



The Power of Language: Gender, Status, and Agency in Performance Evaluations

David G. Smith¹ · Judith E. Rosenstein² · Margaret C. Nikolov³ · Darby A. Chaney⁴

Published online: 3 May 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

In the workplace, women often encounter gender stereotypes and biases that reinforce the existing gender hierarchy, may hinder women's career aspirations and retention, and may limit their ability to be promoted—especially in traditionally male organizations. Long-standing and widely held (although often unconscious) beliefs about gender can reinforce women's perceived lower status position relative to men's. Because men are described/prescribed as agentic (often masculine) and women as communal (often feminine), women leaders are often evaluated as being status-incongruent. We explore the gendered assignment of leader attributes with particular attention to associations of agentic competence (deficiency for women) and agentic dominance (penalty for women). We examined peer evaluations of 4344 U.S. Naval Academy students who are assigned attributes from a predefined list. Although men and women received similar numbers of descriptive (positive) attributes, women received more proscriptive (negative) attributes than did men and these individual attributes were predominantly feminine. These findings offer evidence that women leaders' status incongruity may be associated with perceived competence (agentic deficiency). A contribution of our analysis is theory testing using data from a real-life performance evaluation system. Additionally, our research contributes to our knowledge of gendered language and status characteristics in performance evaluations and can assist researchers and practitioners with developing interventions. Understanding the association of gender status beliefs with evaluation processes may facilitate changing workplace culture to be more gender-inclusive through less biased and stereotypical performance evaluations.

Keywords Gender and leadership · Agentic and communal traits · Stereotype content · Performance appraisal · Military

As women's recruitment and participation in traditionally male occupations increase, research suggests that organizations face challenges in retaining talented women beyond entry-level jobs (Soares et al. 2013; Soyars 2017; Yee et al. 2016). Women in the workforce often encounter gender

stereotypes and biases that reinforce the existing gender hierarchy, which may impede their advancement to higher levels of leadership. Specifically, gender bias and reinforcement of stereotypes in leader performance evaluations may hinder women's career aspirations and retention as well as limit their ability to be promoted.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11199-018-0923-7>) contains supplementary material, which is available to authorized users.

✉ David G. Smith
David.g.smith.phd@gmail.com

¹ U.S. Naval War College, 686 Cushing Road, Newport, RI 02841, USA

² U.S. Naval Academy, 112 Cooper Road, Annapolis, MD 21402, USA

³ U.S. Naval Academy, 572C Holloway Road, Annapolis, MD 21402, USA

⁴ U.S. Marine Corps, Headquarters Battery, 12th Marine Regiment, Okinawa, Japan

In particular, existing gender hierarchies that are predicated on long-standing and widely held gender beliefs may implicitly influence performance evaluations. Gender, as with other social statuses, forms status hierarchies based on the relative values associated with each gender (Berger et al. 1977). For instance, for gender, men are considered higher status and women lower status and for professional position (e.g., leader vs. subordinate), leaders are considered to be higher status, whereas subordinates are lower status. Associated with social status are beliefs that often reinforce the status hierarchy.

Gender status beliefs that reinforce women's perceived lower status position relative to men include stereotype content such as gendered language within performance evaluations for leaders in the form of descriptive and proscriptive

characteristics. *Descriptive* characteristics (generally positive qualities) reinforce who we should be or how we should behave, whereas *proscriptive* characteristics (generally negative qualities) tell us who we should not be or how we should not behave (Prentice and Carranza 2002). These descriptive and proscriptive characteristics may have a gendered, stereotypic quality (i.e., masculine or feminine) such that men and women may differentially receive particular characteristics related to their gender status and role as a leader (Bem 1974; Prentice and Carranza 2002). Given the relative status of role and gender characteristics, leader performance evaluations may implicitly reinforce existing status hierarchies despite efforts to use objective evaluation criteria or meritocratic organizational practices and policies.

To better understand this implicit phenomenon, leader evaluation research examines leader status characteristics based on an agentic-communal dichotomy and finds that agentic characteristics are valued (higher status) whereas communal characteristics are not (lower status) (Abele and Wojciszke 2014; Bakan 1966; Bem 1974; Eagly 1987). Because men are described/prescribed to be agentic and women to be communal, women leaders are often evaluated as being status-incongruent. Women leaders (people of lower gender status in a position of higher status) often receive more proscriptive feedback because they are violating the gender status hierarchy (Rudman et al. 2012). We suggest that female employees in roles that are status-incongruent (e.g., leader) will receive less descriptive and more proscriptive feedback than will male employees who are status-congruent. Further, status-incongruent female employees may receive more masculine proscriptive feedback based on violations of role status and more feminine proscriptive feedback based on gender status. We explore these hypotheses by examining “real world” evaluations of women and men training to be military officers. Because both the leader role and the military are traditionally masculine, men are status-congruent and women are status-incongruent in this domain.

The present research using leader performance evaluations offers contributions to the existing literature on gender, status beliefs, and gender stereotypes in several ways. Whereas most research examining gendered leadership attributions is situated in experimental academic settings or real-world settings with limited access to data, the secondary data analyzed in our research allow us to examine perceptions of leadership performance based on real-world, routine, anonymous evaluations from a military service academy. Not only has the military historically been at the forefront of social change and inclusion, but it is also the largest employer in the United States and thus it is an especially relevant domain in which to assess the potential prevalence of gender bias (Lundquist 2008; Moskos 1993; Sampson and Laub 1996). Our results are consistent with previous experimental findings and support existing theoretical frameworks. Moreover, we are able to

quantify both the breadth and depth of stereotype and gender status beliefs as conveyed in subjective performance evaluations, something not previously feasible without this type of data. Finally, we contribute to the theoretical literature by examining the interrelationship of role status and gender status as distinct bases for evaluating performance.

Gender Stereotypes, Status, and Leadership

The lower retention and advancement of women, especially in traditionally male professions, are often attributed to discrimination and prejudice against women in stereotypically masculine work roles. Stereotypes and expectations about who a leader is “supposed” to be impact how individual leaders are evaluated (Galinsky et al. 2013; Gündemir et al. 2014). These stereotypes and expectations are associated with status characteristics, such as gender (Wagner and Berger 1997). Status characteristics theory (SCT) states that socially significant, salient, and observable characteristics (e.g., gender, race) form status hierarchies based on relative value, competence, and prestige, understood in broadly shared cultural beliefs. Stereotypes and their associated content reflect these cultural beliefs and provide rules for social interaction, evaluation, and judgment. Within the workplace, people with higher status characteristics (e.g., men, Whites) often receive advantage through higher performance expectations, more prestige, and increased influence (Berger et al. 1977).

Status can be ascribed based on individual characteristics (e.g., gender, race, age) or achieved as something that is earned (e.g., leadership position, academic degree, military rank). The gender status hierarchy influences performance expectations such that in higher status positions (e.g., leader), competence (e.g., performance) expectations for women are lower than for men (Ridgeway 2001). Moreover, SCT helps explain why women (lower ascribed status) often lack perceived legitimacy in leadership positions.

Status characteristics are useful in understanding the shared beliefs we hold about who we should be or how we should behave (descriptive) and who we should *not* be or how we should *not* behave (proscriptive), particularly in work roles. When considering a higher status role like leader, descriptive status characteristics are often associated with men and masculinity, whereas proscriptive traits for leaders are often associated with women and femininity. Research on leader performance evaluations using status characteristics is largely based on the agentic-communal dichotomy (Bakan 1966). Agentic behavior is associated with instrumental, task-focused, and goal-oriented characteristics, whereas communal behavior is linked to relationship-oriented, nurturing, and warmth characteristics (Abele and Wojciszke 2014; Bem 1974; Eagly 1987). These agentic qualities and behaviors are characterized as

higher status and communal traits and behaviors as lower status (Abele and Wojciszke 2014).

Agentic leadership qualities and behaviors consist of two distinct constructs related to competence and dominance (Abele and Wojciszke 2014; Rosette et al. 2016). Agentic competence as a leader relates to a person's abilities and skills to lead others in a goal-oriented manner and is typically found in descriptive traits of leaders (Abele and Wojciszke 2014; Rosette et al. 2016). In contrast, agentic dominance is understood as an assertive leader establishing control over others with an emphasis on competitiveness (Abele and Wojciszke 2014; Rosette et al. 2016). Dominance traits of a leader may be either descriptive or proscriptive traits (Rosette et al. 2016; Rudman et al. 2012).

Leader evaluation research examining the competence and dominance constructs finds that women leaders are often evaluated as having either an agentic deficiency (i.e., viewed as lacking competence to be a leader) or an agentic penalty (i.e., penalized for displays of dominance). An agentic deficiency makes it difficult to be hired into or be deserving of a leadership position, and agentic dominance impacts perceived legitimacy in using agentic behaviors in leadership roles (Rosette et al. 2016). Dominance is often perceived as not congruent with stereotypical feminine communality and can lead to perceptions of lack of warmth (not feminine) according to stereotype content research (Eckes 2002; Fiske et al. 2002). Consequently, it is especially challenging for women leaders who must often establish their competence as leaders using agentic characteristics and behavior (Rudman et al. 2012).

The combination of warmth and competence within gender stereotypes reinforces the gender hierarchy for women where they are expected to have lower competence and higher warmth (Eckes 2002; Fiske et al. 2002). The warm but not competent woman (e.g., the “housewife” stereotype) poses no threat to the gender hierarchy. However, a high competence and low warmth woman (e.g., the “career woman” stereotype) challenges the gender hierarchy (Eckes 2002, p. 112). The stereotype that women should not be cold or uncaring (because they should be warm and caring) further penalizes women using a proscriptive characterization. Both dominance and competence as agentic characteristics operate as status maintenance for the gender status hierarchy and provide justification and motivation for penalties in the form of negative performance evaluations or at least less positive evaluations for women compared to men's. This is the double bind women may experience—being negatively evaluated as lacking competence (when perceived as communal) and being too dominant (when perceived as agentic).

Gender status characteristics and stereotype content for military service members are similar to those for leaders. Military leaders are expected to be decisive, independent, confident, and competitive with a command and control style of leadership. These expectations are consistent with those for

male leaders; however, they are inconsistent with expectations for female leaders, who are expected to be helpful, kind, gentle, and emotionally expressive using a participative and collaborative style of leadership (Archer 2013; Boldry et al. 2001; Boyce and Herd 2003; Ebbert and Hall 1993; Francke 1997; Looney et al. 2004; Morgan 2004). Thus, because gender stereotypes for military women are inconsistent with expectations for military leaders, they may contribute to women leaders' negative performance evaluations.

The U.S. Military as a Case Study

The U.S. military offers an ideal environment for directly examining the relationship among gender, status, and stereotype content because it is a social institution that has long been considered a vanguard of social change and has institutionalized role expectations and a formal performance evaluation system (Atkinson 2015; Lundquist 2008). Epitomizing masculine-type work, the military was, until recently, highly gender-segregated limiting women's ability to compete with men on equal footing (Pellerin 2015; Segal et al. 2016). Despite the military's gender integration efforts over the last 40 years, men represent 84% of the active duty forces and are retained at almost twice the rate as women in combat specialties (U.S. Department of Defense 2016). Beyond representation and the type of work, military culture reinforces a hyper-masculine identity with the ideal warrior being brave, unemotional, fit, and ready to fight (Archer 2013; Barrett 1996). This ideal masculine warrior is socialized through basic training and everyday military life, and this paradigm may influence leadership styles or perceptions.

Institutional military structure is premised on an “up-or-out” career model, whereby one is either promoted or separated (i.e., not retained) (Rosen 1992). Evidence suggests that military women and men perform similarly in relevant training and other objective measures such as awards, physical fitness scores, grade point average, military science grades, and rankings (Biernat et al. 1998; Boldry et al. 2001). The military is perceived to be meritocratic because promotion is largely based on expertise and competency; however, subjective factors are also highly relevant (Atkinson 2015; Lundquist 2008).

Our research examines whether stereotype content and status incongruity arise in subjective performance evaluations of those training to be military leaders at the United States Naval Academy (USNA). USNA is one of three U.S. military service academies (the others are the United States Air Force Academy and the United States Military Academy [Army]), which are major military accession sources in addition to Reserve Officer Training Corps (ROTC) and Officer Candidate School (OCS). Military service academies are four-year public colleges where students graduate with

Bachelor of Science degrees. Students receive military, physical, and character training in preparation for commissions as military officers in their respective services. The four-year leader education and development programs are expressly designed to indoctrinate and socialize students into the military profession, with the hope that they will make it a career. For all four years at USNA, students (“Midshipmen”) both work and live in their professional units (“companies”), which results in minimal separation between professional and personal lives and results in students getting to know members of their company on a personal and professional level. The work includes company leadership and organization, dissemination of information from Academy leadership, military leadership training, counseling and guidance, among others. As part of the leadership development process, students are evaluated on their professional competence by superiors and peers in their professional units.

Performance evaluation data on Midshipmen offer a unique opportunity to examine the relationship between gender and assigned leadership characteristics and to assess evidence-based organizational practices. For men and women who are broadly similar with respect to academic standards, physical standards, and military standards, we would expect similar evaluations in the absence of biases and stereotypes. However, theory suggests that gender status beliefs may penalize women in traditionally masculine roles who violate the gender status hierarchy. Peer and upper-class application of subjective leadership characteristics offer insight into how status incongruity and related penalties may be reflected in performance evaluations. Based on previous research, we anticipate that in this military leadership context women will be evaluated differently and more harshly than their male peers will be.

The Present Study

We hypothesize that women training to be military leaders will be perceived as status-incongruent based on their gender and leader role and that this incongruity will be observed in their performance evaluations. Based on SCT, evidence of status incongruity will be associated with women receiving fewer descriptive and more proscriptive characteristics than men (Hypothesis 1). As leaders, men are status-congruent (higher status for gender and role) and expected to be agentic (competent and dominant), whereas women are not status-congruent or expected to be agentic. Therefore, when we consider the gendered component of these characteristics, we anticipate that men will receive more masculine descriptive characteristics whereas women will receive more feminine descriptive characteristics (Hypothesis 2). Also, because women leaders are status-incongruent (lower gender status and higher role status), they will receive more feminine proscriptive

characteristics (agentic deficiency) and masculine proscriptive leadership characteristics (agentic penalty) than men will (Hypothesis 3).

Additionally, analysis at the level of individual characteristics enables us to explore how agentic competence and agentic dominance may be attributed to our participants. We expect that *individual* descriptive and proscriptive characteristics will be assigned consistent with Hypothesis 2 and Hypothesis 3. Building on the theoretical underpinnings of SCT, we explore leader agency through assignment of specific characteristics in performance evaluations. Specifically, we expect that men will be more likely to receive individual masculine descriptive characteristics (i.e., agentic competence and agentic dominance) and that women will be more likely to receive individual feminine descriptive characteristics (i.e., communal) (Hypothesis 4). Finally, we expect that women will be more likely than men will be to receive individual feminine and masculine proscriptive characteristics (Hypothesis 5).

Method

Leader Evaluation Process

Data were drawn from the Midshipmen Aptitude for Commissioning system and merged with demographic and performance (military, academic, and physical) measures drawn from an institutional database with approval from USNA’s Institutional Review Board. Students evaluate one another in multiple ways using the Academy’s Midshipman Aptitude for Commissioning system (this is akin to a 360 degree feedback system in professional context, albeit without subordinate input). At the end of each semester students are required to anonymously rank all of their classmates within their company (i.e., approximately 40 peers per class year) who are of the same year or younger, placing each person into a quintile. Then they must determine who the top three performers in the top quintile are, as well as identify the bottom three performers in the bottom quintile. For each of these six individuals they must make a single selection (one attribute) from a predetermined list of 44 positive and 45 negative characteristics that best describes the individual’s professional and leadership traits (hence referred to as “leadership attributes”). (They may provide leadership attributes for other students as well, but it is not required.)

The leadership attributes available for selection are presented in a single alphabetical list complete with descriptions to raters. (See [online supplement](#) for the complete list including definitions and valence.) The rankings and leadership attributes are intended to capture how other students perceive the target and are largely subjective (United States Naval Academy 2016). Students understand that these evaluations are influential in the assignment of student leadership

positions in conjunction with objective measures (e.g., grades, fitness scores, class standing). Although the evaluations are important and the institution wants students to take the process seriously and provide responses that accurately reflect their observations, it is unclear on exactly what information and criteria students base their evaluations and, for some, it may be more about popularity than a professional evaluation.

Participants

We obtained data on all students (evaluatees) enrolled at USNA in the Spring semester of the 2014–2015 academic year. Because evaluators were anonymous, their demographic data, including gender, were unavailable for this analysis. We excluded students studying abroad ($n = 23$) and students who were foreign nationals or visiting USNA for the semester ($n = 82$).

Many students in high-level leadership positions (called “stripers” because of the insignia they wear) were missing peer evaluation data because they were not ranked by their peers at the end of the semester due to the nature of their positions away from their companies. However, since striper assignment is partially based on class standing, the omission of these stripers might meaningfully skew results. Therefore, for those stripers without peer performance data for the Spring semester, we imputed rankings and attributes from the previous semester ($n = 63$). Forty-five students with striper positions both Spring and Fall semesters had no peer performance measures either semester and were dropped. The resulting dataset comprises 4344 students.

Men composed more than three-quarters of the student body ($n_{\text{men}} = 3349$, 77%; $n_{\text{women}} = 995$, 23%). A majority of the student body was White ($n = 2841$, 65%), with 482 (11%) Hispanic, 304 (7%) African American, 293 (7%) Asian American, 21 (.5%) Native Hawaiian/Pacific Islander, 20 (.5%) Native American, and 383 (9%) “Other.” Classes were approximately equally distributed, with 1031 (24%) seniors, 1055 (24%) juniors, 1114 (26%) sophomores, and 1144 (26%) first years and with differences largely due to attrition. Age was excluded from analysis because all students are required to matriculate by age 23 and graduate by age 27. Although everyone at the Academy must participate in athletics, only about a quarter were varsity athletes ($n = 1108$, 26%), whereas the remainder were involved in intramural and club sports.

Measures

Descriptive and Proscriptive Attributes

The Midshipman Aptitude for Commissioning system identifies 89 leadership attributes Midshipmen can ascribe to one another, and it explicitly assigns them a valence in the context of leadership at the Naval Academy. Because the attributes identified as “positive” are consistent with descriptive

leadership traits, and the attributes identified as “negative” are consistent with proscriptive leadership traits, we labeled the 44 positive attributes “descriptive” and the 45 negative attributes “proscriptive.” The analysis considers the descriptive and proscriptive attributes together because all 89 attributes are available for selection when assigned by evaluators. However, to interpret the assignment of these as distinct descriptive and proscriptive categories, we examine the type of attribution separately.

Feminine and Masculine Attributes

To address gendered assignment of leadership attributes, attributes were labeled as feminine, masculine, or neutral based on gender assignment derived from earlier research (e.g., Bem Sex Role Inventory, Personal Attributes Questionnaire; see [online supplement](#)). Both an undergraduate research assistant and the second author reviewed previous literature for how characteristics were coded. Where our attributes mapped directly onto characteristics identified previously, we used the gender assignment from that research. Where attributes did not map directly, we looked for a closely identified term and used its gender assignment in conjunction with the institutionally provided definition (e.g., “apathetic” is one of our attributes, but does not appear in the research we examined; however, Prentice and Carranza 2002, have “detached,” which we used as a synonym for “apathetic”). If there was disagreement or it was unclear how to code an attribute based on pre-existing literature, we labeled the attribute as neutral. All attribute labeling was reviewed by the first author.

It is worth emphasizing that the attribute gendering is distinct from the descriptive/proscriptive nature of the leadership characteristics. Of the 44 descriptive leadership attributes, 11 were characterized as masculine (analytical, athletic, competent, confident, courageous, decisive, inspiring, logical, practical, proactive, and resourceful), 15 feminine (charismatic, civil, compassionate, dependable, diplomatic, enthusiastic, honest, intuitive, loyal, mature, organized, polished, respectful, self-aware, and team-player), and 18 neutral (articulate, candid, dedicated, diligent, energetic, ethical, industrious, innovative, judicious, level-headed, methodical, principled, resilient, responsible, self-disciplined, self-reliant, thorough, and versatile). Of the 45 proscriptive leadership attributes, 18 were characterized as masculine (abrasive, abusive, apathetic, arrogant, blunt, careless, confrontational, disorganized, egocentric, forgetful, inconsiderate, lethargic, opportunistic, overbearing, ruthless, selfish, sloppy, and stubborn), 10 feminine (excitable, frivolous, gossip, indecisive, inept, panicky, passive, scattered, temperamental, and unpredictable), and 17 neutral (argumentative, complacent, impetuous, inattentive, incurious, indifferent, irresponsible, lackadaisical, mistrustful, sarcastic, sleepy, uncommitted, unprincipled, unproductive, untruthful, vague, and vain).

Counts, Proportions, and Relative Frequencies

The leadership evaluation process produces count data—the number of times each Midshipman was assigned each attribute. For instance, if a Midshipman has a count of 3 for a given attribute (e.g., analytical), then this Midshipman was characterized as such by three other Midshipmen in the company. For each Midshipman in our dataset, we have the number of times she or he was assigned each of the 89 leadership attributes. Broadly (i.e. across attributes), we consider the counts (e.g., the total number of descriptive assignments and the total number of proscriptive assignments). When considering the attributes individually, however, we consider the counts *relatively*; specifically, we examine: (a) breadth (or diversity) of the attribute as indicated by the proportion of the population (4344 Midshipmen [3349 men and 995 women]) to ever receive the attribute (1 = at least once, 0 = never) and (b) depth (or intensity) of the attribute as indicated by the frequency of assignment of that attribute relative to the other 88 attributes (81,774 total attribute assignments [51,699 descriptive and 30,075 proscriptive]). The key distinction in these measures is the denominator, where the denominator for the proportions is the total number of Midshipmen, whereas the denominator for the relative frequencies is the total number of attribute assignments. The proportions and relative frequencies allow us to identify differences, respectively, in how widely an attribute is used to describe the population and how often a particular attribute is used relative to others.

Results

The men and women in our study were comparable in military performance with respect to their cumulative military grade point average ($M_{\text{men}} = 3.18$, $SD = .36$; $M_{\text{women}} = 3.17$, $SD = .38$, on a scale of 0 to 4.0; $p = .354$) and company military ranking ($M_{\text{men}} = 18.5$, $SD = 10.64$; $M_{\text{women}} = 18.3$, $SD = 10.65$, on a scale of 1 to 41; $p = .558$). Therefore, consistent with previous research (e.g., Biernat et al. 1998; Boldry et al. 2001), something other than objective performance presumably accounts for gender differences in the subjective performance evaluations and attribute assignment.

In our discussion of the results that follows, it is important to understand the distinction between attributes and attribute assignments. Suppose a target in our study received analytical twice, competent three times, and all other attributes zero times. This target received two attributes and five attribute assignments. We consider the first measure as the diversity, or variety, of attributes received and the second measure as the intensity of attribute assignment—how, in this case, “positively” the target is viewed.

Given that Midshipmen are required to assign leadership attributes only to the top three and bottom three in their

ranking, one might assume that many Midshipmen are not assigned any attributes. This is not the case. The percentage of men who never received a descriptive attribute was 1.5% compared to 1.6% of women. The percentage of men who never received a proscriptive attribute was 13% compared to 7.4% of women.

The number of descriptive attribute assignments and the number of proscriptive attribute assignments received by men and women were compared using the Wilcoxon rank sum test. The men and women in our study did not differ significantly with respect to the number of descriptive assignments received ($Mdn_{\text{men}} = 10$, $Mdn_{\text{women}} = 9$, $p = .098$); however, the men received significantly fewer proscriptive assignments compared to the women ($Mdn_{\text{men}} = 4$, $Mdn_{\text{women}} = 5$, $p < .0001$), providing partial support of Hypothesis 1.

The numbers of gendered (masculine, feminine, neutral) descriptive and proscriptive attribute assignments received by men and by women were also compared, again using the Wilcoxon rank sum test. The significance level used to test for differences in these comparisons was set at $\alpha^* = .05/6 = .0083$ (standard overall $\alpha = .05$ Type I error rate with a Bonferroni correction to adjust for the six comparisons). Under the attribute classification described previously, there were significant differences in the number of assignments to men and women for masculine descriptive attributes ($Mdn_{\text{men}} = 3$, $Mdn_{\text{women}} = 2$, $p < .0001$) and feminine descriptive attributes ($Mdn_{\text{men}} = 3$, $Mdn_{\text{women}} = 4$, $p < .0001$), providing support for Hypothesis 2. Moreover, whereas women received significantly more feminine proscriptive attributes than did men ($Mdn_{\text{men}} = 0$, $Mdn_{\text{women}} = 1$, $p < .0001$), women did not receive significantly more masculine proscriptive attributes ($Mdn_{\text{men}} = 2$, $Mdn_{\text{women}} = 2$, $p = .010$), providing partial support for Hypothesis 3.

Turning attention now to the individual attributes, the proportion of Midshipmen who were assigned a given attribute (at least once) and the relative frequency of assignment of that attribute were computed separately for men and women. Fisher’s exact test (a small sample alternative to the Chi-square test) was used to test for gender differences in both the proportion and the relative frequency of the individual attributes with significance level $\alpha^* = .05/89 = .00056$ (overall $\alpha = .05$ Type I error rate with a Bonferroni correction to adjust for multiple comparisons). Effect sizes for proportions were calculated using Cohen’s h (an analog to Cohen’s d for proportions) (Cohen 1988). For example, the proportion of men assigned the attribute analytical was .580, whereas the proportion of women assigned analytical was .482; this difference is statistically significant ($p < .0001$, Cohen’s $h = .197$). The relative frequency of analytical also differs significantly by gender ($p < .0001$), where, relative to the other attributes, analytical is used with greater frequency for men than for women (.0598 versus .0430, respectively). (Complete results are available in the [online supplement](#).)

Table 1 summarizes the gender differences in the proportion of Midshipmen ever assigned the individual attributes, categorized by the gendering of the attributes. Statistically significant gender differences were detected on 11 of the 44 descriptive attributes (see Table 1a) and on 10 of the 45 proscriptive attributes (see Table 1b). In partial support of Hypothesis 4, men were more likely to receive 5 masculine descriptive attributes (analytical, competent, athletic, logical, and practical), 1 neutral descriptive attribute (level-headed), and none of the feminine descriptive attributes; whereas women were more likely to receive 1 masculine descriptive attribute (proactive), 1 neutral descriptive attribute (energetic), and 3 feminine descriptive attributes (compassionate, enthusiastic, and organized). In support of Hypothesis 5, women were more likely than men to receive all 10 of the proscriptive attributes for which there was a statistically significant gender difference (selfish, vain, inept, frivolous, gossip, excitable, scattered, temperamental, panicky, and indecisive).

Table 2 shows a similar pattern in gender differences for the relative frequency of attribute assignment. Statistically significant gender differences were detected on 14 of the 44 descriptive attributes (see Table 2a) and on 14 of the 45 proscriptive

attributes (see Table 2b). These results are mainly consistent with the findings for proportions (i.e., Table 1), although there are more attributes that are significant. The fact that we find more significant gender differences among attributes when evaluating the relative frequencies is not surprising when we recognize that a single Midshipman who receives a single attribute many times can strongly impact the relative frequency (attribute counted multiple times), but not the proportion (individual counted once).

As shown in Table 2a, men received 10 descriptive attributes with greater relative frequency, 6 of which are masculine (analytical, competent, athletic, confident, logical, and practical), 3 neutral (versatile, articulate, and level-headed), and 1 feminine (dependable). Women received 4 descriptive attributes with greater relative frequency, of which none are masculine, 1 is neutral (energetic), and 3 are feminine (compassionate, enthusiastic, and organized). Table 2b shows the corresponding results for the proscriptive attributes. Of the 14 proscriptive attributes for which there was a significant gender difference, women received 12 with greater relative frequency (selfish, opportunistic, vain, inept, frivolous, passive, scattered, gossip, excitable, panicky, temperamental, and

Table 1 Significant gender differences in proportions: Descriptive and proscriptive attributes

Attribute	Men receive more		Women receive more	
	Men vs. Women		Men vs. Women	
Type	Attribute	Proportions (<i>p</i>)	Attribute	Proportions (<i>p</i>)
(a) Descriptive attributes				
Masculine	Analytical	.580 vs .482 (< .0001)	Proactive	.082 vs .120 (.0004)
	Competent	.502 vs .401 (< .0001)		
	Athletic	.380 vs .290 (< .0001)		
	Logical	.084 vs .048 (.0001)		
	Practical	.067 vs .035 (.0001)		
Neutral	Level-headed	.195 vs .126 (< .0001)	Energetic	.128 vs .195 (< .0001)
Feminine			Compassionate	.319 vs .562 (< .0001)
			Enthusiastic	.156 vs .257 (< .0001)
			Organized	.077 vs .148 (< .0001)
(b) Proscriptive attributes				
Masculine			Selfish	.151 vs .201 (.0003)
Neutral			Vain	.067 vs .122 (< .0001)
Feminine			Inept	.206 vs .296 (< .0001)
			Frivolous	.128 vs .239 (< .0001)
			Gossip	.057 vs .223 (< .0001)
			Excitable	.066 vs .146 (< .0001)
			Scattered	.070 vs .119 (< .0001)
			Temperamental	.039 vs .105 (< .0001)
			Panicky	.037 vs .096 (< .0001)
			Indecisive	.018 vs .042 (< .0001)

The denominator for these proportions is the total number of Midshipmen: $n_{\text{men}} = 3349$, $n_{\text{women}} = 995$. Fisher's exact test, $\alpha^* = .05/89 = .00056$ (overall $\alpha = .05$ Type I error rate with a Bonferroni correction to adjust for multiple comparisons)

Table 2 Significant gender differences in relative frequency: Descriptive and proscriptive attributes

Attribute	Men receive more		Women receive more	
	Men vs. Women		Men vs. Women	
Type	Attribute	Relative frequencies (<i>p</i>)	Attribute	Relative frequencies (<i>p</i>)
(a) Descriptive attributes				
Masculine	Analytical	.060 vs .043 (<.0001)		
	Competent	.056 vs .038 (<.0001)		
	Athletic	.046 vs .029 (<.0001)		
	Confident	.019 vs .014 (<.0001)		
	Logical	.005 vs .003 (<.0001)		
	Practical	.004 vs .002 (.0001)		
Neutral	Versatile	.016 vs .012 (.0003)	Energetic	.009 vs .013 (<.0001)
	Articulate	.013 vs .010 (<.0001)		
	Level-headed	.013 vs .007 (<.0001)		
Feminine	Dependable	.030 vs .022 (<.0001)	Compassionate	.027 vs .066 (<.0001)
			Enthusiastic	.011 vs .017 (<.0001)
			Organized	.005 vs .009 (<.0001)
(b) Proscriptive attributes				
Masculine	Arrogant	.030 vs .017 (<.0001)	Selfish	.011 vs .015 (<.0001)
Neutral	Irresponsible	.007 vs .004 (<.0001)	Opportunistic	.005 vs .007 (.0005)
			Vain	.004 vs .009 (<.0001)
Feminine			Inept	.023 vs .041 (<.0001)
			Frivolous	.009 vs .018 (<.0001)
			Passive	.006 vs .009 (.0001)
			Scattered	.005 vs .008 (<.0001)
			Gossip	.004 vs .020 (<.0001)
			Excitable	.004 vs .010 (<.0001)
			Panicky	.003 vs .007 (<.0001)
			Temperamental	.003 vs .006 (<.0001)
			Indecisive	.001 vs .002 (<.0001)

The denominator for these relative frequencies is the total number of attribute assignments. Number of descriptive attribute assignments: $n_{\text{men}} = 40,354$, $n_{\text{women}} = 11,345$. Number of proscriptive attribute assignments: $n_{\text{men}} = 22,130$, $n_{\text{women}} = 7945$. Fisher's exact test: $\alpha^* = .05/89 = .00056$ (overall $\alpha = .05$ Type I error rate with a Bonferroni correction to adjust for multiple comparisons)

indecisive). Only 2 proscriptive attributes (arrogant and irresponsible) were assigned with greater relative frequency among men. Again, there is strong support for Hypothesis 5.

Discussion

Based on SCT and status incongruity we predicted that men and women would receive different performance evaluations. Our results show that overall, women received more proscriptive leadership attributes than men do, but a similar number of descriptive leadership attributes (Hypothesis 1). Within the descriptive leadership attributes, we found that women received more feminine attributes and men received more masculine attributes (Hypothesis 2). However, for proscriptive leadership attributes, women received more feminine

attributes while receiving a similar number of masculine attributes (Hypothesis 3). We also found significant gender differences in the individual descriptive attributes, with women more likely to receive 5 attributes (1 masculine, 1 neutral, 3 feminine) and men more likely to receive 6 attributes (5 masculine, and 1 neutral) (Hypothesis 4). As for individual proscriptive attributes, women were more likely to receive 10 attributes (1 masculine, 1 neutral, and 8 feminine) (Hypothesis 5). Consistent with prior meta-analytic research on gender differences in cognitive, communication, and social and personality variables (Hyde 2005), our effect sizes (Cohen's *h*) were relatively small with few exceptions. Although the effect sizes may seem to be small, these differences can result in practical importance in the workplace. Indeed, over time, research shows that small biases against women in performance evaluations can cumulatively result

in large disparities in gender diversity at senior leadership levels (Martell et al. 1996).

We found support for SCT, in that men's higher ascribed gender status is congruent with their higher role status (leader) and women were evaluated as incongruent as lower gender status and higher role status (leader). Examining the collective and individual leadership attributes, we found that men were more likely to receive 6 of the 29 masculine/neutral descriptive attributes and none of the feminine descriptive attributes. Furthermore, not only were women more likely than men to receive only 2 of the masculine/neutral descriptive attributes, but they were also more likely to receive 3 of the 15 feminine descriptive attributes. SCT is further supported by our finding that women were more likely than men were to receive all 10 of the proscriptive attributes for which there was a statistically significant gender difference.

Because the majority of the proscriptive leadership attributes women were more likely to receive were feminine (8 of the 10), it appears that these women may have been evaluated more often on competence (agentic deficiency) than on dominance (agentic penalty). Consistent with previous research, this pattern might imply that these women employ a stereotypical feminine leadership style (communal). If this is the case, it could explain why the women in our study were more likely to be characterized as *inept* because it implies an unspoken questioning of their competence.

Although we hypothesized that women would also receive masculine proscriptive attributes more than men, there was little support for this prediction in our data. This may suggest that either these women are not employing an agentic leadership style or that the dominance penalty is not as prevalent in this context. According to SCT, women who lead using greater agency (dominance) are more likely to receive backlash (agentic penalty) in the form of proscriptive attributes (e.g., abrasive, abusive, argumentative, arrogant, or confrontational) emphasizing the masculine authority they have usurped. With the exception of selfish, no other masculine terms were received more by women.

Because women were generally more likely to receive feminine descriptive and proscriptive leadership attributes, we considered the possibility that evaluators attempt to maintain the gender status hierarchy by evaluating women using attributes that emphasize what women are not—stereotypical masculine leaders. Of note, compassionate was the most commonly assigned attribute of any type to be given to women. Compassionate is a desirable leadership attribute for any leader, regardless of gender, yet it is a characteristic that is generally more associated with women leaders than with men leaders (Parker et al. 2015). Similarly, the leadership attribute, organized, was assigned to women more than to men (Parker et al. 2015). Thus, there is evidence that feminine leadership attributes are being assigned in a way that is consistent with maintenance of the gender status hierarchy.

Our results suggest that women in the military may face a more subtle version of the double bind. Only one masculine proscriptive attribute, selfish, was assigned more often to women whereas we expected more penalties for agentic dominance in the military context. Instead, women were assigned more feminine proscriptive leadership attributes (inept, frivolous, gossip, and excitable), which may be a penalty for being perceived as communal. Of note, the neutral proscriptive leadership attribute, vain, was also more likely to be assigned to women. Personal appearance in the military context is valued and emphasized in terms of professional appearance in uniform. However, women whose personal appearance is observed to be more feminine or somehow overtly enhanced (e.g., cosmetic make-up, nail polish, hairstyle) in ways that may make them feel more professional, could draw attention to their femininity and therefore be evaluated as incongruent with the leader role. We also acknowledge that vain may not be properly categorized as neutral.

Finally, the absence of gender differences between men's and women's cumulative military grade point average and company military ranking indicates that something other than objective performance accounts for gender differences in the subjective performance evaluations and attribute assignment (i.e., bias). However, the possibility exists that some of these evaluations may be grounded in accurate perceptions of leadership. The data do not enable the comparison of attribute assignment to actual performance. For instance, although a person may receive high marks on the aggregate performance measures we have, they may have done it in a way that leads the evaluator to judge the person's leadership style as selfish. However, previous research on applicable leadership traits (e.g., personality traits, intelligence, emotional intelligence, creativity) suggest that any significant differences may be attributed to other evaluative processes such as bias and stereotype content (Baer and Kaufman 2008; Halpern and LaMay 2000; Petrides and Furnham 2000; Schmitt et al. 2008).

Limitations and Future Directions

Because we analyzed real-world data from a current leadership performance evaluation system, there are several limitations to our research. One of the limitations of our dataset is that the evaluator's gender is unknown (so as to provide anonymity in the performance evaluation system). Because maintenance of the gender status hierarchy is conducted by both men and women, it would lead us to expect that there would not be any difference in how men and women assign leadership attributes based on stereotypes (Greenwald and Banaji 1995; Rudman et al. 2012). However, it would be valuable to understand whether men and women assign gendered descriptive and proscriptive leadership attributes differently in the present context. It would also be helpful to understand the

criteria they used because, in some cases, evaluators may have access to more objective performance data allowing for a more comprehensive depiction of how and when proscriptive and descriptive attributes are assigned. Finally, having an accurate depiction of the target's leadership style would enable analysis of who was penalized for agentic styles compared to communal styles of leadership.

Beyond gender, further analysis of attributes may provide more detailed knowledge of how particular attributes relate to each other and factors such as age, race and ethnicity, and important professional qualifications. We contend that an intersectional analysis of gender and race/ethnicity could be of particular interest and importance in today's modern workplace. Multicultural perspectives of leadership performance and evaluations are conspicuously sparse in the literature and would be useful for organizational leaders and human resources managers.

Another area to explore is recent research suggesting that effects of status incongruity and threats to the gender hierarchy in organizations and industries are observable at macro levels. In analyzing leader effectiveness and evaluations, two studies find that gender difference at the organization and industry level moderates leader evaluations (Ko et al. 2015; Paustian-Underdahl et al. 2014). Although our results are consistent with the findings of these macro level analyses that, in more masculine, male-dominated organizations, professions, and industries, men received evaluations as being more effective (e.g., men being evaluated as competent and women being evaluated as inept), examining gender composition at each level of leadership may provide further clarity on status effects in performance evaluations. Beyond historically male-dominated industries such as the military where there are more men than women at all levels, it may be useful to examine industries where there is overall gender balance but where women's representation decreases at successively higher leadership levels (e.g., advertising; pharmaceutical).

Status incongruity and a defense of the gender status quo on a macro level may also help explain why women are more likely to receive vague feedback on performance evaluations that are more closely tied to their communal traits as caregivers than as leaders (Correll and Simard 2016). Consistent with this line of research, women in our study were more likely to be evaluated positively as compassionate and negatively as inept. Future research that includes organization gender composition, evaluator gender, and objective individual performance outcomes may refine the relationship of status, performance, and stereotypes. Particularly useful would be an analysis of the type of language used in performance evaluations based on achievement of a desired outcome.

Finally, longitudinal research examining leadership style and evaluations could provide critical information on employees' outcomes associated with gender status beliefs. Specifically, performance evaluations that can be tied to

retention and promotion outcomes would provide valuable data to practitioners in establishing policy and evaluating best practices.

Practice Implications

The type and amount of evaluation criteria are instrumental to facilitating gendered performance expectations. Research shows that when there is more ambiguity in evaluation criteria and level of performance, evaluators are more likely to rely on stereotyped expectations (Heilman 2012). Additionally, when there is less relevant performance information available for evaluation, evaluators are more likely to infer performance based on stereotypes (Heilman et al. 2004; Swim et al. 1989). The subjective nature of the leadership performance evaluations available to our research participants, along with ambiguity and scant performance criteria, facilitates evaluations based on gender-stereotyped expectations. Alternatively, our results could be evidence of backlash that successful women receive in masculine organizational contexts and male-typed tasks (Heilman et al. 2004).

From a practical perspective, our results add to the wealth of research demonstrating how, in the absence of other information, ambiguous and subjective evaluations facilitate evaluators' use of gender stereotypes. Our data from the existing performance system assumes that participants use appropriate criteria to complete their evaluations. However, the minimal guidance provided may create an environment that allows gender status beliefs to be employed. Creating specific objective criteria based on goals, skills, and outcomes that could be assessed using available tools and metrics may provide more accurate and useful evaluations. Further, traits-based evaluation systems that employ phrases or other pre-selected evaluation content should purposefully select trait language after careful testing to minimize status beliefs and stereotype content.

Finally, our research complements prior work by providing additional evidence as to how status characteristics influence performance perceptions and has important implications for reducing gender inequities throughout the career pipeline. Identifying and removing stereotype content and biased language embedded in job advertisements and recruiting materials is vital to employers seeking to attract and hire diverse talent (Bolukbasi et al. 2016; Gaucher et al. 2011). Additionally, research in performance standards for accountability, promotion, attribution rationalization, and stereotype threat continues to be instrumental in understanding why women may be receiving subtle, if not explicit, messaging that they are not the right fit for the job (Biernat et al. 2010; Castilla 2008; Cuddy et al. 2004; Davies et al. 2005; Kalev et al. 2006; Lerner and Tetlock 1999; Rudman 1998; Rudman and Glick 1999).

Conclusions

Industries and professions are desperately trying to retain talented women who often receive formal and informal messaging that they do not belong and do not fit, as well as are penalized for their authentic leadership style. Even in esteemed institutions such as military service academies with a reputation for producing leaders of character to serve a nation, gender status beliefs are pervasive and may be unknowingly contributing to retention problems when women make career decisions years later. The findings of our research suggest that SCT and status incongruity may be reinforced in a U.S. military leadership context through an institutional, formal performance evaluation system.

In the present paper, we employed SCT and status incongruity to analyze real-world leadership performance evaluation data and found support that women leaders are evaluated with a greater variety of proscriptive attributes. Additionally, our finding that women are evaluated with a limited variety of descriptive leadership attributes provides theoretical nuance. Not only are women penalized for violating the gender status hierarchy by being evaluated with more proscriptive attributes, they are also penalized with fewer types of individual descriptive attributes (less variety). Although this finding is consistent with previous experimental research that women received similar overall numbers of descriptive evaluations (Eagly et al. 1992; Rudman and Glick 2001; Rudman et al. 2012), it also expands what we know about the variety of descriptive evaluations. This reasoning leads us to question if our findings are specific to the military leadership context or might also be observed in other professions and industries—especially those that are historically male or with a hypermasculine culture.

Whereas women received more proscriptive leadership attributes, the type of proscriptive attributes more often used in evaluations were feminine and not the masculine proscriptive attributes that status incongruity and agentic dominance (penalty) would predict. Military women in a hypermasculine culture are challenged to fit in as leaders while also contending with gender stereotypes (Archer 2013). In this cultural context, we expected the agentic dominance penalty to be amplified, with masculine proscriptive attributes outnumbering feminine attributes, but which was not the case. Our data do not enable us to explain this result, but we suggest that there could be two explanations. First, these are college students at a military service academy who may not have adopted a more traditional masculine leadership style. Alternatively, over the course of their time in the military leadership setting, some women may have received sufficient negative feedback and backlash about their agentic leadership style and adapted to a more traditionally feminine leadership style.

Our findings provide important evidence to organizational leaders and human resources managers seeking to develop

transparent evaluation processes that identify, develop, and promote the most talented people, regardless of gender. Research on status characteristics contributes to our knowledge of gendered language in performance evaluations and can assist researchers and practitioners with developing interventions. Understanding how gender status beliefs are associated with evaluation processes may facilitate changing workplace culture to be more gender-inclusive through less biased and stereotypical performance evaluations.

Acknowledgements The authors would like to thank Alice Eagly, Carolyn Judge, Emerald Archer and the Sex Roles reviewers for their thoughtful comments. We would also like to thank Cathy McGuire for her invaluable research assistance with data management. The views of the authors are their own and do not purport to reflect the position of the U.S. Naval War College, the U.S. Naval Academy, the Department of the Navy, or the U.S. Department of Defense.

Compliance with Ethical Standards

Conflicts of Interest The authors have no potential conflicts of interest to disclose for this research.

Ethical Approval The research is based on a secondary data source from the U.S. Naval Academy and was approved by the U.S. Naval Academy Institutional Review Board (IRB).

Informed Consent Informed consent was not applicable as the data was collected from a secondary data source.

References

- Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology*, 50, 195–255. <https://doi.org/10.1016/B978-0-12-800284-1.00004-7>.
- Archer, E. M. (2013). The power of gendered stereotypes in the US marine corps. *Armed Forces & Society*, 39(2), 359–391. <https://doi.org/10.1177/0095327X12446924>.
- Atkinson, R. E. (2015). *Military professionals as guardians of the republic: The hidden promise of Huntington's The Soldier and the State*. Retrieved from https://works.bepress.com/robert_atkinson/24/. Accessed 12 Mar 2018.
- Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior*, 42(2), 75–105. <https://doi.org/10.1002/j.2162-6057.2008.tb01289.x>.
- Bakan, D. (1966). *The duality of human existence*. Chicago: Rand McNally.
- Barrett, F. J. (1996). The organizational construction of hegemonic masculinity: The case of the US navy. *Gender, Work and Organization*, 3(3), 129–142. <https://doi.org/10.1111/j.1468-0432.1996.tb00054.x>.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162. <https://doi.org/10.1037/h0036215>.
- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch Jr., M. (1977). Status characteristics and interaction: An expectation states approach. *Annual Review of Sociology*, 6, 479–508. <https://doi.org/10.1146/annurev.so.06.080180.002403>.
- Biernat, M., Crandall, C. S., Young, L. V., Kobrynowicz, D., & Halpin, S. M. (1998). All that you can be: Stereotyping of self and others in a

- military context. *Journal of Personality and Social Psychology*, 75(2), 301–317. <https://doi.org/10.1037/0022-3514.75.2.301>.
- Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36(7), 855–868. <https://doi.org/10.1177/0146167210369483>.
- Boldry, J., Wood, W., & Kashy, D. A. (2001). Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues*, 57(4), 689–705. <https://doi.org/10.1111/0022-4537.00236>.
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). *Man is to computer programmer as woman is to homemaker? Debiasing word embeddings*. Paper presented at the Advances in Neural Information Processing Systems, Barcelona.
- Boyce, L. A., & Herd, A. M. (2003). The relationship between gender role stereotypes and requisite military leadership characteristics. *Sex Roles*, 49(7–8), 365–378. <https://doi.org/10.1023/A:1025164221364>.
- Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 113(6), 1479–1526. <https://doi.org/10.1086/588738>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum Associates.
- Correll, S., & Simard, C. (2016). Research: Vague feedback is holding women back. *Harvard Business Review*. Retrieved from <https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back>. Accessed 7 Jun 2016.
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues*, 60(4), 701–718. <https://doi.org/10.1111/j.0022-4537.2004.00381.x>.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88(2), 276–287. <https://doi.org/10.1037/0022-3514.88.2.276>.
- Eagly, A. H. (1987). Reporting sex differences. *American Psychologist*, 42, 756–757. <https://doi.org/10.1037/0003-066X.42.7.755>.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111(1), 3–22. <https://doi.org/10.1037/0033-2909.111.1.3>.
- Ebbert, J., & Hall, M.-B. (1993). *Crossed currents: Navy women from WWI to Tailhook*. Washington, DC: Brassey's.
- Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles*, 47(3/4), 99–114. <https://doi.org/10.1023/A:102102092>.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>.
- Francke, L. B. (1997). *Ground zero: The gender wars in the military*. New York: Simon & Schuster.
- Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, 24(4), 498–506. <https://doi.org/10.1177/0956797612457783>.
- Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, 101(1), 109–128. <https://doi.org/10.1037/a0022530>.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. <https://doi.org/10.1037/0033-295x.102.1.4>.
- Gündemir, S., Homan, A. C., de Dreu, C. K., & van Vugt, M. (2014). Think leader, think white? Capturing and weakening an implicit pro-white leadership bias. *PLoS One*, 9(1), e83915. <https://doi.org/10.1371/journal.pone.0083915>.
- Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12(2), 229–246. <https://doi.org/10.1023/A:100902751>.
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 113–135. <https://doi.org/10.1016/j.riob.2012.11.003>.
- 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. <https://doi.org/10.1037/0021-9010.89.3.416>.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, 71(4), 589–617. <https://doi.org/10.1177/000312240607100404>.
- Ko, I., Kotrba, L., & Roebuck, A. (2015). Leaders as males?: The role of industry gender composition. *Sex Roles*, 72(7–8), 294–307. <https://doi.org/10.1007/s11199-015-0462-4>.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>.
- Looney, J., Robinson Kurpius, S. E., & Lucart, L. (2004). Military leadership evaluations: Effects of evaluator sex, leader sex, and gender role attitudes. *Consulting Psychology Journal: Practice and Research*, 56(2), 104–118. <https://doi.org/10.1037/1061-4087.56.2.104>.
- Lundquist, J. H. (2008). Ethnic and gender satisfaction in the military: The effect of a meritocratic institution. *American Sociological Review*, 73(3), 477–496. <https://doi.org/10.1177/000312240807300306>.
- Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist*, 51(2), 157–158. <https://doi.org/10.1037/0003-066X.51.2.157>.
- Morgan, M. J. (2004). Women in a man's world: Gender differences in leadership at the military academy. *Journal of Applied Social Psychology*, 34(12), 2482–2502. <https://doi.org/10.1111/j.1559-1816.2004.tb01988.x>.
- Moskos, C. (1993). From citizens' army to social laboratory. *The Wilson Quarterly* (1976-), 17(1), 83–94. Retrieved from <http://www.jstor.org/stable/40258439>. Accessed 12 Mar 2018.
- Parker, K., Horowitz, J. M., & Rohal, M. (2015). *Women and leadership: Public says women are equally qualified, but barriers persist*. Washington, DC. http://www.pewsocialtrends.org/files/2015/01/2015-01-14_women-and-leadership.pdf. Accessed 22 Mar 2018.
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, 99(6), 1129–1145. <https://doi.org/10.1037/a0036751>.
- Pellerin, C. (2015). *Carter opens all military occupations, positions to women*. <https://www.defense.gov/News/Article/Article/632536/carter-opens-all-military-occupations-positions-to-women/>. Accessed 1 Mar 2016.
- Petrides, K., & Furnham, A. (2000). Gender differences in measured and self-estimated trait emotional intelligence. *Sex Roles*, 42(5–6), 449–461. <https://doi.org/10.1023/A:1007006652>.
- 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. <https://doi.org/10.1111/1471-6402.t01-1-00066>.
- Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, 57(4), 637–655. <https://doi.org/10.1111/0022-4537.00233>.

- Rosen, S. (1992). The military as an internal labor market: Some allocation, productivity, and incentive problems. *Social Science Quarterly*, 73(2), 227–237.
- Rosette, A. S., Koval, C. Z., Ma, A., & Livingston, R. (2016). Race matters for women leaders: Intersectional effects on agentic deficiencies and penalties. *The Leadership Quarterly*, 27(3), 429–445. <https://doi.org/10.1016/j.leaqua.2016.01.008>.
- 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. <https://doi.org/10.1037/0022-3514.74.3.629>.
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, 77(5), 1004–1010. <https://doi.org/10.1037/0022-3514.77.5.1004>.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4), 743–762. <https://doi.org/10.1111/0022-4537.00239>.
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, 48(1), 165–179. <https://doi.org/10.1016/j.jesp.2011.10.008>.
- Sampson, R. J., & Laub, J. H. (1996). Socioeconomic achievement in the life course of disadvantaged men: Military service as a turning point, circa 1940–1965. *American Sociological Review*, 61(3), 347–367. <https://doi.org/10.2307/2096353>.
- Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94(1), 168–182. <https://doi.org/10.1037/0022-3514.94.1.168>.
- Segal, M. W., Smith, D. G., Segal, D. R., & Canuso, A. A. (2016). The role of leadership and peer behaviors in the performance and well-being of women in combat: Historical perspectives, unit integration, and family issues. *Military Medicine*, 181(1S), 28–39. <https://doi.org/10.7205/MILMED-D-15-00342>.
- Soares, R., Bartkiewicz, M. J., Mulligan-Ferry, L., Fendler, E., & Kun, E. W. C. (2013). *Catalyst census: Fortune 500 women executive officers and top earners*. <http://www.catalyst.org/knowledge/2013-catalyst-census-fortune-500-women-executive-officers-and-top-earners>. Accessed 24 Mar 2018.
- Soyars, M. (2017). Firms' productivity rises as women become executives. *Monthly Labor Review*, 140, 1–2. http://www.heinonline.org/HOL/Page?handle=hein:journals/month140&div=6&g_sent=1&casa_token=&collection=journals. Accessed 9 Jun 2017.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105(3), 409–429. <https://doi.org/10.1037/0033-2909.105.3.409>.
- U.S. Department of Defense. (2016). *Population representation in the military services: Fiscal year 2015*. Washington, DC: Department of Defense Printing.
- United States Naval Academy. (2016). *Midshipmen aptitude for commissioning system*. Annapolis, MD. <https://www.usna.edu/Commandant/Directives/Instructions/1000-1999/COMDTMIDNINST-1600.2HMIDSHIPMEN-APTITUDE-FOR-COMMISSION-SYSTEM.pdf>. Accessed 28 Jun 2017.
- Wagner, D. G., & Berger, J. (1997). Gender and interpersonal task behaviors: Status expectation accounts. *Sociological Perspectives*, 40(1), 1–32. <https://doi.org/10.2307/1389491>.
- Yee, L., Krivkovich, A., Kutcher, E., Epstein, B., Thomas, R., Finch, A., ... Konar, E. (2016). Women in the workplace: 2016. <https://www.mckinsey.com/business-functions/organization/our-insights/women-in-the-workplace-2016>. Accessed 24 Mar 2018.