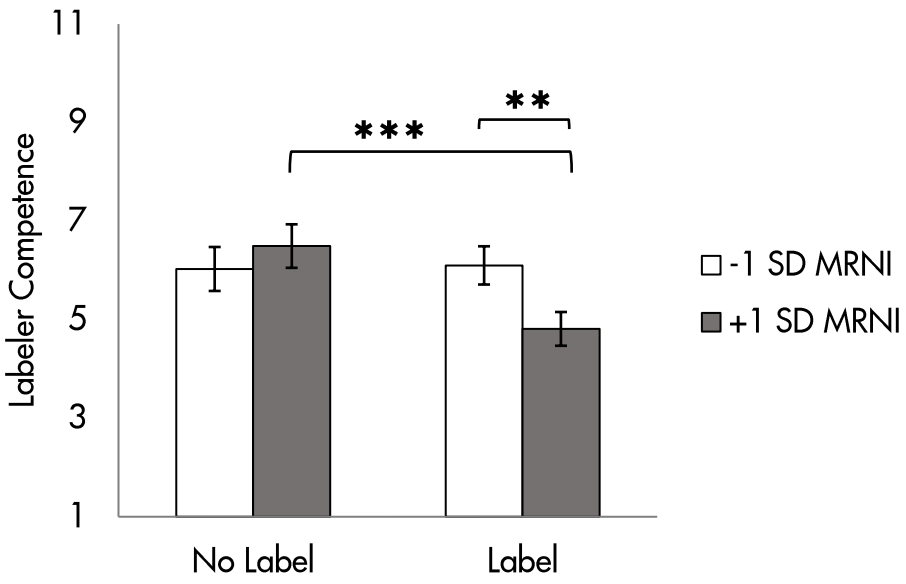


Figure 19. Study 3. Labeler competence by label condition and masculine ideology for the male target.

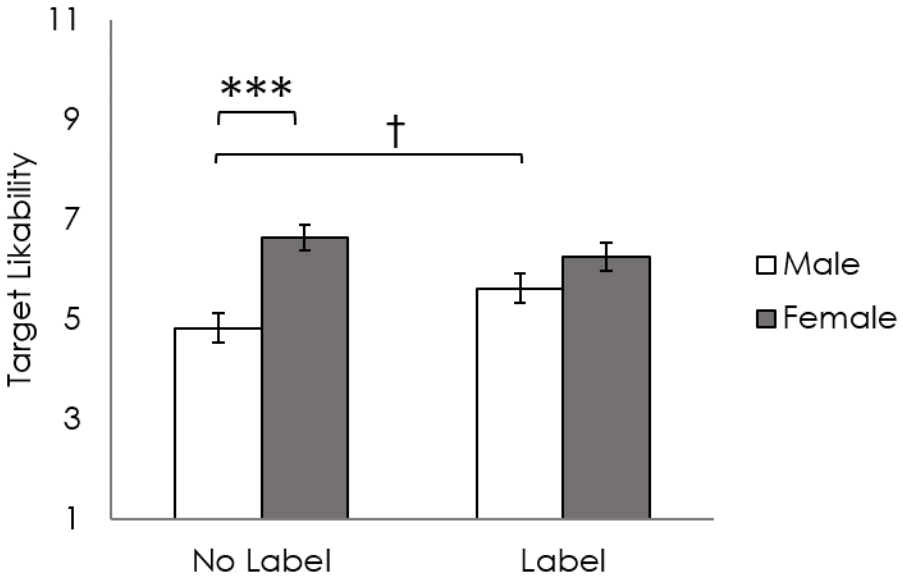


Figure 20. Study 4. Target likability by label condition and target gender.

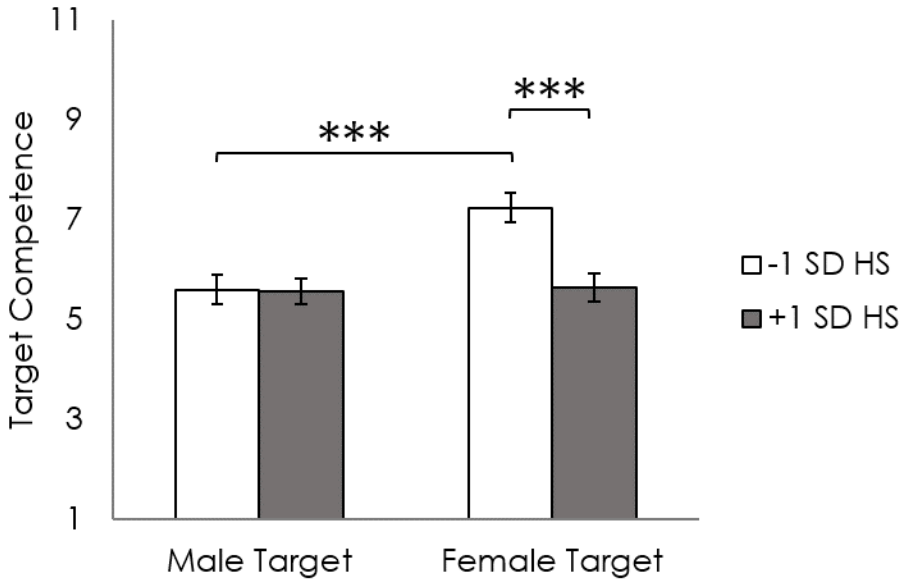


Figure 21. Study 4. Target competence by label condition and target gender.

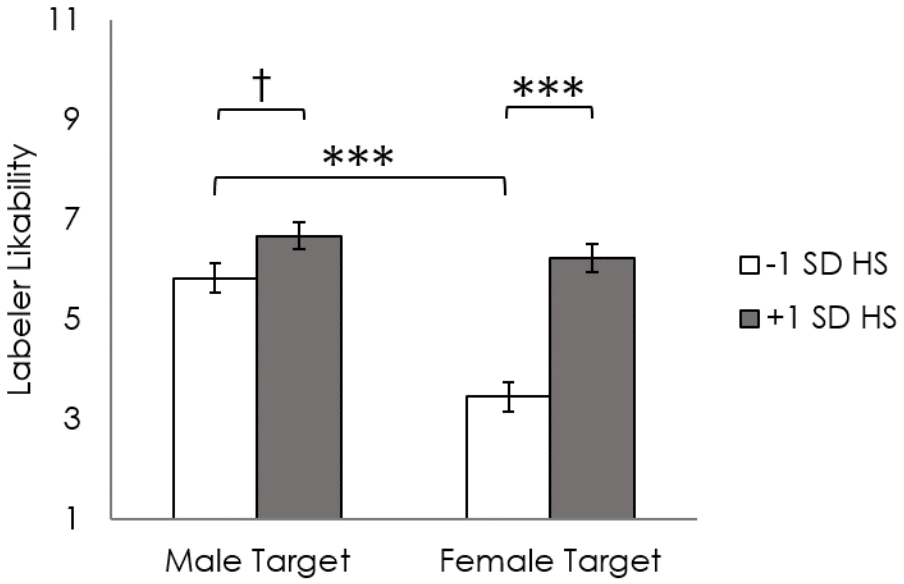


Figure 22. Study 4. Labeler likability by hostile sexism and target gender.

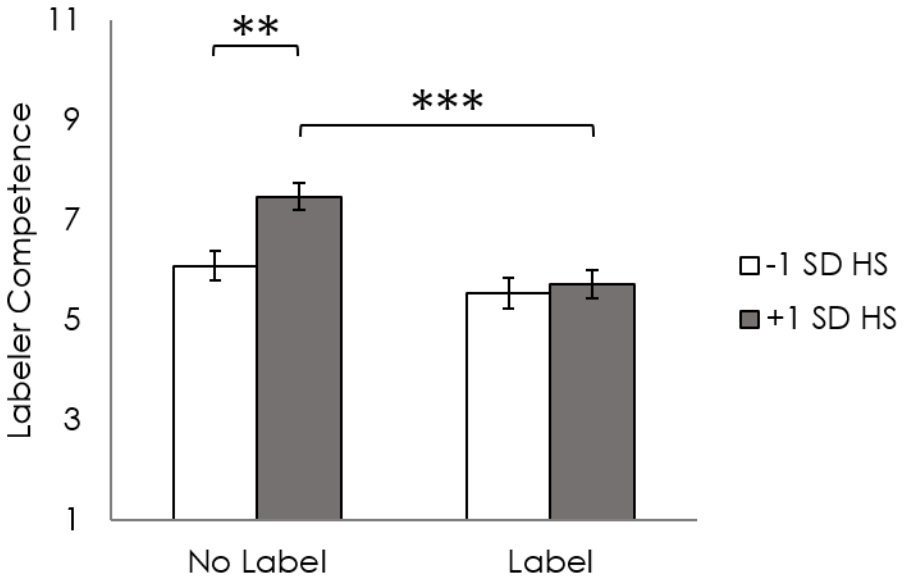


Figure 23. Study 4. Labeler competence by hostile sexism and label condition.



Figure 25. Study 5. Derogatory gendered labels by frequency. The larger the font, the more frequent.

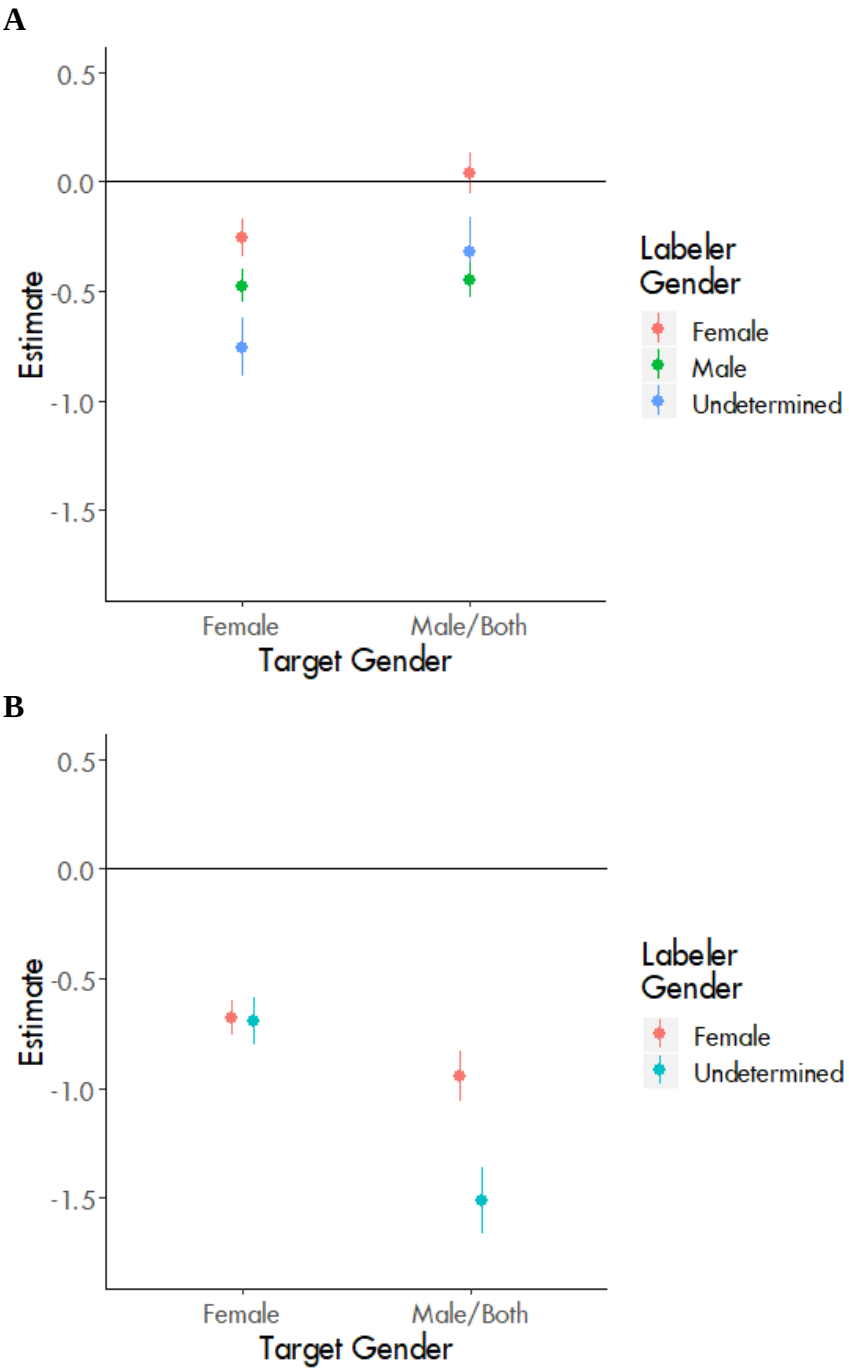


Figure 26. Study 5. Tweet valence by labeler and target gender for (A) Democratic debates and (B) Republican debates.

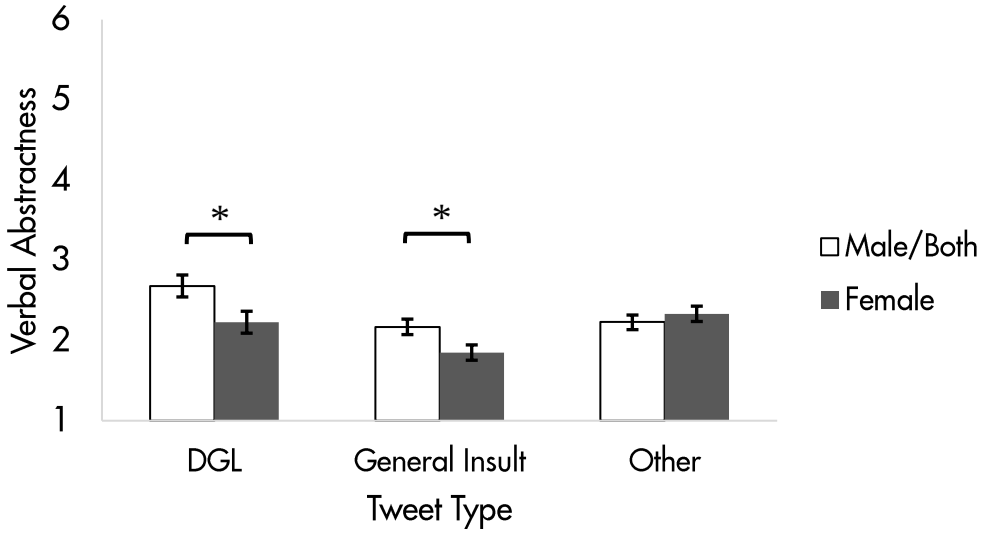


Figure 27. Study 5. Tweet verbal abstractness by target gender and political party of debate.

A)



B)



C)



Figure 28. Study 6. Example tweets (A) Negative; (B) DGL; (C) Positive.

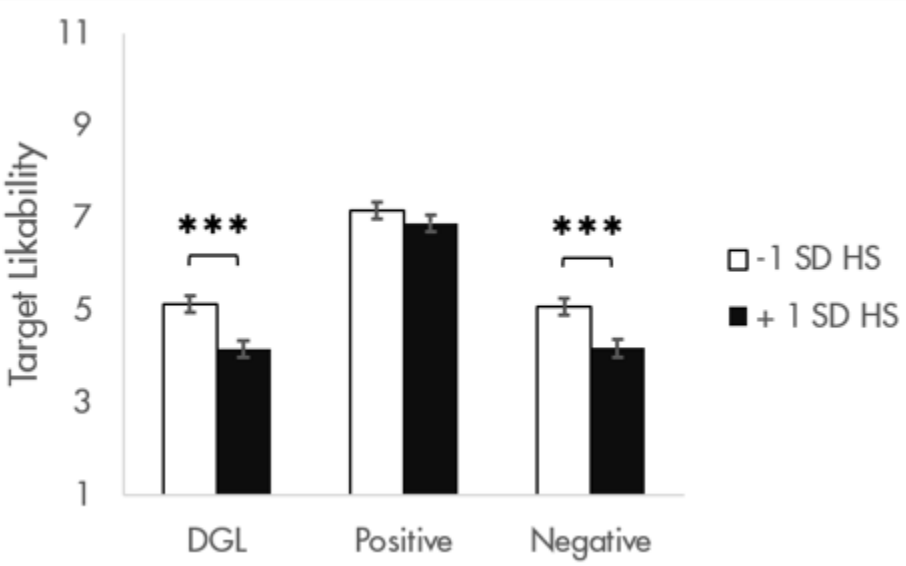


Figure 29. Study 6. Target likability by tweet type and hostile sexism.

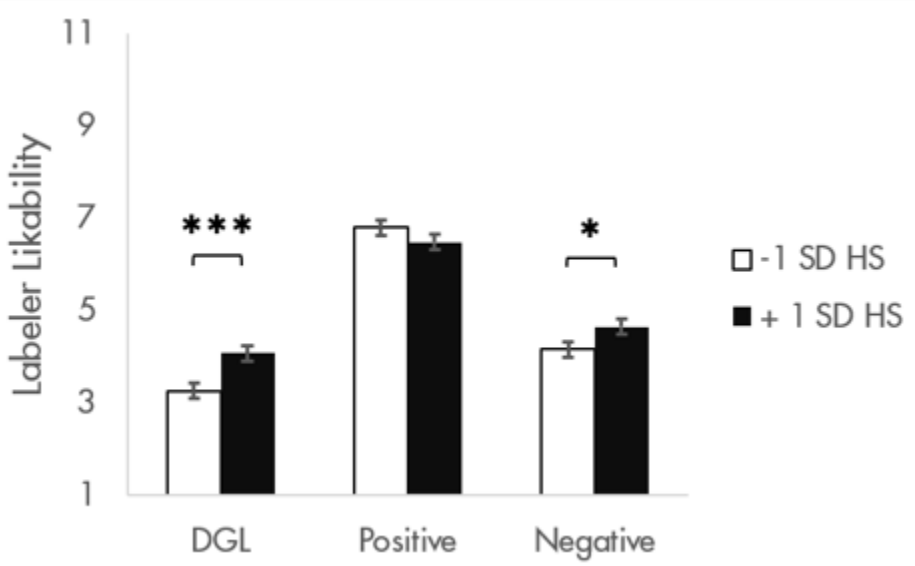


Figure 30. Study 6. Labeler likability by tweet type and hostile sexism.



Figure 31. Study 7. Example of a Negative-Low Abstractness tweet.

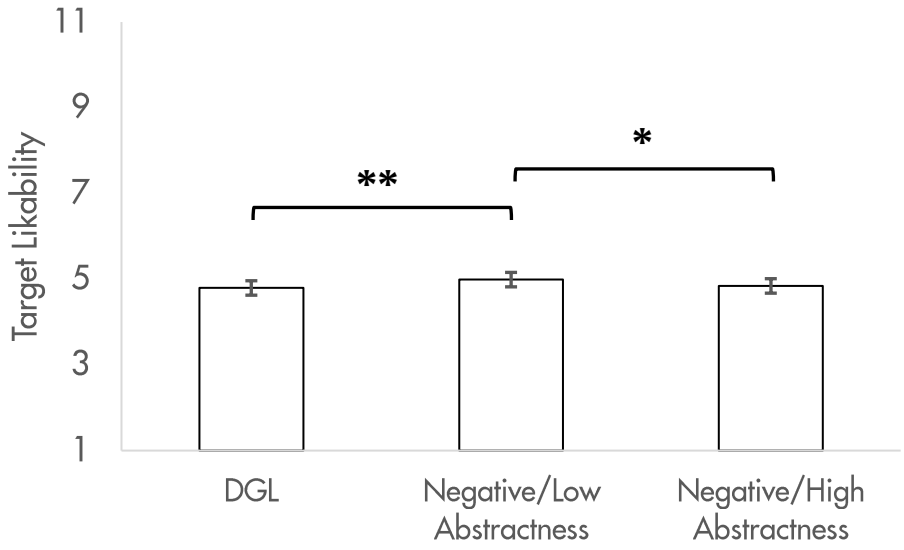


Figure 32. Study 7. Target likability by tweet type.

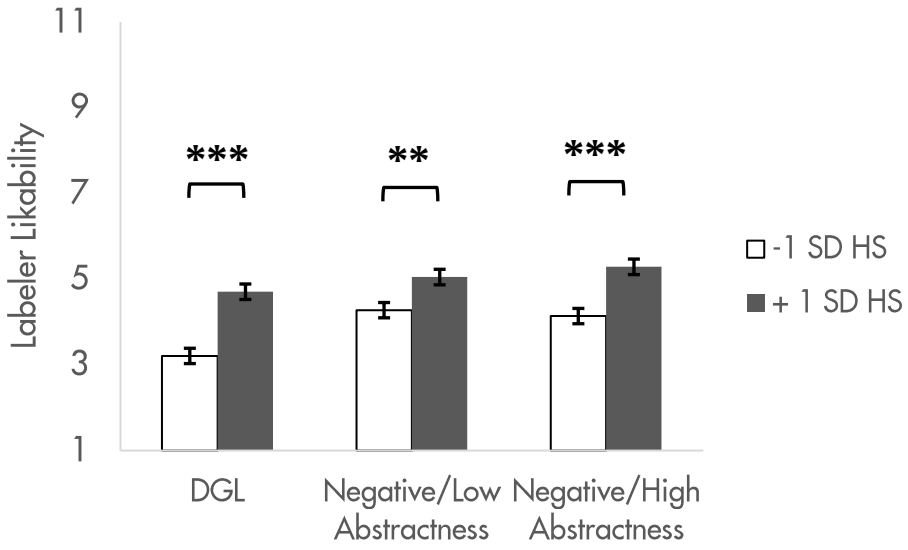


Figure 33. Study 7. Labeler likability by tweet type and hostile sexism.

Appendix A

Coding Instructions

Slurs should be classified as belonging to **one or two** of the categories (assign a one to the slur if it belongs to a category and a zero if it does not). Definitions provided by participants may be used to aid classification, but you should rely on your own understanding of the term.

The category scheme is based on previous research (James, 1998; Saucier et al., 2015). Rather than separate categories by gender (e.g., words that primarily refer to men v. those that primarily refer to women), I decided on categories in which both male-referential and female referential terms can be classified. Note that this makes classification a bit more difficult, as the meaning of terms sometimes depends on what gender they are applied to. Please keep this in mind when performing this task.

Please use the worksheet provided. The spreadsheet contains each slur/insult, accompanied by participants' descriptions (in a single column, separated by a semicolon). There is also a column for any comments or observations that arise when coding the data.

Categories and Descriptions

Homophobic: Homophobic slurs consist of words that challenge or disparage the targeted person's (heterosexual) sexual orientation (e.g., faggot, queer, fairy, dyke, lesbo, butch).

Ethnic: Ethnic slurs consist of words that challenge or disparage the targeted person's race, ethnicity, or culture of origin (e.g., nigger, cracker, redneck, oreo, spic, wigger).

Sexual: Sexual slurs consist of words that refer to the targeted person's sexual behavior, other than sexual orientation. For example, sexual slurs may imply the target is

promiscuous, sexually aggressive, or sexually cold. Note that, when used for men, these terms may not be as derogatory or have quite the same meaning (e.g., horndog, slimeball, stud) as those used primarily for women (e.g., ho, slut, whore).

Strong / Masculine / Aggressive / Mean: These slurs consist of words that imply that the target is strong, masculine, aggressive, and/or mean, other than specifically in regard to sexual behavior (e.g., asshole, bastard, dick). When used for women, these terms often imply aggression particularly towards men (e.g., ballbreaker, cunt, bitch). You might think of this category of words as implying stereotypically masculine traits.

Weak / Feminine / Submissive / Nice: These slurs consist of words that imply that the targeted person is weak, feminine, submissive, and/or (too) nice. When used for men, these terms often imply that the target is weak like a woman (e.g., pussy, sissy, wuss). When used for women, they may be less negative and simply imply that the target is feminine (e.g., priss, sweetheart). You might think of this category of words as implying stereotypically feminine traits.

Unintelligent / Incompetent: Unintelligent / Incompetent slurs consist of words that challenge or disparage the targeted person's intelligence and abilities (e.g., airhead, bimbo, ditz, bonehead, meathead, jackass).

Physical Appearance: Physical Appearance slurs consist of words that focus on the target's physical attractiveness, shape, or stature. Although these may be gendered, many are not (e.g., fat, ugly).

Other: Other slurs consist of words that do not fit into any other category.