Breaking the Glass Ceiling: For one and all?

Francesca Manzi

Utrecht University

Madeline E. Heilman

New York University

Author note

Francesca Manzi, Department of Social, Health & Organisational Psychology, Utrecht

University

Madeline E. Heilman, Department of Psychology, New York University

Correspondence concerning this article should be addressed to Francesca Manzi, Martinus J.

Langeveldgebouw, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

Email: m.f.manzicembrano@uu.nl

Abstract

The current research challenges the assumption that the presence of women in leadership positions will automatically "break the glass ceiling" for other women. We contend that it is not just a female leader's presence, but also her performance, that influences evaluations of subsequent female candidates for leadership positions. We argue that the continued scarcity and perceived mismatch of women with high-level leadership increases gender salience, promoting perceptions of within-group similarity and fostering an evaluative generalization from the performance of a female leader to the evaluations of another, individual woman. In five studies, we demonstrate that the effect of exposure to a female leader on another woman's evaluations and leadership opportunities depends on whether she is successful or unsuccessful (Study 1) and whether she confirms or disconfirms stereotype-based expectations about women's leadership abilities (Study 2). Supporting the role of gender salience and shared group membership in the process, we show that this effect occurs only between women in male gender-typed leadership roles: evaluative generalization does not occur between women in contexts that are not strongly male in gender-type (Study 3) and is not observed between men in male-typed leadership (Study 4). We also explore whether there is evaluative generalization between male leaders in a female-typed context (Study 5). Our results suggest that overcoming gender imbalances in leadership may not be as simple as targeted placement, and that having women in high places should not induce complacency about the elimination of gender bias.

Breaking the Glass Ceiling: For one and all?

Although the proportion of women in top leadership remains low (UN Women, 2019), the visibility of the few women in these male-dominated positions is often assumed to presage the closing of the gender gap and has led many to conclude that women are finally "breaking the glass ceiling." Both in popular culture and in academia, these leaders often are portrayed as trailblazers, or female pioneers who have cleared the path for future aspiring female leaders (e.g., Cohen, Broschak, & Haveman, 1998; Huffman, Cohen, & Pearlman, 2010; Schultheis, 2015). Indeed, the mere presence of a woman in these positions is thought to be unequivocally beneficial for other women, making "the impossible possible" not only for the leader herself, but for other women as well (e.g., Traister, 2016).

But is this truly the case? Does the mere presence of a female leader open the door for other women? Past work has established the role of exposure to counter-stereotypical group members, such as female leaders, in weakening stereotype-based beliefs. This research has focused primarily on how stereotypes are affected by the generalization of information from a counter-stereotypical group member to beliefs about the group as a whole - in other words, a "person-to-group" generalization (e.g., Crawford, Sherman, & Hamilton, 2002; Dasgupta & Asgari, 2004; Fiske & Neuberg, 1990; Johnston & Hewstone, 1992; Lai et al., 2014; Rothbart, 1981). According to this perspective, the presence of a highly competent female leader should challenge, and therefore weaken, people's stereotype-based beliefs about the (lack of) abilities of women in leadership. However, whether such exposure effectively "breaks the glass ceiling", boosting opportunities for other individual women, remains largely unclear. Further, although person-to-group generalizations have been widely examined in past research, much less is known

about whether people generalize from one individual group member to another individual member of the group – in other words, a "person-to-person" generalization.

The goal of the present research is to examine the process of evaluative generalization and, more specifically, to determine whether people make evaluative generalizations between individual women in the context of leadership. We aim to test the widely held assumption that exposure to a woman in a traditionally male leadership position is beneficial for other women striving to become leaders. In addition, we hope to lend insight into the nature of person-to-person generalization and the conditions that regulate when it is likely to occur.

In a series of studies, we examine whether and how the presence of a woman in a top leadership role affects people's evaluations and recommendations regarding another woman who is a candidate for a leadership position. We contend that when exposure to a woman in a top leadership role heightens the degree to which gender is made salient to perceivers, it will promote an evaluative generalization between the female leader and a female candidate for leadership. If our ideas are correct, then exposure to a woman in a top leadership role will not always lead to evaluative generalization, nor will it always lead to positive outcomes for aspiring female leaders. In fact, there may be situations in which the presence of a female leader impedes rather than facilitates other women's attainment of their leadership goals.

Gender bias against female leaders

The invisible barriers that prevent women from reaching the upper echelons of leadership have been referred to as the "glass ceiling" – a metaphor that continues to resonate among women more than 30 years after the term was coined (Hymowitz & Schellhardt, 1986). Research consistently shows that female leaders, as well as those aspiring to become leaders, often face challenges that men do not, particularly in domains historically dominated by men (Eagly, 2007;

Ellemers, Rink, Derks, & Ryan, 2012; Heilman, 2001). One of the most well-documented

challenges to women's upward mobility in these areas is the persistence and pervasiveness of

stereotypes that portray women as not having "what it takes" to be a good leader (Eagly &

Karau, 2002; Koenig, Eagly, Mitchell, & Ristikari, 2011, Schein, 1973, 2001).

People's beliefs about the characteristics of men and women tend to be organized along

two general dimensions: agency and communion. Agency comprises attributes such as

achievement orientation (e.g., able, successful), assertiveness (e.g., dominant, forceful), and

autonomy (e.g., independent, self-reliant); while communality denotes consideration for others

(e.g., caring, helpful), affiliation with others (e.g., sociable, likable), and emotional sensitivity

(e.g., tender, sensitive) (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972;

Diekman & Eagly, 2000; Heilman, Manzi, & Braun, 2015; Hentschel, Heilman, & Peus, 2019).

These dimensions constitute the core content of gender stereotypes, which depict men as agentic

and women as communal (Bakan, 1966; Eagly, Wood, & Diekman, 2000). Despite the many

advances that women have made over the previous decades, women continue to be seen as more

communal and less agentic than men (Eagly, Nater, Miller, Kaufmann, & Sczesny, 2019).

High-level leadership has been historically dominated by men and is typically

characterized in masculine terms (Gaucher, Friesen, & Kay, 2011; Koenig et al., 2011).

Congruity models of gender discrimination maintain that the mismatch between the attributes

typically ascribed to women and the agentic qualities thought necessary for success as a leader

fosters the belief that women are not equipped to effectively handle leadership roles and

consequent expectations that they will be incompetent in enacting them (Eagly & Karau, 2002;

Heilman, 1983, 2012). They further contend that these negative performance expectations and

presumptions of incompetence induce bias against aspiring female leaders (Heilman, 2001,

2012). There is much evidence supporting these ideas. Compared to men, women are less likely

to be selected or promoted for leadership positions, especially those for which agency is

considered most key (Koenig et al., 2011; Lynness & Heilman, 2006). Even when women have

demonstrated their ability and have achieved high-level positions, stereotype-based performance

expectations continue to haunt them, adversely affecting evaluations of their leadership and

performance (Ellemers et al., 2012).

Cracks in the glass-ceiling

Despite the power of gender stereotypes, there is reason to believe that the presence of

the few women who have reached higher levels of leadership might be beneficial for aspiring

female leaders, as is often assumed. Classic research on the mere exposure effect has shown that

exposure to a novel stimulus elicits a positive response toward similar stimuli (Monahan,

Murphy, & Zajonc, 2000; Zajonc, 1968; 2001; Zebrowitz, White, & Wieneke; 2008), suggesting

that simply being exposed to a female leader might lead to more positive attitudes towards other

women in leadership. However, seeing a woman in a position of power also conveys important

information that goes beyond mere exposure. Specifically, it provides information about

women's leadership abilities that otherwise would have been unavailable, demonstrating that

women, like men, can lead successfully (DeVaro & Waldman, 2012; Milgrom & Oster, 1987).

Importantly, this information deviates from stereotypical depictions of women.

There is a large body of evidence demonstrating that exposure to individuals who

strongly defy their group's stereotypes can lead to a revision of people's stereotypes about that

group (Johnston & Hewstone, 1992; Rothbart, 1981; Weber & Crocker, 1983). For example,

positive interactions with an individual from a negatively stereotyped group have been shown to

decrease prejudice towards the group as a whole (Pettigrew & Tropp, 2006). Similarly, observing

someone succeed in a counter-stereotypical domain can dampen perceivers' stereotypes and attenuate negative attitudes towards the group (Bless, Schwarz, Bodenhausen, & Thiel, 2001; Bodenhausen, Schwarz, Bless, & Wänke, 1995; Critcher & Risen, 2014; Dasgupta & Greenwald, 2001; Hewstone, Hassebrauck, Wirth, & Waenke, 2000).

It is reasonable, then, to assume that the mere presence of a woman in a position of power will attenuate gender stereotypes about women's lesser leadership competence. Research on the effects of exposure to female leaders lends support to this idea. For example, there is evidence that observing women in high-profile, male-typed leadership positions not only weakens some stereotypes about women (e.g., women are emotional), but also activates some counter-stereotypical beliefs (e.g., women are assertive) (Dasgupta & Asgari, 2004).

In line with this idea, the metaphor of "breaking the glass ceiling" suggests that once one woman has reached the top, positive outcomes for another aspiring female leader will ensue. This metaphor has been very powerful. Public policies designed to give preference to women over equally qualified men (e.g., affirmative action) or to increase the number of women in male-dominated domains (e.g., gender quotas) are often predicated on the assumption that this effect will "snowball" to the point that other policies directed at increasing gender equality will be unnecessary. However, despite the pervasiveness of beliefs about the salutary effects of breaking the glass ceiling, there is little evidence to support it. Research has yet to determine whether the presence of a woman in leadership beneficially impacts the evaluations of other individual women – evaluations that are essential to determining whether their aspirations to become leaders are fulfilled.

Female leaders, gender salience, and evaluative generalization

Although the evidence might point to the benefits of exposure to female leaders for other women in the abstract, it is uncertain whether the presence of a single stereotype-defying woman is enough to curb stereotype-based expectations and, as a consequence, reduce biased evaluations of another, unrelated, individual woman. Indeed, while the processes involved in person-to-group generalizations have been well documented, much less is known about person-to-person generalizations and the processes underlying them.

Generalization from one group member to another appears to be susceptible to situational variation, and prior investigations are suggestive about the conditions under which it is likely to occur. Perceived similarity has been shown to promote such generalizations. Person-to-person generalizations have been documented between individuals that share physical attributes, suggesting that physical similarity may be enough for individuals to transfer their evaluations from one individual to another (Gawronski & Quinn, 2013; Lewicki, 1985, Ranganath & Nosek, 2008). But the degree to which individuals are perceived to be similar to one another is affected by factors beyond physical attributes. For example, members of outgroups are often seen as more homogenous, and therefore more similar to one another, than members of one's own group (Quattrone & Jones, 1980). These perceptions of within-group similarity lead to a greater likelihood of making direct generalizations between members of outgroups than ingroups (Chen & Ratliff, 2015; Henderson-King & Nisbett, 1996; Ratliff & Nosek, 2011).

An important factor contributing to perceptions of within-group similarity is group membership salience. Past research suggests that heightening the cognitive availability of a person's group membership (e.g., a target's race or gender) increases the degree to which she or he is perceived by others as similar to other members of the group and different from members

of other groups (Kanter, 1977; Tajfel, 1969; Taylor, Fiske, Etcoff, & Ruderman, 1978). To the

degree that group membership is made salient, general group impressions (e.g., stereotypes)

become associated with all members of the group, who are seen as interchangeable with one

another. In such cases, information about any specific group member is less likely to be

remembered in a way that is uniquely associated with that member (Crawford et al., 2002).

Because it promotes the perception of similarity among group members, group member salience

is likely to facilitate person-to-person generalization.

Gender group membership is not always salient; there are many aspects of an individual

that compete for prominence in person perception. However, there is reason to believe that when

women attain leadership positions in traditionally male domains, their gender will be highly

salient. Perceptual biases toward novelty heighten salience for group memberships that are

infrequent or unfamiliar (Kanter, 1977; Oakes & Turner, 1986; Risen, Gilovich, & Dunning,

2007; Taylor & Fiske, 1978), and members of rare or unfamiliar groups are often seen as less

differentiated and more homogenous than those of familiar groups (Linville & Fischer, 1993;

Ostrom & Sedikides, 1992; Quattrone & Jones, 1980). Women still constitute a minority in

male-typed leadership, rendering their presence in such roles infrequent and novel. As a result,

their gender is likely to draw attention. Moreover, because of the perceived lack of fit between

female stereotypes and leadership requirements, it is not just the scarcity of women leaders, but

also the contradiction with expectations their presence connotes, that is likely to heighten the

salience of their gender.

To the degree that the gender of a woman in a leadership position is salient, perceptions

of within-group homogeneity between her and other individual women should be high, and

perceptions of variability between her and other women should be low. Consequently, perceivers

should be more likely to see a female leader as similar to other individual women and should be less likely to differentiate between them. Under these conditions, any one woman is likely to be viewed as a member of her gender group rather than as an individual, appearing to others as representatives of women at large.

These ideas are significant for women who are seeking access to leadership positions: it implies that when the gender of a female leader is salient, there will be a tendency to see the aspiring woman as similar to her – to see the two of them as interchangeable members of their gender group. Thus, exposure to a counter-stereotypical exemplar such as a female leader may not only affect perceptions of women as a group (e.g., gender stereotypes), as past research has shown. It may also trigger evaluative generalization between individual group members, with perceivers generalizing from their evaluations of a female leader to the evaluations of another woman.

The critical role of female performance

If the salience of a female leader's gender increases the degree to which she is perceived as representative of or similar to other women, her individual actions and behaviors should have a considerable impact on how aspiring female leaders are perceived. However, contrary to common assumptions, the nature of that impact may not always be beneficial.

Previous research demonstrating the beneficial effects of exposure to female leaders are based on the supposition that female leaders will be successful leaders. In fact, in most empirical studies aimed at examining the effects of exposure to counter-stereotypical exemplars it is difficult to disentangle performance from exposure. Participants are typically exposed to an individual who is in a counter-stereotypical role and, at the same time, is successful in that role, thereby conflating the counter-stereotypicality of the situation (e.g., a woman in a leadership

position) with the counter-stereotypicality of the outcome (e.g., a woman being successful in a male-typed role). Thus, it is unclear whether to be beneficial, exposure to a female leader may require not only her presence, but also her success in the role.

When women take on leadership positions, success is not assured; there is always the possibility of failure. As with all leaders, individual women will vary in their performance and some will not be successful. Moreover, research suggests that failure –both real and perceived— is even more likely for female than male leaders.

Research on the "glass cliff" phenomenon has shown that women are more likely than men to be appointed as leaders in times of an economic crisis or when the position is associated with a higher risk of failure (Haslam & Ryan, 2008; Ryan & Haslam, 2005), especially in male-typed contexts (Bruckmüller & Branscombe, 2010). Although women and men are likely to perform similarly in these circumstances, it is the women who are more often filling these precarious leadership roles and therefore appear more prone to failure.

Not only are female leaders more likely than male leaders to actually fail because of the glass cliff phenomenon but, unless unequivocally successful, their performance is more likely to be perceived as unsuccessful. Because success in male-typed domains is inconsistent with stereotype-based expectations, it is easily discounted, not attended to, or simply not given much weight in evaluation (Plaks, Stroessner, Dweck, & Sherman, 2001; Perry, Davis-Blake, & Kulik, 1994; Uhlmann & Cohen, 2005). Indeed, research has demonstrated that women in male-typed domains are judged more harshly than men when their performance is sub-par (Rossette & Livingston, 2012), when they make mistakes (Bongiorno, Bain, & David, 2013; Brescoll, Dawson, & Uhlmann, 2010), and when their performance fluctuates over time (Heilman, Manzi, & Caleo, 2019). This negativity bias is evident even when the source of poor performance is

ambiguous, with female leaders being held responsible for bad outcomes to a higher degree than their male counterparts (The Rockefeller Foundation, 2016).

Given that female leaders are more likely than male leaders to actually be in positions in which failure is preordained, or to prompt perceptions that they are unsuccessful, it is important to consider the effect of exposure to female leaders who are not successful on people's reactions to other women striving for leadership. If our contentions are correct, and exposure to female leaders affects the evaluations of other women, then not only should a female leader's success be beneficial, but her failure should be detrimental to the evaluations of other women aspiring to become leaders.

Overview of the current research

The research presented here tests the assumption that the presence of a female leader will inevitably break the glass ceiling for other women by exploring the process by which the presence and performance of a woman in a position of leadership affects the evaluations of an aspiring female leader. We propose that in traditionally male contexts – those that make gender salient – exposure to a female leader will affect reactions to another woman seeking access to leadership. We expect that the combination of gender salience and shared group membership creates the perfect context for generalization from a perceiver's evaluations of a female leader to his or her evaluations of another woman seeking access to leadership.

We further propose that it is not merely the presence of the female leader, but also how she performs that will determine the nature of the response to a female candidate for a leadership role. Specifically, we expect that exposure to a woman in a male-typed leadership role will benefit a female candidate for a leadership position when the leader has been successful, but negatively affect a female candidate when the leader has been unsuccessful.

We test these predictions in five studies. We first examine whether, in a male-typed context, the performance of a female (but not male) leader is generalized to a female candidate for a leadership position (Study 1). We then explore the role of stereotype-based beliefs in the content of these evaluative generalizations. Specifically, we predict that the performance of a female leader will affect the degree to which a female candidate for leadership is seen as having the qualities necessary to succeed in leadership, and that these perceptions, in turn, lead to more positive or negative evaluative outcomes for that female candidate (Study 2). To further test our contention that evaluative generalization occurs because of a combination of gender salience and shared group membership, we then seek to establish that generalization occurs between women when the female leader is in a male-typed leadership position, but not in a position that is not considered to be strongly male in gender-type (Study 3), and to demonstrate that these generalizations are observed for female, but not for male candidates (Study 4). Although the aim of this research is to examine the effects of exposure to female leaders on other women aspiring to leadership roles, we conclude by examining whether men, too, are the targets of evaluative generalization when the leadership context is counter-stereotypical for them (Study 5)[1].

Pilot Study

We have proposed that evaluative generalization between women in leadership occurs, in part, because of the salience of a female leader's gender. Thus, before testing our hypotheses, we sought to examine whether a woman's gender is indeed salient to perceivers when she is in a traditionally male leadership position. If this is the case, then gender, as a category, should be activated more often when people are exposed to a female than a male leader.

---

[1] Materials are available in appendices (see supplemental materials). Data and syntax are available at https://osf.io/y37th/.

Method

Participants and design.

Two hundred (124 female, 74 male, 2 undetermined)[2] participants with a mean age of

19.36 years were recruited for course credit from the subject pool of a large Northeastern

university. Of these participants, 45.5% identified as Asian, 30% as White, 12.5% as Hispanic,

7.5% as Black and 3.5% as other racial categories. Participants were randomly assigned to read

about a female leader or male leader.

Procedure.

Participants were told the study was about how different amounts and types of

information affect the way people characterize others (see appendix A for details). They were

asked to review information about the CEO (either male or female) of a male-typed organization

(either a steel mill or a tool manufacturing company). After reading a brief account of the CEO

and the company (see appendix B for stimuli), participants were asked to describe the person

they read about.

Dependent Measures.

Gender salience. To assess the degree to which the gender of the leader was salient to

participants, we created a binary measure based on their answers to two open-ended questions

asking participants to describe the CEO. In the first question, participants were to imagine that

they had to describe the person they reviewed to someone who had no information about this

person. They were asked to write down 5 words that would provide the most accurate description

of the person they read about. The second question required participants to write a brief

---

[2] Data for the pilot study was collected throughout the course of one academic semester. Because the resulting
sample was largely female, we continued collecting data from male participants to ensure a more balanced gender
distribution across the two conditions, aiming for a total of 200 participants.

description of the person reviewed (see appendix C for measures). To create a measure of gender

salience, we coded participants' responses to both questions for the inclusion of words explicitly

alluding to the gender of the target. Examples of these words were "man", "woman", "male",

"female", "guy", and "girl". Responses that included 1 or more of these words were coded as 1.

If no gender words were used, the answers were coded as 0.

Manipulation and stimulus checks. We asked participants to select the name of the CEO

from a list of three options to check the manipulation for leader gender. To ensure that

participants viewed the company as male-typed, we also asked them to rate how masculine or

feminine they thought the company was on a scale of 1 (masculine) to 7 (feminine).

Results

Preliminary Analyses.

All participants answered correctly when asked to select the name of the person they had

read about, indicating that they were aware of the CEO's gender. Analyses also confirmed that

participants viewed the company as male in gender-type ($M = 3.22$, $SD = 1.49$), rating it

significantly below the midpoint of the scale (4), $t(199) = 7.45$, $p < .001$. Analyses testing for

differences between male and female participants indicated no significant main effects or

interactions of participant gender on our dependent measure when included as an additional

predictor in the model. In addition, there were no differences in gender salience depending on the

particular male-typed company participants read about (a steel mill or a tool manufacturing

company). All data were therefore combined for the analyses reported.

Gender Salience.

We ran a logistic regression to test whether the likelihood of mentioning gender when

describing a leader in a male-typed company would differ depending on the leader's gender.

Leader gender was dummy-coded (female as 0 and male as 1). The analyses confirmed that participants were 47% more likely to spontaneously refer to the target's gender when they were describing a female leader than when they were describing a male leader, $B = -.76$, $SE = .34$, $Wald\ \chi^2(1) = 5.11$, $p = .02$. Specifically, 32% of participants spontaneously alluded to the leader's gender when she was a woman, compared to 18% when the leader was a man.

Study 1

In Study 1, we sought to test the idea that there is an evaluative generalization from the performance of a female leader to the evaluations of a female candidate. Results from the pilot study confirmed that gender is activated to a greater extent when people are exposed to female leaders than when they are exposed to male leaders in a traditionally male setting. Given that gender is salient for women in male-typed leadership positions, and that both female leaders and aspiring female leaders belong to the same gender group, we expected evaluative generalization when participants were exposed to a female leader but not to a male leader. Specifically, we predicted that a leader's successful performance will lead to more positive evaluations of a female candidate than a leader's unsuccessful performance when the leader is a woman but not when the leader is a man. (Hypothesis 1.1). We also examined whether the presence of a female leader is more beneficial to other women than the presence of a male leader. We predicted that exposure to a successful female leader will be more beneficial to the evaluations of a female candidate for a leadership position than exposure to a successful male leader (Hypothesis 1.2), but that exposure to an unsuccessful female leader will be more detrimental for the evaluations of a female candidate for a leadership position than exposure to an unsuccessful male leader (Hypothesis 1.3).

Method

    Participants and design.

One hundred and forty-two (86 female, 56 male) participants[3] with a mean age of 19.53

years were recruited through the participant pool of a large Northeastern university. Of these

participants, 55.6% identified as Asian, 23.9% as White, 9.9% as Hispanic, 4.9% as Black, and

4.2% as other racial categories. Two additional participants completed the study but were

excluded from analyses after incorrectly responding to a manipulation check. The study was a 2

x 2 between-subjects design, with leader gender (male or female) and leader performance

(success or failure) as the two independent variables. Participants were randomly assigned to one

of the four experimental conditions[4].

    Procedure.

The study was said to be about employment decisions. Participants were told they would

be making initial screening decisions about candidates being considered for high-level jobs (see

appendix A for details). They were presented with descriptive information about a male-typed

organization (a steel manufacturing company) and were told that the current CEO of the

company, either male or female (Michael or Patricia Walden), was departing. To strengthen our

manipulation and ensure that gender was attended to, we also included a portrait photo of the

CEO[5]. To manipulate the leader's success or failure, we provided a composite of newspaper

---

[3] A total of 144 participants were recruited for Study 1 over the course of one semester. Sample size for studies 1
through 3 was determined by the number of students who had participated in each study by the end of one academic
term. This time frame was chosen based on past data collection experience using the participant pool for course
credit. We estimated that during this time we would obtain data from at least 140 participants, yielding about 35
participants per condition in Study 1.
[4] All consent forms, debriefing forms, stimuli and procedures were approved by New York University's Institutional
Review Board (IRB# 13-9505/IRB# FY2016-479) before carrying out this program of research.
[5] The photos were pretested on a sample of college students to ensure they were matched in perceived age,
attractiveness, and competence.

clippings that addressed the CEO's on-the-job performance, ostensibly to provide participants

with information about "the current situation of the company". The headlines were designed to

either signal the leader's success (e.g. "CEO Patricia [Michael] Walden exceeding expectations")

or failure (e.g. "CEO Patricia [Michael] Walden falling short"). After reviewing the newspaper

clippings, participants were asked to evaluate a female candidate for the position. Information

about the candidate included schools attended, degrees earned and history of work experience. A

photo was also attached to the profile to ensure that participants distinguished between the

candidate and the leader (see appendix B for stimuli). Participants then completed a brief

questionnaire containing the dependent measures and were thanked for their participation and

debriefed.

Dependent Measures.

Perceived job-fit of the female candidate. Responses to three questions were aggregated

to create a measure of perceived job-fit of the female candidate for the leadership position.

Participants were asked to rate how competent, effective, and qualified they thought the

candidate was to be the next CEO of the company. All responses were on a scale from 1 to 7.

The composite showed high reliability (α=.90).

Screening recommendation for the female candidate. Participants also were asked how

strongly they would recommend that the candidate be kept in the applicant pool on a scale from

1 (not at all strongly) to 7 (very strongly).

The specific questions used for each measure are reported in appendix C.

Manipulation and stimulus checks. As a check of our leader gender manipulation we

asked participants to select the name of the departing CEO from a list of three options. They

were also asked to rate how this leader had performed on a seven-point scale (1=very poorly,

7=very well) as a manipulation check for success and failure of the leader.

Results

    Preliminary Analyses.

    All participants correctly indicated the name of the departing CEO and were therefore

aware of the leader's gender. The leader performance manipulation also had its intended effect.

With the exception of two participants (excluded from analyses[6]), responses were consistent with

condition, rating the failing leader's performance below the midpoint of the scale (4), and the

successful leader's performance above the midpoint of the scale. In addition, analyses of the

performance ratings indicated that the participants rated the leader's performance more favorably

in the success conditions ($M = 6.62$, $SD = 0.80$) than in the failure conditions ($M = 1.77$, $SD =$

0.70), $t(140) = 38.40$, $p < .001$, $d = 6.49$, CI = [-5.10 – -4.60].

    There were no significant main effects or interactions with participant gender on any of

the measures, so data for male and female participants were combined for all subsequent

analyses.

    Dependent Measures.

    Perceived job-fit of the female candidate. A two-way analysis of variance (ANOVA) of

the perceived job-fit ratings for the female candidate yielded a significant main effect of leader

performance, $F(1, 138) = 7.11$, $p = .01$, $\eta_p^2 = .05$, and a significant interaction between leader

performance and leader gender, $F(1, 138) = 5.07$, $p = .03$, $\eta_p^2 = .04$. Pairwise comparisons

provided support for H1.1. Female candidates were perceived to be a significantly better fit for

the leadership position when participants had been exposed to a successful rather than an

---

[6] Two participants rated the failing leader's performance above the midpoint of the scale (4)

unsuccessful female leader, $t(138) = 3.45$, $p = .001$, $d = .59$, CI = [-1.31 – -0.36], but the

performance of a male leader had no effect on the job-fit ratings of the female candidate, $t(138) =$

0.30, $p = .77$, $d = .05$, CI = [-0.54 – 0.40]. Furthermore, consistent with H1.3, the female

candidate was rated as significantly less fit for the leadership role when a failing leader was

female than male, $t(138) = 2.21$, $p = .03$, $d = .38$, CI = [-1.00 – -0.06]. However, H1.2 was not

supported: being exposed to a successful female (vs. male) leader did not lead to significantly

higher ratings of the female candidate's job-fit, $t(138) = 0.97$, $p = .33$, $d = .17$, CI = [-0.24 –

0.71]. Means and standard deviations for each condition are reported in Table 1.

Screening recommendation for the female candidate. A two-way ANOVA yielded a

main effect of leader performance, $F(1, 138) = 6.32$, $p = .01$, $\eta_p^2 = .04$, and a significant

interaction between leader performance and leader gender, $F(1, 138) = 4.43$, $p = .04$, $\eta_p^2 = .03$.

The results of the pairwise comparisons paralleled those for perceived job-fit. They provided

support for H1.1, indicating that female candidates received significantly more positive

screening recommendations after participants had read about a successful than a failing female

leader, $t(138) = 3.25$, $p = .001$, $d = .55$, CI = [-1.40 – -0.34], but their screening

recommendations were unaffected by the performance of a male leader, $t(138) = 0.29$, $p = .77$, d

= .05, CI = [-0.60 – 0.45]. Moreover, In line with H1.3, when the leader had performed poorly,

screening recommendations were significantly more negative for female candidates when the

leader was a woman than a man, $t(138) = 2.09$, $p = .04$, $d = .36$, CI = [-1.09 – -0.03], and

contrary to H1.2, screening recommendations were not significantly more positive for female

candidates following exposure to a successful female leader relative to those following exposure

to a successful male leader, $t(138) = 0.89$, $p = .37$, $d = .15$, CI = [-0.29 – 0.77]. See Table 1 for

means and standard deviations of each condition.

Discussion

Results from Study 1 provide support for the idea that there is an evaluative generalization from female leaders to female candidates for leadership. Our findings demonstrate that exposure to a woman, but not a man, in a male-typed leadership position directly affects the evaluations of other women aspiring to that position. In particular, they show that the performance of a female leader is critical in the evaluations of another woman. Specifically, female candidates were thought to be less suitable and were less highly recommended for a leadership position following exposure to an unsuccessful than a successful female leader. Our findings also indicated that when people had been exposed to poorly performing leaders, it was more harmful than beneficial for female candidates to follow a woman than a man but, contrary to our hypothesis, exposure to a successful female leader did not boost the evaluations of a female leadership candidate. These results provide preliminary evidence that, as we had predicted, exposure to women in male-typed leadership roles does not necessarily benefit other women. They also suggest an asymmetry – that the negative effect of leadership failure is greater than the positive effect of leadership success.

Study 2

In Study 1 we provided preliminary evidence that the presence of a female leader is not unequivocally positive for aspiring female leaders. Our findings reveal that there is an evaluative generalization between women in male-typed leadership, whereby the performance of a female leader has significant effects on the perceptions of another woman's job-fit and screening recommendations. But what exactly is being generalized from a female leader to a female candidate? In Study 2 we sought to examine the role of stereotype-based beliefs in evaluative

generalization and the extent to which stereotype-relevant information is transferred between women in leadership.

Gender stereotypes depict women as high in communality but low in agency. We expected an unsuccessful female leader to confirm negative beliefs about women's lack of agency and a successful leader to challenge these beliefs. We furthermore anticipated that these stereotype-based reactions to a female leader's performance would carry over to the perceptions of female candidates' agency. Specifically, we predicted that female candidates would be perceived as less agentic following exposure to an unsuccessful woman leader than following exposure to a successful woman leader. In line with congruity models of gender discrimination (Eagly & Karau, 2002; Heilman, 1983; 2012), which posit that agency perceptions are key to biased leadership evaluations, we expected that the characterization of the female candidate as less agentic would, in turn, negatively affect evaluations of her fit for a leadership position and her screening recommendations. Figure 1 depicts the hypothesized model.

Method

Participants and design.

Two hundred and thirty-seven (149 female, 86 male, 2 undetermined)[7] participants with a mean age of 19.47 years were recruited from a large Northeastern university. Of these participants, 41.4% were Asian, 33.3% were White, 13.1% were Hispanic, 4.2% were Black, and 5.5% identified as another racial group. Three additional participants completed the study but were excluded from analyses after failing manipulation checks. Participants were randomly assigned to either read about a successful female leader or an unsuccessful female leader.

---

[7] Following the recommendation of Kline (2011), we sought to recruit at least 200 participants to test the proposed structural equation model. After the desired sample size was obtained, we continued collecting data from male participants to ensure a more gender-balanced sample.

Procedure.

The procedure closely followed the one used in Study 1, but in this case, participants only reviewed information about a female leader who had succeeded or failed as the CEO of a male-typed company (a steel mill). They were then asked to evaluate a female candidate for the position (see appendix B).

Dependent Measures.

We included the same two dependent measures as Study 1: a scale of perceived job-fit comprised of the same three items (ratings of how competent, effective, and qualified they thought the candidate was to be the next CEO of the company $\alpha = .84$) and a screening recommendation about whether to retain the applicant for further consideration. The manipulation and stimulus checks also were the same.

In addition to evaluating the fit of the candidate for the leadership position and providing a screening recommendation, participants were asked to rate their general impressions of the candidate on a series of traits. All items were chosen to denote traits that have been consistently associated with agency – a stereotypical description often used to describe men but not women (Eagly & Karau, 2002, Heilman, 2012) (see appendix C). A scale of the perceived agency of the female candidate ($\alpha = .86$) was created by combining ratings on four 7-point bipolar adjective scales (timid – bold, emotional – rational, hesitant – not-hesitant, uncertain – certain) and five scales asking participants to directly rate the leader's decisiveness, forcefulness, achievement-orientation, leadership ability, and strength on a scale of 1 (not at all) to 7 (very much). Higher ratings indicated higher perceived agency.

To ensure that the perceived agency scale was statistically distinct from the perceived job-fit scale, we performed confirmatory factor analyses (CFA). As expected, a two-factor model

(with the items included in each scale loading onto separate factors) was a significantly better fit to the data than a one-factor model (with all items loading onto one factor), $\Delta\chi^2$ (1, N = 237) = 163.94, p = .001[8]. These analyses confirmed that the perceived agency and perceived job-fit scales measure different constructs.

Results

Preliminary Analyses.

Two participants were excluded after incorrectly selecting the leader's name, and one additional participant was excluded after inaccurately rating the leader's performance[9]. In line with our manipulation, ratings of the leader's performance on a scale of 1 ("very poorly") to 7 ("very well") indicated that participants perceived successful performance (M = 6.84, SD = 0.43) to be significantly more positive than unsuccessful performance, (M = 1.76, SD = 0.79), t(235) = 61.37, p < .001, d = 8.01, CI = [-5.24 – -4.91]. No significant main effects or interactions were found when including participant gender into the analyses of our dependent measures. The analyses that follow are therefore collapsed by participant gender.

Dependent Measures.

Structural equation modeling was used to test the above-described model of the relationship between exposure to a female leader's success or failure and a female candidate's perceived agency ratings, job-fit ratings, and screening recommendations. In Figure 2, unstandardized coefficients are given for each path. This model provided a very good fit to the

---

[8] In addition to distinguishing between perceived job-fit and perceived agency, CFAs also suggested that the perceived agency scale may be multifaceted, something that has been discussed in previous literature on the agency construct (see Hentschel, Heilman, & Peus, 2019). Specifically, the data showed a pattern by which the job-fit items loaded onto a first factor, timid – bold, emotional – rational, hesitant – not-hesitant, and uncertain – certain loaded onto a second factor, and decisiveness, forcefulness, achievement-orientation, leadership ability, and strength loaded onto a third factor. Analyses replacing the 9-item perceived agency scale with each of these two factors in the hypothesized model revealed the same pattern of results as those presented here.
[9] One participant rated the failing leader's performance above the midpoint of the scale (4)

data: χ^2 (2, N = 237) = 1.80, p = .41; CFI = 1.0; TLI = 1.0; $RMSEA$ = 0.00. A significant indirect

effect indicated that, in line with our hypotheses, perceived agency of the female candidate

mediated the generalization from a female leader's performance to the perceptions of a female

candidate's fit for the job, b = .197, SE = .064, p = .002. As predicted, unsuccessful (vs.

successful) performance of a female leader led to the perception that a female candidate for the

leadership role was lower in agency (b = .375, SE = .114, p = .001) and, in turn, was thought to

be a worse fit for the leadership position (b = .525, SE = .059, p < .001). In addition, lower

perceptions of job-fit led to more negative screening recommendations for the female candidate

(b = .809, SE = .058, p < .001).[10]

Discussion

These results provide further support for the idea that there is evaluative generalization

between women in male-typed leadership. In addition, they demonstrate that stereotype-based

beliefs about women's (lack of) agency play a part in this process. Specifically, exposure to an

unsuccessful female leader led people to perceive a female leadership candidate as less agentic

(e.g., less dominant, independent, self-reliant) than exposure to a successful female leader, which

in turn led to the perception that she was less suited for the role and to a more negative screening

decision. These results suggest that exposure to a stereotype confirming or disconfirming female

leader affects perceptions, evaluations, and decisions about a female leadership candidate.

---

[10] Additional analyses comparing mean evaluations of the female candidate after exposure to a successful or unsuccessful female leader followed the expected pattern of results. Mean agency ratings for the female candidate were significantly lower following exposure to an unsuccessful (M = 5.09) than successful (M = 5.47) female leader, t(235) = 3.30, p = .001. In line with Study 1, job-fit perceptions of the female candidate were significantly worse following exposure to an unsuccessful (M = 5.28) than successful (M = 5.66) female leader, t(235) = 3.13, p = .002. Although the screening recommendation was lower after participants were exposed to an unsuccessful (M = 5.31) than successful (M = 5.55) female leader, this test did not reach significance, t(234) = 1.63, p = .104.

Study 3

Studies 1 and 2 demonstrated that evaluative generalization between female leaders occurs, and that stereotype-based beliefs about the female leadership candidate play a role in the generalization process. In Study 3 we focused on gender salience. Although our pilot study demonstrated that gender is more salient for female than male leaders in male-typed contexts, we have not yet investigated the role that this enhanced gender salience plays in the evaluative generalization process. We have posited that shared group membership, by itself, does not prompt evaluative generalization – that gender salience is a necessary condition for evaluative generalization to occur. This study sought to test the role of gender salience by varying the gender-type of the leadership position.

Given that novelty and lack of fit perceptions are key to salience, gender should be most salient when women are leaders in contexts where their presence is both scarce and incongruent with gender stereotypic expectations—that is, when the leadership position is viewed as strongly male in gender-type. Therefore, we hypothesize that the performance of a female leader will affect the evaluations of a female candidate for a leadership position when the context of leadership is male in gender-type, but not when the context is female in gender-type (Hypothesis 3.1).

Method

Participants and design.

We recruited one hundred and fifty-seven (103 female, 54 male)[11] participants with a mean age of 19.49 from a large Northeastern university. Of these participants, 47.1% identified

---

[11] Data for Study 3 was collected throughout the course of one academic semester, resulting in a total of 162 participants.

as Asian, 33.1% as White, 10.8% as Hispanic, 3.8% as Black, and 5.1% identified as other racial

categories. Five additional participants completed the study but were excluded from analyses

after responding incorrectly to a manipulation check. The study was a 2 x 2 between-subjects

design with performance of a female leader (success or failure) and gender-type of context (male

or female company) as the independent variables. Participants were randomly assigned to one of

four experimental conditions.

Procedure.

The procedure for Study 3 was similar to that of Study 1. Participants were told they

would be evaluating and making screening decisions regarding candidates being considered for a

high-level job and were presented with information about a female leader who had been

successful or unsuccessful in the role. We reasoned that an effective way to manipulate gender

salience without changing the position was to vary the gender-type of the context. Though it was

unlikely that that the position itself would be viewed as female-typed (high-level leadership is

generally characterized as male in gender-type), the description of the organization could be

manipulated so participants would perceive the leadership context as more or less male gender-

typed, and hence increase or decrease the salience of the female leader's gender. The company

always was a manufacturing company but was either male (a tool manufacturing company) or

female (a beauty products manufacturing company) in gender-type. After receiving information

about the company (see appendix B for stimuli) and the departing female leader's performance,

participants were asked to evaluate a female leadership candidate. Participants were then thanked

for their participation and debriefed.

Dependent Measures.

The dependent measures were the same as those of Study 1: a job-fit scale including the same three items ($\alpha = .83$) and a screening recommendation (see appendix C). We used the same manipulation checks as in previous studies. To check the manipulation for gender-type of context, participants were asked to rate how masculine or feminine they thought the company was on a bipolar adjective scale of 1 ("masculine") to 7 ("feminine").

Results

Preliminary Analyses.

One participant responded incorrectly when asked to indicate the leader's name and was therefore excluded. Four participants were excluded from subsequent analyses after rating the leader's performance inaccurately[12]. Participants' ratings of the departing leader's performance on a scale of 1 ("very poorly") to 7 ("very well") indicated that participants rated successful performance ($M = 6.77$, $SD = 0.53$) significantly more positively than unsuccessful performance, ($M = 1.86$, $SD = 0.81$), $t(155) = 44.71$, $p < .001$, $d = 7.18$, CI $= [-5.13 – -4.69]$. In addition, participants rated the beauty products company as significantly more feminine ($M = 6.17$, $SD = 0.87$) than the home improvement company ($M = 3.89$, $SD = 1.18$), confirming that the gender-type manipulation was successful. $t(155) = 13.77$, $p < .001$, $d = 2.21$, CI $= [-2.61 – -1.95]$.

No significant main effects or interactions were found when including participant gender into the analyses of our dependent measures. The analyses that follow are therefore collapsed by participant gender.

Dependent Measures.

Perceived job-fit of the female candidate. We conducted a two-way ANOVA of the job-fit ratings of the female candidate with leader performance and gender-type of context as the

---

[12] Four participants rated the failing leader's performance above the midpoint of the scale (4)

independent variables. The analyses revealed a significant main effect of leader performance, $F(1,153) = 6.52$, $p = .01$, $\eta_p^2 = .04$, and gender-type of context, $F(1,153) = 8.70$, $p = .004$, $\eta_p^2 = .05$. These effects were qualified by a marginally significant interaction between leader performance and gender-type of context, $F(1, 153) = 3.51$, $p = .06$, $\eta_p^2 = .02$. Pairwise comparisons conducted to test our hypothesis supported our prediction, indicating that job-fit perceptions of the female candidates were affected by the performance of a female leader only when the context was male-typed. Replicating previous findings, when the context of the leadership position was male in gender-type, the female candidate was deemed a significantly better fit following exposure to a successful than unsuccessful female leader, $t(153) = 3.14$, $p = .002$, $d = .51$, CI $= [-1.09 - -0.25]$. However, when the context was not male-typed, the performance of a female leader had no effect on the job-fit perceptions of the female candidate for the position, $t(153) = 0.48$, $p = .63$, $d = .08$, CI $= [-0.53 - 0.32]$. See Table 2 for means and standard deviations of all conditions.

     Screening recommendation for the female candidate. Screening recommendations followed the same pattern as the job-fit ratings. A two-way ANOVA of leader performance and gender-type of context on the screening recommendations of the female candidate yielded a significant main effect of leader performance, $F(1,153) = 5.61$, $p = .02$, $\eta_p^2 = .04$, and a significant interaction between leader performance and gender-type of context, $F(1,153) = 3.83$, $p = .05$, $\eta_p^2 = .02$. As hypothesized, female candidates for a leadership position in the male-typed context received significantly more positive screening recommendations after participants were exposed to a successful than an unsuccessful female leader, $t(153) = 3.06$, $p = .003$, $d = .49$, CI $= [-1.33 - -0.29]$. However, when the gender-type of the context was not male-typed, the performance of a female leader had no significant effects on participants' screening

recommendations of the female candidate, $t(153) = 0.29$, $p = .77$, $d = .05$, CI = [-0.60 – 0.45]. Means and standard deviations for each condition are reported in Table 2.

Discussion

Results from Study 3 indicate that evaluative generalization occurred only in contexts where being a woman is novel and thought to be incongruent with the leadership role, and the degree to which gender is salient is therefore heightened. Indeed, when there was no longer what is typically thought to be a mismatch between a female leader and the role (i.e., when the leadership context was not strongly male in gender-type), we did not find evidence of evaluative generalization from female leaders to female candidates. These results are consistent with the idea that gender salience is a necessary condition for the evaluative generalization process to unfold – that evaluative generalization from the performance of a female leader to the evaluation of a female candidate depends not only on their shared gender, but also on the salience of their gender.

Study 4

Thus far, our studies support the idea that the evaluative generalization from the performance of a female leader to perceptions of a female candidate for a leadership position occurs when there is a combination of gender salience and shared group membership. However, we have tested our predictions only with female candidates, leaving several important possibilities unexplored. It is possible that the salience of a female leader's gender affects the subsequent evaluations of any individual, regardless of group membership. If so, then exposure to a female leader who has failed or has succeeded would differentially affect reactions to male as well as female candidates to leadership positions. Furthermore, it is possible that the necessity for gender salience is limited to evaluative generalization among women, but that such

generalization occurs among men whenever there is group-based similarity, regardless of gender salience. If so, then evaluative generalization should occur between male leaders and male leadership candidates. To rule out these alternative conceptualizations of the evaluative generalization process and provide further support for our ideas, in Study 4 we included a male leader and a male candidate to our experimental design.

The goal of Study 4 was both to provide a replication of our previous findings and to further explore the role of gender salience and shared group membership on evaluative generalization, demonstrating the necessity of each to this process. Because of gender salience, we expected that exposure only to a woman, not a man, in a male-typed leadership position would affect a subsequent candidate's evaluations, and because of the necessity of shared group membership, we expected that the exposure to the female leader would affect only the evaluation of female candidates. Moreover, because a male leader's gender is unlikely to be salient in a male-typed leadership position, we expected that exposure to a male leader would have no effect on the evaluations of candidates, even when the candidate is male and the two of them share the same group membership. We therefore hypothesized that candidate evaluations will vary when the leader is successful or unsuccessful only when the leader and candidate both are female (and both gender salience and shared group membership are present, Hypothesis 4.1a), not when the leader is female and the candidate male (when there is high gender salience, but no shared group membership, Hypothesis 4.1b), nor when the leader is male and the candidate is male (when there is low gender salience but shared group membership, Hypothesis 4.1c).

In addition, Study 4 enabled us to retest our hypotheses about the benefits of exposure to a female leader relative to a male leader. Specifically, it allowed us to determine whether the failure to find positive effects for exposure to a successful female leader observed in Study 1

would be repeated in the present study. As in Study 1, we predicted that exposure to a female

leader to result in more positive evaluations and outcomes for female candidates than exposure to

a male leader when the leader was successful (Hypothesis 4.2), and we expected exposure to a

female leader to result in more negative evaluations and outcomes for female candidates than

exposure to a male leader when the leader was unsuccessful (Hypothesis 4.3).

Given that participants rated both male and female candidates in this study, we also were

able to obtain participants' preferences between them for further screening. In line with our

hypotheses for the ratings measures, we expected leader performance to make a difference in the

likelihood of choosing a female candidate over a male candidate only when the leader was a

woman. Specifically, we expected that the likelihood of choosing a female candidate will be

greater when participants are exposed to a successful (vs. unsuccessful) female leader (H4.4), but

that no differences will emerge as a result of a male leader's performance (H 4.5). We also

expected that when a leader has failed, female candidates will be less likely to be chosen if the

leader is female rather than male (H4.6), and when a leader has succeeded, female candidates

will be more likely to be chosen if the leader is female than male (H4.7).

Method

Participants and design.

Two hundred and one (114 female, 87 male)[13] participants with a mean age of 19.67

years were recruited from a large Northeastern university. Of these participants, 45.8% were

Asian, 36.3% were White, 6.5% were Hispanic, 4.5% were Black, and 6% identified as another

racial group. Five additional participants completed the study but were excluded from analyses

---

[13] A simulation analyses indicated that 160 participants were needed to obtain 85% power for detecting the expected effect (with $\eta_p^2 = .05$) in a 2x2x2 mixed design. After reaching the desired sample size, we continued recruiting male participants to obtain a more balanced gender distribution in our final sample.

after giving an incorrect response to a manipulation check. The study was a 2 x 2 x 2 mixed

design with leader gender (female or male) and leader performance (success or failure) as the

two between-subject variables, and candidate gender (female or male) as the within-subject

variable. Participants were randomly assigned to one of the four between-subject conditions.

Procedure.

The procedure was the same as Study 1 except that participants were asked to review

both a female and a male candidate for the CEO position. Again, participants learned about a

male or female leader who had been successful or unsuccessful in the leadership position. They

then reviewed two potential candidates. The information provided about the candidates was

designed to be parallel in its content (e.g. schools attended, degrees earned, and relevant work

experience) and its presentation was counterbalanced to appear equally often for male and

female candidates in each condition. The order in which the candidates were seen by participants

also was counterbalanced. We attached a photograph to the candidate profile to reinforce our

gender manipulation and ensure that participants distinguished between the candidates and the

leader[14] (see appendix B). After reading about each candidate, participants were asked to

complete a brief questionnaire containing our dependent measures. Once participants had

completed the study, they were thanked for their participation and debriefed.

Dependent Measures.

The questionnaire used to evaluate the candidates was identical to Studies 1, 3, and 4 and

was given immediately after participants read about each candidate. The same three items were

used to compose the scale of perceived job-fit of each candidate ($\alpha = .83$ for job-fit of the female

---

[14] The photos were again pretested on a sample of college students to ensure they were matched in perceived age, attractiveness, and competence

candidate and α = .80 for job-fit of male candidate) and we again asked participants to indicate the strength of their screening recommendation.

In addition, a new measure was included which asked participants to make a choice between the candidates ("candidate preference"). After reviewing both candidates, participants answered the question: "If you had to choose between the two candidates you've seen so far, which one would you select to undergo formal review?". The answer was a forced choice between the female candidate and the male candidate.

The specific questions asked for each measure are reported in appendix C. The manipulation and stimulus checks were the same as those used in previous studies.

Results

Preliminary Analyses.

All but three participants (excluded from subsequent analyses) chose the name of the departing CEO correctly, indicating that our leader gender manipulation was successful. The leader performance manipulation also had the intended effect. With two exceptions (excluded from subsequent analyses[15]), participants responded consistently with condition, rating the failing leader's performance below the midpoint of the scale (4), and the successful leader's performance above the midpoint of the scale. An analysis of the performance ratings further confirmed that participants rated the departing leader's performance significantly more positively when they succeeded ($M = 6.83$, $SD = 0.43$) than when they failed ($M = 1.59$, $SD = 0.65$), $t(198) = 67.25$, $p < .001$, $d = 9.58$, CI = [-5.39 − -5.08].

---

[15] One participant rated the failing leader's performance above the midpoint of the scale (4) and one participant rated the successful leader's performance below the midpoint of the scale.

Analyses including participant gender in the model revealed a significant main effect of participant gender for the measures pertaining to the screening recommendation. Specifically, male participants in this study were harsher in their screening recommendations than female participants. No significant interactions were found between participant gender and any of our independent variables, so we collapsed across participant gender for all analyses presented here.

Dependent Measures.

We ran a three-way repeated measures ANOVA, with leader gender and leader performance as the between-subjects factors and candidate gender as the within-subjects factor on the perceived job-fit ratings and the screening recommendation ratings, and followed up with pairwise comparisons to test our specific hypotheses. We used logistic regression to test differences in the binary candidate preference. The model included leader gender (effect coded with male as -1 and female as 1), leader performance (effect coded with success as -1 and failure as 1), and the interaction between the two as predictors of candidate preference (dummy coded with 0 as choosing the male candidate and 1 as choosing the female candidate).

Perceived job-fit of the female and male candidate. Analyses yielded a significant interaction between leader gender and leader performance, $F(1,197) = 5.09$, $p = .03$, $\eta_p^2 = .03$, a significant interaction between leader performance and candidate gender, $F(1, 197) = 5.95$, $p = .02$, $\eta_p^2 = .03$, and a significant three-way interaction, $F(1,197) = 4.29$, $p = .04$, $\eta_p^2 = .02$, that qualified the two lower order interactions and indicated that the effect of leader performance was affected by the particular combinations of leader gender and candidate gender.

Pairwise comparisons provided support for our hypotheses. In line with H4.1a, exposure to a successful (vs. unsuccessful) leader resulted in higher job-fit ratings for candidates when the leader and candidate both were female, $t(197) = 3.36$, $p = .001$, $d = 0.48$, CI = [-0.99 – -0.26].

However, as predicted, leader performance had no effect on job-fit ratings when the leader was female and the candidate male, $t(197) = 0.29$, $p = .77$, $d = 0.04$, CI = [-0.28 – 0.38] (H4.1b), or when the leader was male and the candidate male, $t(197) = 1.15$, $p = .25$, $d = 0.16$, CI = [-0.14 – 0.52] (H4.1c). Thus, as we had expected, neither gender salience nor shared group membership on their own produced differences in ratings; only when both elements were present did the success or the failure of the leader impact a candidate's perceived job-fit.

Additional pairwise comparisons were conducted to test H4.2 and H4.3. Consistent with Study 1, the job-fit ratings of the female candidate were significantly lower after exposure to a failing female (vs. male) leader, $t(197) = 2.33$, $p = .02$, $d = .33$, CI = [-0.79 – -0.07]. Also consistent with the results of Study 1, but counter to our hypotheses, exposure to a successful woman (vs. man) did not have a significant effect on the job-fit ratings of the female candidate, although results showed a trend in the predicted direction (H4.2), $t(197) = 1.77$, $p = .08$, $d = .21$, CI = [-0.04 – 0.69]. All means and standard deviations are presented in Table 3.

Screening recommendation for the female and male candidate. Analyses revealed a significant interaction between leader gender and performance, $F(1,197) = 6.83$, $p = .01$, $\eta_p^2 = .03$, a significant interaction between leader performance and candidate gender, $F(1, 197) = 5.30$, $p = .02$, $\eta_p^2 = .02$, and a significant three-way interaction, $F(1,197) = 8.99$, $p = .003$, $\eta_p^2 = .04$.

We again found support for H4.1a, H4.1b, and H4.1c, indicating that both gender salience and shared group membership were necessary for a leader's performance to affect a leadership candidate's screening recommendations. Specifically, exposure to a successful (vs. unsuccessful) leader resulted in more favorable screening recommendations of candidates when both the leader and the candidate were female, $t(197) = 3.36$, $p = .001$, $d = .48$, CI = [-1.18 – -0.31], but not when the leader was female and the candidate was male, $t(197) = 0.82$, $p = .41$, $d =$

.12, CI = [-0.23 – 0.56], or the leader and candidate were both male, $t(197) = 1.44$, $p = .15$, $d = .21$, CI = [-0.11 – 0.68].

In contrast to earlier findings, H4.2 and H4.3 both were supported. Not only were female candidates recommended less highly after exposure to a failing female than male leader, $t(197) = 2.92$, $p = .004$, $d = .42$, CI = [-1.09 – -0.21], but they were rated more highly following exposure to a successful female than male leader, $t(197) = 2.30$, $p = .02$, $d = .33$, CI = [0.07 – 0.94]. Means and standard deviations for each condition can be found in Table 3.

Candidate preference. Analyses yielded a significant interaction between leader gender and leader performance, $B = -.33$, $SE = .14$, $Wald\ \chi^2(1) = 5.36$, $p = .02$. Providing support for H4.4, the likelihood of choosing a female candidate over a male candidate was significantly higher following exposure to a successful (vs. unsuccessful) female leader, $B = -.49$, $SE = .21$, $Wald\ \chi^2(1) = 5.61$, $p = .02$, but the performance of the male leader had no effect on candidate preference, $B = .18$, $SE = .20$, $Wald\ \chi^2(1) = 0.79$, $p = .37$ (H4.5). Consistent with the pattern of our previous results, and supporting H4.7, when a leader had failed, female candidates were less likely to be selected if the failing leader was female than male, $B = -.41$, $SE = .20$, $Wald\ \chi^2(1) = 3.93$, $p = .05$. In addition, we failed to find support for H4.6: when the leader had been successful, female candidates were no more likely to be selected if the successful leader was female than male, $B = .26$, $SE = .20$, $Wald\ \chi^2(1) = 1.67$, $p = .19$ (see Figure 3).

Discussion

As we had predicted, evaluative generalizations were consistently indicated only among women in male-typed leadership. Finding that exposure to a female leader affected the evaluations and job outcomes of a woman, but not a man, seeking access to leadership, suggests that gender salience of the leader is not, by itself, the driving factor of evaluative generalizations;

shared group membership is required. The results also support the idea that shared group membership, by itself, is not sufficient to produce evaluative generalization: unlike female candidates, male candidates were unaffected by the performance of same-gender leaders. Thus, these findings build on those of Study 3, giving further credence to our contention that evaluative generalization is triggered by the combined effects of gender salience and shared group membership.

These findings strengthen the evidence for evaluative generalization among women in male-typed leadership. However, they suggest that the effects of exposure to a successful or unsuccessful female leader do not follow the symmetrical pattern we had hypothesized. While unsuccessful performance by a female leader had a consistently negative effect on the evaluations of a female candidate, the benefits of exposure to a successful female leader were not as clear-cut. Successful performance by a female leader resulted in more positive outcomes for a female candidate in the case of the screening recommendation, but the predicted pattern of results was not evident for perceptions of job-fit or indications of candidate preference.

Study 5

Because the objective of this research is to determine the effect of exposure to female leaders on evaluations of other women aspiring to be leaders, we have until now tested our ideas about evaluative generalization with a focus on women. However, if our ideas are correct, evaluative generalization should also occur for men when they are in positions that make their gender salient. The objective of Study 5 was to test the effects of gender salience and shared group membership on the evaluation of male leadership candidates.

To conduct this study, it was necessary to identify a leadership position for which being a man makes gender salient. This was a difficult task. Most leadership roles, and particularly high-

level ones, tend to be highly masculine in gender-type – they are heavily populated by men and

seen as requiring attributes that are strongly associated with men (Eagly & Karau, 2002; Koenig

et al., 2011; Lyness & Heilman, 2007). As a consequence, gender salience for men in top

leadership is likely to be rare, and it is not clear that it is ever as pronounced as it is for female

leaders. To test whether evaluative generalization occurs among men, it was therefore necessary

to select a context that could effectively counteract the perceived masculinity of the leadership

role. With this in mind, we chose to test our hypotheses in a highly feminine leadership context.

To be as conservative as possible in the test of the generalization process for male

leaders, we chose to conduct a study that mirrored Study 4 and to state parallel hypotheses. Our

design and hypotheses were analogous to those of Study 4. We expected that in a female-typed

context the performance of a male, but not female, leader would differentially affect the

evaluations and job outcomes of male leadership candidates. Specifically, we predict that a

leader's successful (vs. unsuccessful) performance will lead to more positive ratings of a

leadership candidate when both the leader and the candidate are male (Hypothesis 5.1a), but not

when the leader is male and the candidate is female (Hypothesis 5.1b), nor when both the leader

and the candidate are female (Hypothesis 5.1c).

Paralleling our hypotheses for women in male-typed leadership, we also predicted that

the evaluation of male candidates will be more positive after exposure to a successful male (vs.

female) leader (Hypothesis 5.2), and more negative after exposure to an unsuccessful male (vs.

female) leader (Hypothesis 5.3).

Finally, we predicted that the likelihood of choosing a male candidate over a female

candidate to undergo further review would be greater after exposure to a successful (vs.

unsuccessful) male leader (Hypothesis 5.4), but expected no difference in candidate preference

depending on the performance of a female leader (Hypothesis 5.5). We also predicted a greater likelihood of choosing a man over a woman after exposure to a successful male (vs. female) leader (Hypothesis 5.6), and a lower likelihood of choosing a man over a woman after exposure to an unsuccessful male (vs. female) leader (Hypothesis 5.7).

Method

Participants and design.

Two hundred and nineteen (146 female, 71 male, 2 undetermined)[16] participants with a mean age of 19.43 years were recruited from a large Northeastern university. 38.4% participants identified as Asian, 27.4% as White, 12.3% as Hispanic, 6.4% as Black, and 14.6% identified as another racial group. Four additional participants were excluded for responding incorrectly to one or more manipulation checks. The study used a 2 x 2 x 2 mixed design identical to Study 4. Participants were randomly assigned to one of the four between-subject conditions.

Procedure.

The procedure and materials were identical to those of Study 4 except for the leadership context – in Study 5, participants read about the departing CEO (male or female) of a female-typed company. Data from Study 3 suggested that participants viewed the beauty products company as highly female in gender-type ($M = 6.17$, on a 1 to 7 "masculine-feminine" bipolar scale). We therefore used the same female-typed company as in Study 3 to test our hypotheses (see appendix B for stimuli). The information provided about the female and male leadership candidates was the same as in Study 4 (see appendix B). After completing a questionnaire for each candidate, participants were thanked and debriefed.

---

[16] We used the same simulation analyses as in Study 4 to determine our minimum sample size (160 participants). Given the low number of male participants in the resulting sample, we continued recruiting male participants to ensure a more balanced gender distribution.

Dependent Measures.

We used the same dependent measures and manipulation checks as in Study 4 ($\alpha = .87$

for job-fit of the female candidate and $\alpha = .85$ for job-fit of male candidate; see appendix C)

Results

Preliminary Analyses.

Four participants responded to the CEO gender and/or performance manipulation checks

in a way that was inconsistent with condition and were therefore excluded from analyses[17]. In

line with our manipulations, successful leader performance was rated significantly more

positively ($M = 6.78$, $SD = 0.58$) than unsuccessful leader performance ($M = 1.83$, $SD = 0.72$),

$t(216) = 56.14$, $p < .001$, $d = 7.64$, CI = [-5.12 – -4.78].

No consistent main effects or interactions were found when we included participant

gender into the analyses for our dependent measures. We therefore combined the data for female

and male participants.

Dependent Measures.

We conducted a three-way repeated measures ANOVA on the perceived job-fit and

screening recommendations of the candidates, with leader gender and performance as the

between-subjects factors and candidate gender as the within-subjects factor. Again, to test our

specific hypotheses we followed up with paired comparisons. Candidate preference was analyzed

with logistic regression. Leader gender was effect coded with female as -1 and male as 1, leader

performance was effect coded with success as -1 and failure as 1, and the likelihood of choosing

---

[17] Two participants responded incorrectly to both manipulation checks. One participant selected the wrong name for
the departing leader. Another participant rated the failing leader's performance above the midpoint of the scale (4).

the male candidate over the female candidate was dummy coded with 0 as choosing the female

candidate and 1 as choosing the male candidate.

   Perceived job-fit of the female and male candidate. The results of the ANOVA

indicated a significant main effect of candidate gender, whereby female candidates received

more positive evaluations than male candidates, $F(1,215) = 21.40$, $p < .001$, $\eta_p^2 = .09$. There also

was a significant three-way interaction, $F(1,215) = 8.45$, $p = .004$, $\eta_p^2 = .04$.

   Mirroring the pattern of results for female leaders, pairwise comparisons provided

support for H5.1a, H5.1b, and H5.1c. Exposure to a successful (vs. unsuccessful) leader led to

more positive fit ratings for the leadership candidate when both leader and candidate were male,

$t(215) = 2.22$, $p = .03$, $d = .29$, CI = [-0.83 – -0.05], but not when the leader was male and

candidate was female, $t(215) = 0.21$, $p = .83$, $d = .03$, CI = [-.32 – 0.39]. Furthermore, a female

leader's performance did not affect the perceptions of job-fit of the female candidate, $t(215) =$

0.95, $p = .35$, $d = .13$, CI = [-0.54 – 0.19]. Also consistent with our findings for female leaders,

our data failed to provide support for H5.2 but supported H5.3. Specifically, perceptions of job-

fit of the male candidate were significantly lower after exposure to an unsuccessful man (vs.

woman), $t(215) = 2.23$, $p = .03$, $d = .30$, CI = [0.05 – 0.84], but were not significantly higher

after exposure to successful man (vs. woman), $t(215) = 0.94$, $p = .35$, $d = .13$, CI = [-0.58 – 0.21].

Means and standard deviations are presented in Table 4.

   Screening recommendation for female and male candidate. Analyses revealed only a

main effect of candidate gender, $F(1,215) = 20.44$, $p < .001$, $\eta_p^2 = .09$. Specifically, screening

recommendations were significantly more positive for the female candidate than the male

candidate, regardless of the gender or performance of the previous leader. The three-way

interaction, critical to our predictions, was not statistically significant, $F(1,215) = 0.42$, $p = .52$,

$\eta_p^2 = .002$.

     Candidate preference. A logistic regression yielded a marginally significant interaction

between leader gender and leader performance, $B = -.27$, $SE = 0.12$, *Wald χ^2*$(1) = 3.26$, $p = .07$.

We conducted simple effects analyses to test our specific hypotheses. These analyses did not

provide support for our hypotheses: the likelihood of choosing a man over a woman was not

significantly affected by either the gender or the performance of a previous leader (all ps > .15).

Discussion

     Results for Study 5 were mixed. Although the performance of a male leader significantly

affected the perceptions of job-fit of a male leadership candidate, it had no effect on screening

recommendations or candidate preferences. These findings show that exposure to women and

men in counter-stereotypical leadership contexts do not have fully analogous effects. While

evaluative generalization appears to occur for men and women leaders alike, the breadth of this

generalization and its outcomes seems to be more equivocal for men than for women.

General Discussion

     More than ever before, women have reached high-level leadership positions, giving rise

to the belief that female leaders have finally broken the glass ceiling not only for themselves, but

also for women as a group. Putting this belief to the test, we examined if, when, and how

exposure to a woman in top leadership affects the evaluations of another woman seeking access

to leadership. In line with our predictions, we find consistent evidence that there is an evaluative

generalization between women in leadership, and that it is not only the presence, but also the

performance of a female leader that determines her effect on the evaluations of subsequent women.

Our results indicate that differences in a leader's performance affected reactions to a woman candidate for a leadership position when the leader was female, not male, and the leadership position was male in gender-type. Reactions to women candidates were not affected by the performance of a male leader or of a female leader in a domain that is not male-typed. These results support our argument that gender salience, produced by the scarcity of women in leadership and the perceived incongruence of a woman in a traditionally male role, is central to the evaluative generalization process. Importantly, our results indicated that evaluative generalization is not the product of gender salience alone. The absence of evaluative generalization from the performance of female leaders to male candidates demonstrated that shared group membership also is necessary. However, finding that there was no evaluative generalization from male leaders to male leadership candidates in a male-typed context made clear that shared group membership is not sufficient to induce evaluative generalization; gender salience also is required.

In addition, our results suggest that gender stereotypes affect the content of evaluative generalization between female leaders. When the female leader fulfilled stereotype-based expectations and was not successful, female candidates were characterized as less agentic than when the female leader challenged stereotype-based expectations and was successful. Our data further showed that these perceptions of lesser agency were associated with more negative outcomes for female candidates, promoting perceptions that they are a worse fit for the leadership position and resulting in less favorable screening recommendations.

Our results demonstrate that the evaluations of a woman aspiring to become a leader are more positive after exposure to a successful female leader than after exposure to an unsuccessful female leader. Yet to fully test assess the question of whether the presence of a female leader indeed "breaks the glass ceiling", we also examined whether following a woman is better for a female leadership candidate than following a man. Our results, which confirmed that the answer to this question depended upon whether the performance of the female leader was successful or unsuccessful, generally indicated no. When a female leader had been successful, her presence tended to have little effect on the fit evaluations, screening recommendation, and preference for female leadership candidates as compared to when the successful leader was male. Moreover, when a female leader was unsuccessful, her presence had detrimental effects – in such cases, it was actually worse for a female candidate to follow a woman than a man. Thus, although we had predicted that the effect of a female leader's performance would be symmetrical, our results indicated that only the negativity associated with failing female leaders was consistently generalized. These findings strongly suggest that for women aspiring to become leaders, the gender and performance of their predecessors matter. Notably, our results also suggest that the failure of a female leader may have a disproportionate impact, influencing evaluations of other women more than the success of a female leader.

To further test our ideas, in Study 5 we examined the effects of exposure to male leaders in a female-typed domain. The results were mixed. Paralleling our findings for women, the performance of a man, but not a woman, in a female-typed leadership context affected whether a male, but not female, candidate was seen as a good fit for the position. Also, in line with the research focusing on women, there was indication that it was the negativity of failure but not the positivity of success that was generalized. However, in contrast to the job-fit findings, the

performance of a male leader had no impact on whether the male candidate would be

recommended or preferred over a female candidate to continue in the selection process. Thus,

while analogous results were found for one outcome, they were not found for the others.

It is interesting to consider why the pattern of results for male leaders converged with

those for female leaders only some of the time. It is possible that, in top leadership, gender

simply is not as salient for men as it is for women. Top leadership tends to be inherently male in

gender-type, and even in a female-typed leadership context, a male CEO may not seem to be as

novel or incongruous as a female CEO in a male-typed context. Moreover, unlike women in

male-typed companies, men are not scarce in top leadership of female-typed companies.

Although the share of female leaders tends to be higher in companies that are thought of as

female in gender-type (e.g., cosmetics, lingerie), men continue to be overrepresented in their

boards and executive teams (Cheng, 2017). Thus, exposure to female and to male leaders in

gender-incongruent organizational contexts may not be equivalent in their effects on gender

salience, and consequently may lead to weaker effects for male than for female leadership

candidates. This difference in gender salience strength may therefore account for the predicted

effect on job-fit perceptions, but not on the more outcome-oriented effects of screening

recommendations or choices about whether the candidate should remain in the selection process.

To provide a more stringent test of our ideas about the effects of exposure to a successful or

unsuccessful man on a male candidate, it would be useful to move away from top leadership and

focus on positions, roles, and contexts that are unequivocally female in gender-type (e.g.,

nursing, early childhood education, the domestic sphere).

It is important to note that a potential limitation to the generalizability of these results

relates to the specific characteristics of our sample. Although the ethnic and racial composition

of our participants was quite diverse across studies (about 45% Asian, 30% White, 10% Hispanic, and 5% Black), our sample consisted mostly of college students from a large northeastern university of the United States. This implies that our results might be limited to a relatively young, well-educated sample, one that is less likely than the general population to endorse stereotype-based beliefs (Farley, Steeh, Krysan, Jackson, & Reeves, 1994; Spence & Hahn, 1997; Whitley, 1999). Further studies should examine whether the results found here replicate among older, less educated and/or non-US based participants, and to examine these effects in a real-life context, for example, an organization or a political election. It would also be beneficial to directly examine the role of stereotype endorsement and/or specific types of sexism on people's tendency to engage in evaluative generalizations between female leaders. We would expect these findings to be more pronounced among participants who hold more negative views about women in leadership.

Theoretical implications

Our findings extend earlier work on the association between group membership salience and perceptions of similarity among group members (e.g., Kanter, 1977; Tajfel, 1969; Taylor et al., 1978), as well as research on the effects of exposure to counter-stereotypical exemplars (e.g., Critcher & Risen, 2014; Dasgupta & Greenwald, 2001; Johnston & Hewstone, 1992; Rothbart, 1981). Bridging these literatures, our work provides evidence that mere exposure to a single counter-stereotypical exemplar (such as a female leader) can affect the evaluations of other individual group members. It also provides important insight into the processes underlying these "person-to-person" evaluative generalizations and the conditions under which they occur.

Past research suggests that people often are reluctant to use information about one person to judge another individual (Banaji & Bhaskar, 2000), although there is some evidence that direct, person-to-person generalization does occur under some circumstances. Evaluative generalization has been found between outgroup but not ingroup members (e.g., Chen & Ratliff, 2015; Henderson-King & Nisbett, 1996; Ratliff & Nosek, 2011) and for implicit but not explicit attitudes (Ranganath & Nosek, 2008). Our findings also demonstrate that person-to-person generalization occurs. However, contrary to prior findings, our results indicate that when gender is salient, both outgroup (men) and ingroup (women) members are quite willing to generalize between female leaders, and that they do so even when using explicit measures. Future research should consider whether people are aware that they are engaging in person-to-person generalization and, if so, the extent to which they believe these generalizations are justified.

Our findings further indicate that person-to-person generalizations are not bound by physical or "objective" similarity, as previous research has suggested (e.g., Gawronski & Quinn, 2013; Lewicki, 1985, Ranganath & Nosek, 2008). The results demonstrate that evaluative generalizations occur between individual women when their shared gender is made salient, but not when it is not salient (e.g., in female-typed leadership roles). This suggests that it is the *perception* of similarity, not physical similarity per se, that is critical to the evaluative generalization process, and that context can be critical in determining when these perceptions of similarity emerge.

Although the main goal of this program of research was to examine person-to-person generalizations in the context of female leadership, our findings may have implications for women in other counter-stereotypical domains. There already is some evidence for this in the medical field: after an unfavorable outcome involving a female surgeon, physicians were found

to be less likely to refer their patients to another female surgeon, whereas similar outcomes involving a male surgeon had no impact on physician referrals (Sarsons, 2017). Furthermore, to the extent that gender salience elicits perceptions of within-group similarity, exposure to a female leader may also have consequences for women vying for similar, but not identical positions. It would be interesting, for example, to examine the effects of exposure to a female CEO on the prospects of woman applying for middle or lower management positions or even for women in different occupational domains. Moreover, if group membership salience gives rise to evaluative generalization between individuals belonging to the same group, this phenomenon might also occur among other occupational minorities. For example, the evaluations of a Latino political leader might affect the likelihood of other Latinos being elected into office. Lastly, it would be important to test our ideas using different manipulations of gender salience. If our model is correct, the results we demonstrated in these studies should be evident whatever the source of gender salience for the observer.

As we mentioned in the introduction, exposure to counter-stereotypical group members can weaken general stereotypes about that group. Though results from Study 2 suggest that gender stereotypes play an important role in the content of what is being generalized between a female leader and a female leadership candidate, our studies do not address whether stereotypes themselves are affected during the generalization process. Thus, there is the possibility that evaluative generalizations are not as direct as we have contended, and that they instead occur via the strengthening or weakening of stereotypes. That is, exposure to a counter-stereotypical exemplar (e.g., a successful female leader) could lead to an erosion of gender stereotypes which, in turn, reduces biased evaluations of a female leadership candidate. Likewise, exposure to a stereotypical exemplar (e.g., an unsuccessful female leader) could solidify stereotype-based

beliefs and then lead to more negative evaluations of an aspiring female leader. Although past research suggests that it may take more than a one-time exposure to a counter-stereotypical target to elicit a lasting change in stereotypes (Weber & Crocker, 1983), future research should examine the extent to which the weakening or strengthening of stereotypes is involved in person-to-person generalizations.

If, however, evaluative generalization occurs independently of stereotype change, as we are positing, our findings have important implications for congruity models of gender discrimination. These models propose that gender bias against women in traditionally male domains results from a perceived lack of fit between their performance-related attributes and the requirements of the job to be done (Eagly & Karau, 2002; Heilman, 2001, 2012). However, we have shown that when evaluative generalization occurs, presumptions about the performance-related attributes of a woman are affected not only by gender stereotypes, but also by the behavior of a female predecessor. Thus, expectations about whether a woman will be a competent leader and the evaluation of her actual performance might, under certain circumstances, be shaped directly by the performance of another woman. Future research should examine the extent to which exposure to a single stereotype confirming or disconfirming woman takes precedence over more general stereotype-based expectations in triggering lack of fit perceptions.

Generalization of success vs. generalization of failure

Our results suggest that the generalization from the performance of a female leader to the evaluations of another woman did not follow the symmetrical pattern we had predicted. Across all studies, we found strong, consistent evidence that exposure to a failing female leader was more detrimental to another woman's evaluations and leadership opportunities than following a

male leader. However, the expected boost from following a successful female (vs. male) leader was not evident. While we found one result evidencing this "boost" in Study 4, we did not find it for other dependent variables in Study 4 or for any of the outcomes in Study 1. These findings suggest that the processes underlying evaluative generalizations among female leaders are not necessarily symmetrical, and that while a previous woman's failure easily generalizes to an aspiring female leader's evaluation, the success of a previous woman leader is not as readily generalized.

Although counter to our predictions, these results are consistent with past research on the differential weighting of negative vs. positive information, with negative information carrying more weight than positive information in a vast array of domains (for a review, see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). However, how people weigh and use negative vs. positive information to form their impressions of others depends, to a large extent, on the degree to which it is consistent with group-based stereotypes. Regardless of its valence, stereotype-consistent information is more readily attended to and remembered (Plaks, et al., 2001; Uhlmann & Cohen, 2005). In line with this idea, there is research to suggest that only stereotype-consistent information will be generalized across members of a group (Stangor & McMillan, 1992). For example, negative (but not positive) behaviors displayed by a racial outgroup member have been shown to negatively affect people's implicit attitudes towards another racial outgroup member (Chen & Ratliff, 2015; Ratliff & Nosek, 2011).

Thus, one possible explanation for the asymmetry we found is that by confirming stereotype-based expectations about women's lack of effectiveness in male-typed occupations, unsuccessful women are seen as more representative of other female leaders than successful women. Whereas the failure of a female leader is likely to reinforce stereotyped beliefs and

incompetence perceptions and therefore easily generalize to another woman, success might only refute gender stereotypes for that individual female leader. Studies have shown that under certain conditions, strong counter-stereotypical exemplars are put into a different sub-category, separate from other members of their group. This has been referred to as "subtyping", a process by which exposure to individuals who are incongruent with stereotypes about their group leads to the perception that these individuals are exceptions, and not representative of their group (e.g., Weber & Crocker, 1983; Hantzi, 1995). If this happened in our studies, exposure to a female leader who is successful would have thwarted the generalization process, accounting for our results. Future research could test this idea by examining whether successful female leaders are seen as outliers. If so, then interventions aimed at bolstering the positive effects of exposure to successful female leaders might consider including information that highlights stereotype-consistent information about these leaders (e.g., their communal attitudes or behavior) in addition to their stereotype-inconsistent success.

It also is possible the roots of the observed asymmetry lie earlier in the generalization process. Indeed, the confirmation or disconfirmation of stereotypes may have had differential effects on the extent to which gender was salient, with gender being more salient after exposure to a stereotype-confirming woman (an unsuccessful leader) than a stereotype-disconfirming woman (a successful leader). This may explain the differences we observed in the generalization of success and failure: if gender is not salient to participants after being exposed to a successful female leader, then we would not expect evaluative generalization to occur.

A different body of literature offers yet another possible explanation for the asymmetry we observed in the generalization of success vs. failure. Research has demonstrated that women who defy gender stereotypes by being successful in male-typed positions are disliked and

perceived as self-centered, cold, and manipulative, and that these negative characterizations are detrimental to their career outcomes (Heilman, Wallen, Fuchs & Tamkins, 2004; Rudman, 1998; Rudman & Glick, 1999). According to this literature, a successful female leader may be perceived as skilled, but unsuitable for a leadership position due to interpersonal deficits. It is possible, then, that exposure to a successful female leader may simultaneously help *and* hinder the evaluations of another woman through the generalization of both positive traits (e.g., higher agency) and negative traits (e.g., higher interpersonal hostility, lower likeability). If this is the case, then it may be that the benefits of exposure to a female leader's success are dampened by the simultaneous generalization of negative attributes. Future research should examine whether these negative conceptions of successful women in male-typed roles generalize to other women and, if so, whether they account for the absence of generalization for success that we observed in these studies.

Practical implications

These findings not only provide novel theoretical insights, but they also have significant real-world implications. In recent years, an increasing number of countries and organizations have taken important steps to promote gender equality in many traditionally male settings. These efforts have included a strong push for the implementation of gender diversity initiatives for governmental positions, executive boards, and recruitment for STEM fields. Although targeted placement policies such as quotas and affirmative action have been vital for the increase of women in leadership, our findings suggest that overcoming the persistent gender imbalance in high-power roles may require additional or alternative measures. Providing female exemplars, in and of itself, is no panacea for gender imbalance, and can even undermine the intended effects of these policies when women leaders are not successful. Our findings therefore provide important

lessons and valuable insights for those designing interventions aimed at counteracting gender inequity.

Our results also shed light on an interesting phenomenon sometimes observed in work settings, whereby White men are asked to step in after the perceived failure of a woman and/or racial minority. Correlational evidence suggests that when a company's performance declines during the tenure of an occupational minority, these leaders are more likely to be replaced by a White male (Cook & Glass, 2014). Though this phenomenon has been described as the "savior effect", our findings suggest that it may not be the most accurate description of what is occurring. That is, it may not be the case that the failure of an occupational minority leads to an enhancement of a White male leadership candidate, per se; rather, as our results suggest, it may instead be that an unsuccessful minority leader hampers the evaluations of subsequent minority candidates, thereby tipping the scales in favor of White male candidates. Future research should examine whether similar evaluative generalizations are observed between other occupational minorities, and whether this generalization may even cross group boundaries under conditions in which general minority status becomes the salient group category.

These findings also have important implications for female leaders themselves. While exposure to a successful female leader has been shown to positively impact women's self-perceptions of competence (Lockwood, 2006; Stout, Dasgupta, Hunsinger & McManus, 2011; Young, Rudman, Buettner & McLean, 2013), exposure to an unsuccessful female leader might reinforce women's negative stereotypes about their own abilities. Women may therefore be more reluctant to take on a leadership role after another woman has performed poorly and perhaps be more anxious about their own performance being heavily scrutinized. Moreover, past research shows that when negatively stereotyped group members are led to believe that their performance

will be used to assess their group, they feel threatened, and their performance is impaired as a result (Shapiro & Neuberg, 2007). If female leaders are aware of their perceived role as representatives of women in leadership, their actual performance might suffer, which, in turn, may fuel the negative effects of evaluative generalization. The threat of becoming a representative of their gender group may further deter women from taking on leadership roles altogether. Future research should examine the extent to which women expect evaluative generalization to occur, and whether these expectations affect their motivation to become leaders and their ability to perform well once they are in these positions.

More generally, our results show that evaluative generalization can be a subtle, yet powerful source of gender disparities. They indicate that, in traditionally male settings, men enjoy the advantage of being judged on their own individual merits while women must pay the price for another woman's missteps. Future research should consider what can be done to offset the negative effects of evaluative generalization. Our results suggest that one of the most straightforward ways to avoid evaluative generalization among women would be to reduce the extent to which gender is made salient. It therefore would be interesting to determine whether organizational descriptions of leadership roles as less exclusively male result in decreased evaluative generalization among female leaders. Our results also suggest that the work culture and its implications for the interpretation of failure can affect reactions based on evaluative generalization. A work culture in which talent is viewed as malleable and mistakes are welcomed as an opportunity for growth rather than a sign of incompetence (see Canning et al., 2020; Emerson & Murphy, 2015) may be an effective antidote to the negative effects of a female leader's failure on another woman's evaluations.

Conclusion

In sum, our findings demonstrate that exposure to female leaders does not automatically "break the glass ceiling" and lead to positive consequences for other women, as is often assumed. The combination of gender salience and shared group membership created a context in which the performance of a female leader was generalized to the evaluations of a female candidate for a leadership position. However, the effects were not equally powerful for leader success and failure. While the consequences for women who followed an unsuccessful female leader were unequivocally negative, exposure to a successful female leader rarely was found to give rise to evaluative generalizations that had positive consequences for aspiring women. Thus, our findings shed light on an additional challenge for female leaders. They imply that, unlike men, women in leadership are not always judged on the basis of their qualifications and accomplishments. Instead, irrelevant information – in this case the performance of another, unrelated woman – can shape their evaluations and unduly influence their career prospects.

References

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Chicago: Rand McNally.

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139 –175). Cambridge, MA: Harvard University Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.

Bless, H., Schwarz, N., Bodenhausen, G. V., & Thiel, L. (2001). Personalized versus generalized benefits of stereotype disconfirmation: Trade-offs in the evaluation of atypical exemplars and their social groups. *Journal of Experimental Social Psychology*, *37*(5), 386-397.

Bodenhausen, G. V., Schwarz, N., Bless, H., & Wänke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals. *Journal of Experimental Social Psychology*, *31*(1), 48-63.

Bongiorno, R., Bain, P. G., & David, B. (2014). If you're going to be a leader, at least act like it! Prejudice towards women who are tentative in leader roles. *British Journal of Social Psychology*, *53*(2), 217-234.

Brescoll, V. L., Dawson, E., & Uhlmann, E. L. (2010). Hard Won and Easily Lost The Fragile Status of Leaders in Gender-Stereotype-Incongruent Occupations. *Psychological Science*, *21*(11), 1640-1642.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972).

 Sex-role stereotypes: A current appraisal. *Journal of Social issues*, *28*(2), 59-78.

Bruckmüller, S., & Branscombe, N. R. (2010). The glass cliff: When and why women are

 selected as leaders in crisis contexts. *British Journal of Social Psychology*, *49*(3), 433-

 451.

Canning, E. A., Murphy, M. C., Emerson, K. T., Chatman, J. A., Dweck, C. S., & Kray, L. J.

 (2020). Cultures of genius at work: Organizational mindsets predict cultural norms, trust,

 and commitment. *Personality and Social Psychology Bulletin*, *46*(4), 626-642.

Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black–White

 biracial group members. *Social Psychological and Personality Science*, *6*(5), 544-550.

Cheng, M. (2017). Women are making over the beauty industry's boy's club. FiveThirtyEight.

 Retrieved from https://fivethirtyeight.com/features/women-are-making-over-the-beauty-

 industrys-boys-club/

Cohen, L. E., Broschak, J. P., & Haveman, H. A. (1998). And then there were more? The effect

 of organizational sex composition on the hiring and promotion of managers. *American

 Sociological Review*, 711-727.

Cook, A., & Glass, C. (2014). Above the glass ceiling: When are women and racial/ethnic

 minorities promoted to CEO? *Strategic Management Journal*, 35, 1080-1089.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype

 formation, and the interchangeability of group members. *Journal of Personality and

 Social Psychology*, *83*(5), 1076.

Critcher, C. R., & Risen, J. L. (2014). If he can do it, so can they: Exposure to counterstereotypically successful exemplars prompts automatic inferences. *Journal of Personality and Social Psychology*, *106*(3), 359.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642-658.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology*, *81*(5), 800.

DeVaro, J., & Waldman, M. (2009). The Signaling Role of Promotions: Further Theory and Empirical Evidence. *Journal of Labor Economics*, 30(1), 91-147.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and social psychology bulletin*, *26*(10), 1171-1188.

Eagly, A. H. (2007). Female leadership advantage and disadvantage: Resolving the contradictions. *Psychology of women quarterly*, *31*(1), 1-12.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American psychologist*, 75(3), 301–315.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The developmental social psychology of gender*, 123-174.

Ellemers, N., Rink, F., Derks, B., & Ryan, M. K. (2012). Women in high places: When and why promoting women into top positions can harm them individually or as a group (and how to prevent this). *Research in Organizational Behavior*, *32*, 163-187.

Emerson, K. T., & Murphy, M. C. (2015). A company I can trust? Organizational lay theories moderate stereotype threat for women. *Personality and Social Psychology Bulletin*, *41*(2), 295-307.

Farley, R., Steeh, C., Krysan, M., Jackson, T., & Reeves, K. (1994). Stereotypes and segregation: Neighborhoods in the Detroit area. *American Journal of Sociology*, *100*(3), 750-780.

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. *Advances in experimental social psychology*, *23*, 1-74.

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.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*(1), 120-125.

Hantzi, A. (1995). Change in stereotypic perceptions of familiar and unfamiliar groups: The

    pervasiveness of the subtyping model. *British Journal of Social Psychology*, *34*(4), 463-

    477.

Haslam, S. A., & Ryan, M. K. (2008). The road to the glass cliff: Differences in the perceived

    suitability of men and women for leadership positions in succeeding and failing

    organizations. *The Leadership Quarterly*, *19*(5), 530-546.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In Staw, B., &

    Cummings, L. Eds. *Research in organizational behavior* (Vol. 5). Greenwich, CT: JAI.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's

    ascent up the organizational ladder. *Journal of Social Issues*, 57(4), 657–674.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational

    Behavior*, 32(2012), 113–135.

Heilman, M. E., Manzi, F. & Braun, S. (2015). Presumed incompetent: Perceived lack of fit and

    gender bias in recruitment and selection. In A. Broadbridge & S. Fielden (Eds.),

    Handbook of Gendered Careers in management: getting in, getting on, getting out.

    Edward Elgar Publishing.

Heilman, M. E., Manzi, F., & Caleo, S. (2019). Updating impressions: The differential effects of

    new performance information on evaluations of women and men. *Organizational

    Behavior and Human Decision Processes*, *152*, 105-121.

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.

Henderson-King, E. I., & Nisbett, R. E. (1996). Anti-Black prejudice as a function of exposure to the negative behavior of a single Black person. *Journal of Personality and Social Psychology*, *71*(4), 654.

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: a current look at men's and women's characterizations of others and themselves. *Frontiers in psychology*, *10*, 11.

Hewstone, M., Hassebrauck, M., Wirth, A., & Waenke, M. (2000). Pattern of disconfirming information and processing instructions as determinants of stereotype change. *British Journal of Social Psychology*, *39*(3), 399-411.

Huffman, M. L., Cohen, P. N., & Pearlman, J. (2010). Engendering change: Organizational dynamics and workplace gender desegregation, 1975–2005. *Administrative Science Quarterly*, *55*(2), 255-277.

Hymowitz, C., & Schellhardt, T. D. (1986, March). The glass ceiling: Why women can't seem to break the invisible barrier that blocks them from the top jobs. *The Wall Street Journal*, 1D.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: (3) Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28*, 360-386.

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.

Kline, R. B. (2011). Principles and Practice of Structural Equation Modeling, 3rd edn Guilford Press. *New York*.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological bulletin*, *137*(4), 616.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., ... & Frazier, R. S. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765.

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*(3), 563.

Linville, P. W., & Fischer, G. W. (1993). Exemplar and abstraction models of perceived group variability and stereotypicality. *Social Cognition*, *11*(1), 92-125.

Lockwood, P. (2006). "Someone like me can be successful": Do college students need same-gender role models?. *Psychology of Women Quarterly*, *30*(1), 36-46.

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluation and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91, 777–785.

Milgrom, P., & Oster, S. (1987). Job discrimination, market forces, and the invisibility hypothesis. *The Quarterly Journal of Economics*, 453-476.

Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science*, *11*(6), 462-466.

Oakes, P., & Turner, J. C. (1986). Distinctiveness and the salience of social category memberships: Is there an automatic perceptual bias towards novelty?. *European Journal of Social Psychology*, *16*(4), 325-344.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*(3), 536.

Perry, E. L., Davis-Blake, A., & Kulik, C. T. (1994). Explaining gender-based selection decisions: A synthesis of contextual and cognitive approaches. *Academy of Management Review*, *19*(4), 786-820.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751.

Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: preferences for stereotypic versus counterstereotypic information. *Journal of Personality and Social Psychology*, *80*(6), 876.

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology*, *38*(1), 141.

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*(3), 249-254.

Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and

transfer. *Personality and Social Psychology Bulletin*, *37*(12), 1692-1703.

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype

formation. *Personality and Social Psychology Bulletin*, *33*(11), 1492-1502.

Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects of

organizational performance on leaders with single versus dual-subordinate

identities. *Journal of Experimental Social Psychology*, *48*(5), 1162-1167.

Rothbart, M. (1981). Memory processes and social beliefs. In D. Hamilton (Ed.), *Cognitive

processes in stereotyping and intergroup behavior* (pp. 145-181). Hillsdale, N.J.:

Erlbaum.

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.

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.

Ryan, M. K., & Haslam, S. A. (2005). The glass cliff: Evidence that women are over-represented

in precarious leadership positions. *British Journal of management*, *16*(2), 81-90.

Sarsons, H. (2017). Interpreting signals in the labor market: evidence from medical referrals.

[Manuscript submitted for publication]. Booth School of Business, University of

Chicago.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management

　　　characteristics. *Journal of Applied Psychology*, 57, 95–100.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in

　　　management. *Journal of Social issues*, *57*(4), 675-688.

Schultheis, E. (2015, May 25). How women's groups are betting on "The Hillary Clinton

　　　Effect". The Atlantic. Retrieved from

　　　http://www.theatlantic.com/politics/archive/2015/05/how-womens-groups-are-betting-on-

　　　the-hillary-clinton-effect/457839/

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats:

　　　Implications of a multi-threat framework for causes, moderators, mediators,

　　　consequences, and interventions. *Personality and Social Psychology Review*, *11*(2), 107-

　　　130.

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

　　　college students. *Psychology of Women Quarterly*, *21*(1), 17-34.

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.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using

　　　ingroup experts to inoculate women's self-concept in science, technology, engineering,

　　　and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, *1*(S1), 173-191.

Traister, R. (2016, July 29). Hillary Is Poised to Make the 'Impossible Possible' — for Herself and for Women in America. *New York Magazine*. Retrieved from https://www.thecut.com/2016/07/hillary-poised-to-make-the-impossible-possible.html

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, 11, 249-288.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778.

The Rockefeller Foundation. (2016). Does the media influence how we perceive women in leadership?. Retrieved from https://assets.rockefellerfoundation.org/app/uploads/20161028122206/100x25_MediaLanguage_report1.pdf

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16, 474–480.

United Nations Women. (2019). *Facts and figures: Leadership and political participation.* Retrieved from http://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures#notes

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, *45*(5), 961.

Whitley, B. E. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology*, *77*(1), 126.

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of

female role models on women's implicit science cognitions. *Psychology of Women

Quarterly*, *37*(3), 283-292.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social

Psychology*, *9*(2), 1-27.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in

Psychological Science*, *10*(6), 224-228.

Zebrowitz, L. A., White, B., & Wieneke, K. (2008). Mere exposure and racial prejudice:

Exposure to other-race faces increases liking for strangers of that race. *Social

Cognition*, *26*(3), 259-275.

Table 1

Study 1: Means and Standard Deviations of Female Candidate's Perceived Job-fit and Screening

Recommendation after Exposure to a Female or Male Leader

Measure	Female leader		Male leader	
	Failure	Success	Failure	Success
Perceived job-fit	4.70 (1.17)	5.53 (0.94)	5.23 (0.91)	5.29 (1.00)
Screening recommendation	4.79 (1.25)	5.67 (1.12)	5.35 (1.03)	5.43 (1.09)

Table 2

Study 3: Means and Standard Deviations of Female Candidate's Perceived Job-fit and Screening

Recommendation after Exposure to a Female Leader in a Male-typed or Non Male-Typed

Leadership Context

Measure	Male-typed context		Non male-typed context	
	Failure	Success	Failure	Success
Perceived job-fit	4.71 (1.04)	5.38 (0.78)	5.44 (1.04)	5.54 (0.89)
Screening recommendation	4.55 (1.45)	5.36 (0.87)	5.23 (1.20)	5.31 (1.08)

Table 3

Study 4: Means and Standard Deviations of Female and Male Candidate's Perceived Job-fit and

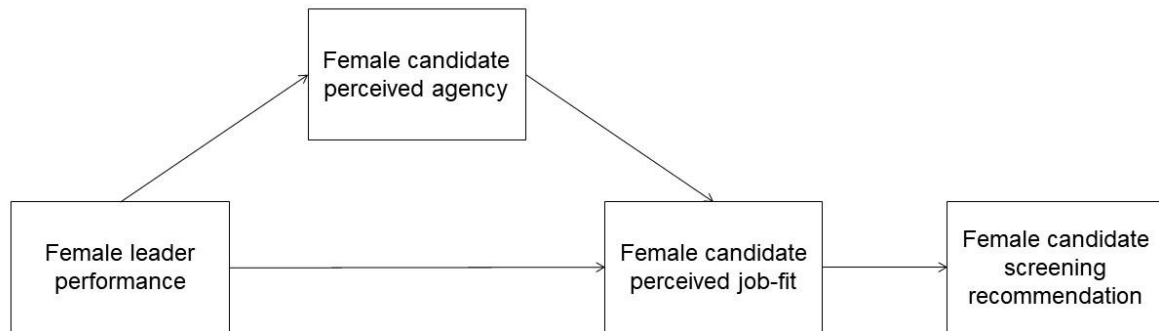*Screening Recommendation after Exposure to a Female or Male Leader in a Male-Typed*

Context

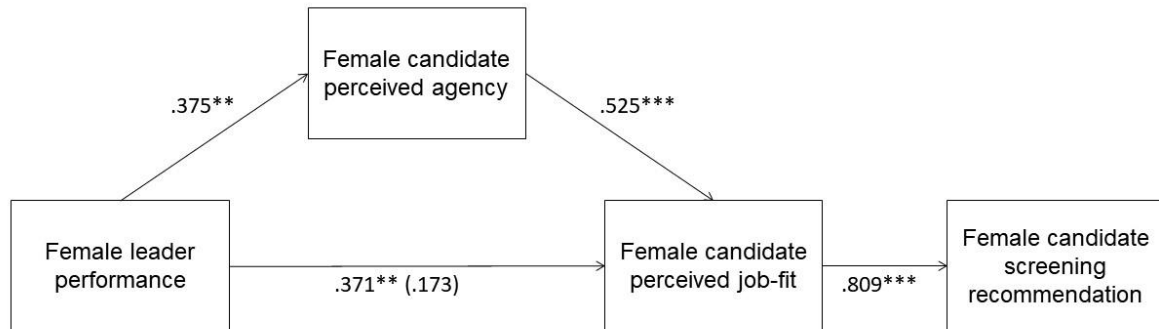|	Female candidate				Male candidate			
	Female leader		Male leader		Female leader		Male leader	
Measure	Failure	Success	Failure	Success	Failure	Success	Failure	Success
Perceived job-fit	5.09	5.71	5.52	5.38	5.59	5.54	5.45	5.26
	(0.95)	(0.79)	(1.01)	(0.93)	(0.74)	(0.82)	(0.91)	(0.86)
Screening	4.98	5.73	5.63	5.22	5.71	5.55	5.61	5.32
recommendation	(1.22)	(0.94)	(1.19)	(1.06)	(0.89)	(1.05)	(1.00)	(1.08)

Table 4

Study 5: Means and Standard Deviations of Female and Male Candidate's Perceived Job-fit

after Exposure to a Female or Male Leader in a Female-Typed Context

Measure	Female candidate				Male candidate			
	Female leader		Male leader		Female leader		Male leader	
	Failure	Success	Failure	Success	Failure	Success	Failure	Success
Perceived job-fit	5.52 (0.93)	5.69 (0.95)	5.51 (0.69)	5.47 (1.19)	5.37 (0.98)	5.18 (1.11)	4.92 (1.02)	5.36 (1.06)

Figure 1. Study 2: Model predicting evaluative outcomes for a female candidate (perceived job-fit and screening recommendation) from a female leader's performance through perceived agency of the female candidate.

Figure 2. Study 2: Unstandardized regression coefficients for model predicting evaluative

outcomes for a female candidate from a female leader's performance through perceived agency

of the female candidate. The unstandardized regression coefficient between female leader

performance and perceived job-fit of female candidate when adjusting for perceived agency of

the female candidate is shown in parenthesis.
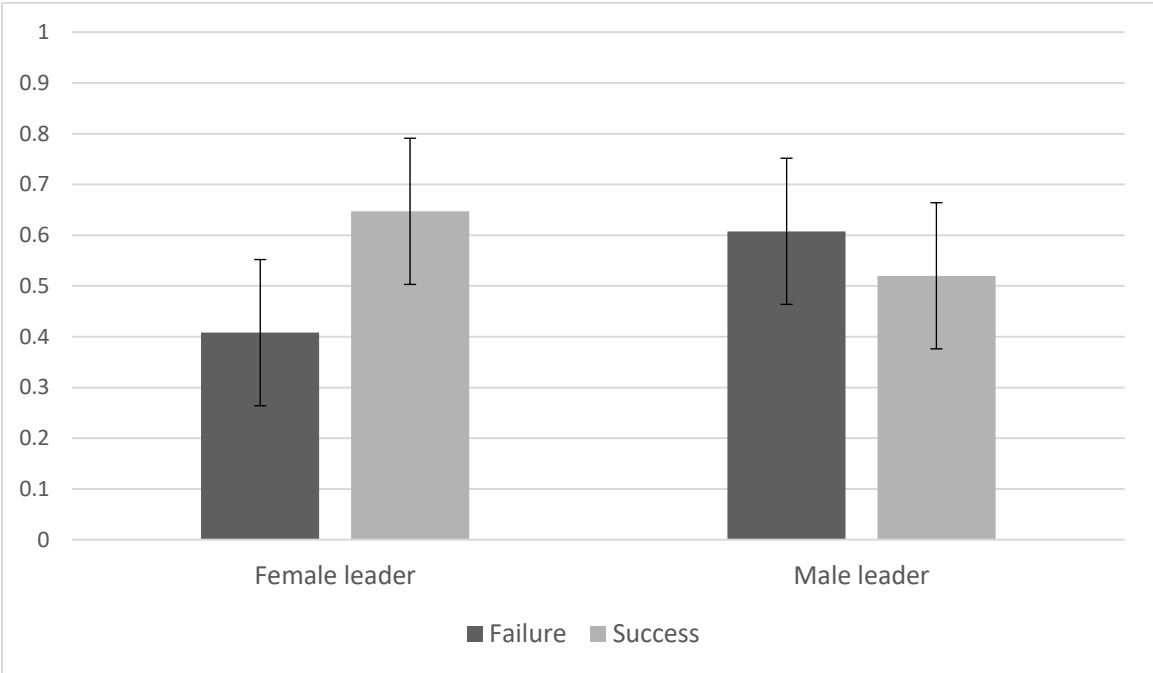
* p < .05

** p < .01

*** p < .001

Figure 3. Study 4: Likelihood of choosing a female candidate over a male candidate after exposure to a female or male leader.