



Do voters judge the performance of female and male politicians differently? Experimental evidence from the United States and Australia

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

De Geus, R. J. ORCID: <https://orcid.org/0000-0003-3053-2123>, McAndrews, J. R., Loewen, P. J. and Martin, A. (2021) Do voters judge the performance of female and male politicians differently? Experimental evidence from the United States and Australia. *Political Research Quarterly*, 74 (2). pp. 302-316. ISSN 1938-274X doi: <https://doi.org/10.1177/1065912920906193> Available at <http://centaur.reading.ac.uk/100265/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <https://doi.org/10.1177/1065912920906193>

To link to this article DOI: <http://dx.doi.org/10.1177/1065912920906193>

Publisher: Sage

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Do Voters Judge the Performance of Female and Male Politicians Differently? Experimental Evidence from the United States and Australia

Political Research Quarterly
2021, Vol. 74(2) 302–316
© 2020 University of Utah



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1065912920906193
journals.sagepub.com/home/prq



Roosmarijn A. de Geus¹ , John R. McAndrews² , Peter John Loewen²,
and Aaron Martin³

Abstract

Do gender stereotypes about agency affect how voters judge the governing performance of political executives? We explore this question using two conjoint experiments: one conducted in the United States and the other in Australia. Contrary to our expectations, we find no evidence in either experiment to suggest that female political executives (i.e., governors, premiers, and mayors) receive lower levels of credit than their male counterparts for positive governing performance. We do find evidence that female executives receive less blame than male executives for poor governing performance—but only in the U.S. case. Taken together, our findings suggest that the stereotype of male agency has only a limited effect on voters' retrospective judgments. Moreover, the results indicate that—when performance information is presented in unframed, factual terms—agentic stereotyping by voters does not, in itself, present a serious obstacle to the re-election of women in powerful executive positions.

Keywords

gender, retrospective voting, conjoint experiment, United States, Australia

Introduction

What if men are by physiology or temperament more adapted to exercise authority or to issue command?

—Tony Abbott, former prime minister of Australia¹

I just don't think she has a presidential look, and you need a presidential look.

—Donald Trump on Hillary Clinton²

The idea that women have to fit certain stereotypes; that's a weight around the ankle of every ambitious woman I've ever met. [...] This should be called out for what it is: a cultural, political, economic game that is being played to keep women in their place.

—Hillary Clinton³

Women who run for political office face a different environment than their male counterparts (Dittmar 2014; Lawless 2015). Female candidates must overcome preexisting conceptions about their suitability for office based

on stereotypes related to their gender (Bauer 2019; Fridkin, Kenney, and Woodall 2009; Koch 2000; Mo 2015; Paul and Smith 2008; Sanbonmatsu 2002; Sanbonmatsu and Dolan 2009; Sawyer 2002). While men are thought to be assertive, confident, and independent—the traditional traits of leadership—women as a social group are thought to be kind, other-serving, and compassionate (Eagly and Karau 2002; Huddy and Capelos 2002; Huddy and Terkildsen 1993b; Koenig et al. 2011). These gender stereotypes not only affect public perceptions of women's suitability for office and leadership, but also affect perceptions of their political views (Devroe and Wauters 2018; Kelley and McAllister 1983; Koch 2000), their electability (Funk, Hinojosa, and Piscopo 2017; O'Brien 2015; Thomas 2018; Thomas and Bodet

¹Oxford University, UK

²University of Toronto, ON, Canada

³The University of Melbourne, VIC, Australia

Corresponding Author:

Roosmarijn A. de Geus, Nuffield College, Gwilym Gibbon Centre for Public Policy, Oxford University, New Road, Oxford OX1 1NF, UK.
Email: Rose.degeus@nuffield.ox.ac.uk

2013), and their policy competence (Herrick and Sapieva 1998; Lawless 2004; Holman, Merolla, and Zechmeister 2011, 2016, 2017). Female candidates need to portray competence and leadership while simultaneously complying with stereotypes about their gender (Bauer 2017b; Cassese and Holman 2018; Teele, Kalla, and Rosenbluth 2018). In short, women who participate in politics face significant hurdles that their male counterparts do not.

While much is known about how gender stereotypes affect voters' judgments of female candidates and their suitability for office, comparatively little is known about how these stereotypes influence voters' judgments of women once they have *obtained* positions of significant power and responsibility. This limitation is significant as increasing numbers of women have been selected to be mayors, governors, cabinet ministers, and—in some countries—presidents and prime ministers (Barnes and O'Brien 2018; Krook and O'Brien 2012; O'Brien 2015). In other words, women do not just run for office; in an increasing number of instances, they also hold office and govern (A. C. Alexander and Jalalzai 2018; Schwindt-Bayer and Reyes-Housholder 2017). Despite an extensive literature on retrospective voting, it remains an open question whether gender stereotypes affect voters' judgments of women's performance in these positions. If voters judge the performance of women in executive positions differently from how they judge the performance of their male counterparts, then this may affect the re-election chances of women in executive roles.

In this paper, we focus on the potential implications of one gender stereotype in particular—namely, the preconception of males as agentic—and investigate whether this stereotype influences voters' retrospective judgments of executive performance. Specifically, we suggest that the gendered nature of perceptions of “agency” may lead voters to attribute lower levels of credit and blame to women executives compared to their male counterparts. If voters assume that men are more agentic—that is, “in control”—than women, then voters are expected to be more likely to attribute governing outcomes (both good and bad) to male executives than to female executives.

We test our assumptions through the use of a conjoint experiment that we conducted in both the United States and Australia. Respondents were asked to evaluate the performance of a male or female executive (mayor, governor, or state premier) in a particular governing domain (employment, crime, education, and child poverty). Each respondent saw two profiles. In the experiment, we varied the gender of the executive, the performance domain, the performance outcome (neutral, positive, or negative), and the leadership style of the executive (neutral, agentic, or communal). We further provided information on the executive's party and demographic background. This setup allows us to provide respondents with a rich

executive profile in which social desirability bias is reduced. Furthermore, the variation of multiple attributes means that our findings are averaged across a wide range of profile permutations, increasing the external validity of our results.

We find only limited evidence of agentic stereotyping in citizens' retrospective judgments. First, contrary to our expectations, we show that female and male executives receive similar levels of credit from citizens for positive governing performance. This is true in both the U.S. and Australian experiments. Second, we find that male executives receive higher levels of blame than female executives for negative performance outcomes—but only in the U.S. case. Thus, while we find some evidence consistent with the idea that voters attribute greater responsibility for performance outcomes to men than women, this evidence of agentic stereotyping is limited to the United States and only to negative performance. The upshot of these findings, we argue, is that female executives do not appear to be systematically disadvantaged when it comes to voters' retrospective judgments of their performance—at least when information about their performance is presented to voters in unframed, factual terms (i.e., in the absence of efforts by journalists or rival political actors to frame the performance information in a particular light). This should motivate female executives to emphasize positive governing performance in re-election bids and encourage political parties to select women for positions of executive power.

The paper proceeds as follows. First, we review the literature on gender stereotypes. We then set out our theoretical expectations, describe our experimental design, and present our results. We conclude by discussing our findings and identify important avenues for future research.

Gender Stereotypes and Female Candidates

Stereotypes about the male and female genders affect evaluations of political candidates (see Bauer 2019 for an extensive literature review). Men are thought to possess “agentic” character traits, such as assertiveness, confidence, and independence. Women, on the contrary, are often thought of in “communal” terms and are associated with character traits such as kindness, empathy, gentleness, and wanting to serve others (Eagly and Karau 2002; Huddy and Terkildsen 1993a, 1993b; Koenig et al. 2011). These gender-related stereotypes are significant because the conventional characteristics of politicians, especially political executives, overlap with the male gender stereotype, but contradict common stereotypes about women. The incompatibility between gender stereotypes and traditional leadership traits has been shown to pose a barrier

to the advancement of women not only in politics but also in the workplace more generally (Heilman and Eagly 2008; Heilman and Parks-Stamm 2007; Wellington, Kropf, and Gerkovich 2003).

A range of experimental studies show the negative effects of gender stereotypes on women's success in the political arena (D. Alexander and Andersen 1993; Huddy and Terkildsen 1993b, 1993a; Leeper 1991; Matland 1994; Mo 2015; Rosenwasser and Dean 1989; Rosenwasser and Seale 1988; Sapiro 1981). These studies underscore how voters associate particular character traits more with male than female candidates and make inferences about leadership and executive suitability and policy competence on the basis of these associations. In large part, these assumptions harm female candidates because women are thought to be less suitable for political office—including, in particular, executive positions related to defense and national security (Herrick and Sapieva 1998; Holman, Merolla, and Zechmeister 2011, 2016, 2017; Lawless 2004).

More recently, a growing body of research finds no clear evidence of voter bias toward female candidates. A range of studies in the United States and Canada find that when women run for office they win at the same rates as male candidates (Burrell 1992; Darcy, Welch, and Clark 1994; Fox 2006; Fox and Oxley 2003; Lawless and Pearson 2008; Sevi, Arel-Bundock, and Blais 2019; Smith and Fox 2001). Extensive research by Dolan (2004, 2010, 2014a, 2014b), Dolan and Lynch (2016), and Brooks (2013) further finds that gender stereotypes do not affect voters' judgments of politicians once partisanship and incumbency status are taken into account (also see Hayes and Lawless 2015, but see Schneider and Bos 2016). The Australian experience is similar: while females candidates have historically garnered fewer votes than men (Kelley and McAllister 1983), this gap in vote share has shrunk considerably in recent years owing to changes in social norms (A. King and Leigh 2010), potentially due to increased exposure to female politicians acting as role models (on role models, see A. C. Alexander and Jalalzai 2018; Morgan and Buice 2013).

These newer findings highlight the complicated and potentially evolving picture of how gender stereotypes influence voters' choice of political candidates. One explanation for these changes is that the strength of gender stereotypes among voters may have diminished over time (Hayes 2011). Many studies that find evidence of gender stereotypes were conducted in the 1980s and 1990s, whereas more recent studies find no such effects. Nevertheless, several recent experimental studies still report evidence of gender stereotypes (Fridkin and Kenney 2009; Mo 2015; Paul and Smith 2008; Sanbonmatsu 2002). Work by Bauer (2015a, 2015b, 2017a, 2019) suggests that gender stereotypes are present

among segments of the population, but need to be activated to be of effect. Thus, gender stereotypes are increasingly thought to affect judgments of female candidates under certain, but not all, conditions. Studies of political campaigns and media reporting indicate that the effects of gender stereotypes may be conditional on the political environment in which campaigns take place and the messages that candidates send to voters (Holman, Merolla, and Zechmeister 2011, 2016, 2017; Lawless 2004; Bauer 2015a, 2018).

Gender Stereotypes and Retrospective Judgments of Governing Performance

Much of the literature on gender stereotypes has explored how gender stereotypes affect perceptions of women's suitability for office. Comparatively little is known, however, about the potential effect of gender stereotypes on voters' judgments of women who have obtained office. Specifically, it remains unknown whether gender stereotypes affect the way in which voters attribute responsibility and blame to executives for governing outcomes. The literature that chiefly focuses on this question, the retrospective voting literature, has overlooked the possibility that gender stereotypes may affect this process. The historical dominance of men in executive office has meant that the vast majority of studies of retrospection are based on voters' judgments of men, either explicitly or implicitly. All seminal works in the field cover time periods in which incumbents were exclusively—or almost exclusively—men (Fiorina 1981; Key 1966; Kramer 1971). Perhaps most notably, every observational study of retrospective evaluations of the U.S. president to date has necessarily involved a male executive. Even today, an observational study of U.S. state governors (88% male) or U.S. mayors (78% male) would only include a small proportion of women executives.⁴ Furthermore, while some scholars have used experimental methods to explore retrospective judgments of political executives, these experiments have tended to feature vignettes of real-world male executives like Barack Obama (Newman 2013) or experimental treatments that do not specify the executive's gender (Rudolph 2006).

There is reason to believe that gender is particularly important in the executive domain, the focus of these studies of retrospective voting. Work by Schwindt-Bayer and Reyes-Housholder (2017) and A. C. Alexander and Jalalzai (2018) has shown that positions of executive power tend to be associated even more strongly with masculine character traits when compared to positions of legislative power. Whereas legislators design laws and operate as part of a collective, executives enact laws, wield power, and take decisions individually. Political

executives further tend to be more visible and receive higher levels of attention than any single member of the legislature. As a consequence, the symbolic representational effects (role model effects) are particularly strong for executive politicians when compared to legislators (A. C. Alexander and Jalalzai 2018; Morgan and Buice 2013). This leads us to expect that gender stereotypes might also affect voters' judgments of the governing performance of executives.

Theoretical Expectations

A key premise of retrospective voting is that voters evaluate executives on the basis of their performance in office. In keeping with this literature, we use the term "performance" here as a shorthand for changes in the economic and social conditions during an executive's tenure in office.⁵ In our experiments (described below), we distinguish between three types of performance outcomes: negative performance in which conditions worsened, neutral performance in which there was no change in prevailing conditions, and positive performance in which conditions improved. In line with the retrospective judgments literature, we expect that—relative to neutral performance—voters will respond favorably to executives associated with positive performance and respond negatively to executives associated with negative performance.

Given that male executives are thought of in more agentic terms than female executives (Eagly and Karau 2002; Koenig et al. 2011), we argue that voters are more likely to attribute past governing performance to male rather than female executives—that is, to treat men as more responsible than women for changes that occur in economic and social conditions during their tenure. Following from this, we expect that positive performance will have a stronger positive effect for male executives:

Hypothesis 1 (H1): Compared to neutral performance, positive governing performance has a more positive effect on public approval of male executives than female executives.

By the same logic, we expect that negative performance will have a stronger negative effect for male executives—based again on the idea that agentic stereotyping will lead voters to treat them as more responsible for the outcome. Thus,

Hypothesis 2 (H2): Compared to neutral performance, negative governing performance has a more negative effect on public approval of male executives than female executives.

H1 and H2 are both based on the same premise: if men are seen to be more agentic by voters, this should result in higher levels of attribution of responsibility for governing performance to male versus female executives.

At first glance, the expectation in H2—namely, lower levels of blame attributed to female executives compared to male executives—appears to run counter to some research that finds that female party leaders are less likely than their male counterparts to hold on to power under certain conditions (O'Brien 2015; Thomas 2018). However, the shorter tenure of female party leaders is partially shaped by the fact that women's paths to power differ from those of men. Women are more likely to be selected as party leaders when the party is thought to perform badly already or when it is facing particularly strong competition—a phenomenon often referred to as the "glass cliff" (Bruckmüller and Branscombe 2010; Funk, Hinojosa, and Piscopo 2017; O'Brien 2015; but see Thomas 2018). Moreover, the greater likelihood of female leaders leaving office could, in theory, be precipitated by pressure within their party or criticism in the press. In other words, a comparatively shorter tenure of female party leaders does not, in itself, provide direct evidence about how voters judge their conduct in office. Instead, how voters evaluate the performance of female executives remains an open question and will be tested directly in H2.

Data and Method

We use a conjoint experiment to test our hypotheses in two countries, the United States and Australia. The vast majority of studies on gender stereotypes focus on the United States (but see Herrick and Sapieva 1998; Mo 2015; Ono 2018). By running similar experiments in two different countries, we can begin to assess the generalizability of the findings. We include Australia in the present study as a broadly similar case to the United States. Like the United States, the Australian system is leader-centered (Bean 1993; Bean and Mughan 1989; Goot 2008; Kefford 2013; McAllister 2011), English-speaking, British in its colonial origins, and typically a two-party system. What is more, data from the World Value Survey suggest that attitudes toward the role of men and women in politics and society are relatively similar in the two countries.⁶

Despite these similarities, Australia has more experience with women in elected office than the United States. Levels of female representation in the legislature are somewhat higher in Australia currently than in the United States (32% in Australia vs. 20.6% in the U.S. House and Senate combined).⁷ With the exception of South Australia, every Australian state and territory has had a female head of government (premier or chief minister, comparable to

the position of a state governor in the United States). Australia has also had a female chief executive at the federal level, Prime Minister Julia Gillard. Having said this, however, access to high-level positions remains restricted in both countries: currently, women govern in only two out of eight states and territories in Australia; in the United States, there are women governors in only six of the fifty states.⁸ In short, despite some differences, we believe that the Australian case offers an important opportunity for cross-national replication.

To evaluate H1 and H2, we use a conjoint experimental design (Hainmueller, Hopkins, and Yamamoto 2014). Conjoint experiments perform well in terms of external validity (Hainmueller, Hangartner, and Yamamoto 2015) and have previously been used to investigate gender-based biases (Eggers, Vivyan, and Wagner 2018; Teele, Kalla, and Rosenbluth 2018). We preregistered our design, hypotheses, and analysis plan.⁹ We used a commercial sample provider (Qualtrics) in both the United States and Australia to construct a sample that is broadly representative of the national adult population in terms of age, gender, and region (in the U.S. case, race). Sample details are provided in the supplementary material.¹⁰ The U.S. experiment was fielded in June 2018 to a sample of 525 citizens. The Australian experiment was conducted in October 2018 and involved a sample of 607 citizens.

The design and text of the two experiments were very similar, differing only with respect to a few small changes needed to adapt the information to the particular country's context. All respondents saw two profiles of a political executive. In Australia, respondents saw profiles of two hypothetical state premiers; in the United States, respondents saw one profile of a hypothetical city mayor and one profile of a hypothetical state governor. Respondents read and evaluated each executive profile separately (see supplementary materials for the stimuli). In each profile, we randomly assigned several personal characteristics of the executive: including most notably gender, but also party, tenure in office, prior profession, and leadership style. In addition, we varied two aspects of the executive's performance in office: the outcome (i.e., negative, neutral, or positive) and the policy domain in question (i.e., crime, child poverty, unemployment, and education).¹¹ Table 1 reports all of the attributes—and their associated categories—that were randomized in the experiments.

Following each vignette, we asked respondents how much they approved of how the executive was handling their job. The exact question wording varied only with respect to the executive's title and gender: "On a scale of 0–100, how much do you approve or disapprove of the way in which the governor/mayor/premier is handling her or his job?" We further asked respondents how likely they would be to vote for the executive and asked respondents

to indicate how well they thought a range of character traits (e.g., strong leader, competent, honest) described the executives.

The use of an experimental design is motivated by the fact that the actual number of women in real-world executive positions remains small. This limited population of female executives makes inferences about the effects of performance less feasible in an observational setting. What is more, the small number of female political executives potentially limits the generalizability of observational findings because voters' opinions might be closely related to a particular female executive. By contrast, an experimental design—and the conjoint design used here in particular—has several advantages. First, we can present respondents with hypothetical female executives without being constrained by the real-world underrepresentation of women in these offices. The hypothetical nature of the stimulus simultaneously helps us avoid a scenario in which respondents make associations with any particular female executive.

Second, in a conjoint design, we can provide respondents with a range of attributes on which to base their judgments. This potentially reduces social desirability bias that may have occurred had respondents had only one or two pieces of information on which to base their judgments. Third, various authors have cautioned that the use of experimental designs may inflate the effect of gender stereotypes because experimental respondents operate in a low-information environment (Andersen and Ditonto 2018; Dolan 2014a; Hayes 2011; D. C. King and Matland 2003; Koch 1999). By increasing the information available to respondents through a conjoint design, we enhance the realism of the experiment—potentially avoiding the false positives that may be more likely in a low-information experimental setting.

Fourth, by averaging our estimates of treatment effects across a wide range of profiles, we increase the generalizability of our estimates—compared again to the more traditional experimental approach in which many of these profile characteristics would necessarily be fixed by design. In developing our experimental design, however, we opted against the common paired conjoint design in which respondents are simultaneously presented with two profiles (usually side-by-side on a single page) and forced to choose between the two. Instead, we adopted the single-profile conjoint design so as to focus respondents' judgments on evaluating the executive, rather than seeking to simulate an election contest.

All attributes in the conjoint were randomly and independently assigned. We report average marginal component effects (AMCEs) to estimate the effects of the randomized attributes and report average component interaction effects (ACIEs) to estimate how these component effects depend on the gender of the executive (Hainmueller

Table 1. Attributes Included in Conjoint Experiment.

Attribute	Categories
Age	40 45 50 55
Gender	Male Female
Party	Democrat (United States) Republican (United States) Liberal (Australia) Labor (Australia)
Incumbency	Has been governor/mayor/state premier for X years: 1 3 5
Prior profession	Lawyer Doctor Business owner Engineer
Performance domain	Student test scores (Australia and Governor vignette United States) Employment rate (Australia and Governor vignette United States) Child poverty rate (Mayor vignette United States) Crime rates (Mayor vignette United States)
Performance outcome	Gone down Stayed the same Gone up
Leadership style	He or she makes important decisions on his or her own He or she likes to work with others when making important policy decisions No information

and Hopkins 2015; Hainmueller, Hopkins, and Yamamoto 2014). All models are estimated using ordinary least squares (OLS) regression. We adjust for the fact that each respondent reviewed two separate executive profiles—and thus contributed two observations to the analysis—by clustering the standard errors by participant. In the U.S. experiment, we pool the responses from the mayor and governor profiles to increase our statistical power.¹² We also ran our main analyses for the mayor and governor profiles separately and found no significant differences. These analyses are provided in the supplementary materials.¹³

Results

Figures 1 and 2 show the effect of each conjoint attribute on job approval. Figure 1 presents the results from the U.S. experiment; Figure 2 reports the results from the Australian experiment. Both figures are organized in the same manner: the left-most panel presents the results for female executives only; the middle panels show the results for male executives only; and the right-most panel shows the differences in the component effects between the two.

The estimates for the right-most panel were obtained by pooling the female and male executive profiles and then estimating a model in which all components were interacted with executive gender. Thus, if the confidence interval of an estimate in the right-most panel includes zero, this indicates that male and female executives were evaluated similarly on the basis of the attribute in question. All model estimates are available in the supplementary material in the form of regression tables.

We begin by examining the effects of positive and negative governing performance on job approval. The left-most and middle panels in both Figures 1 and 2 show that the performance conditions exerted effects in the expected directions; consistent with the observational work on retrospective judgments, we find that a worsening situation has a negative effect on approval ratings compared to the neutral performance condition. Similarly, as one would expect, an improving situation has a positive effect: voters react more favorably to an improving situation than to the neutral performance condition. These findings are true for both female and male executives in both the U.S. and Australian experiments.

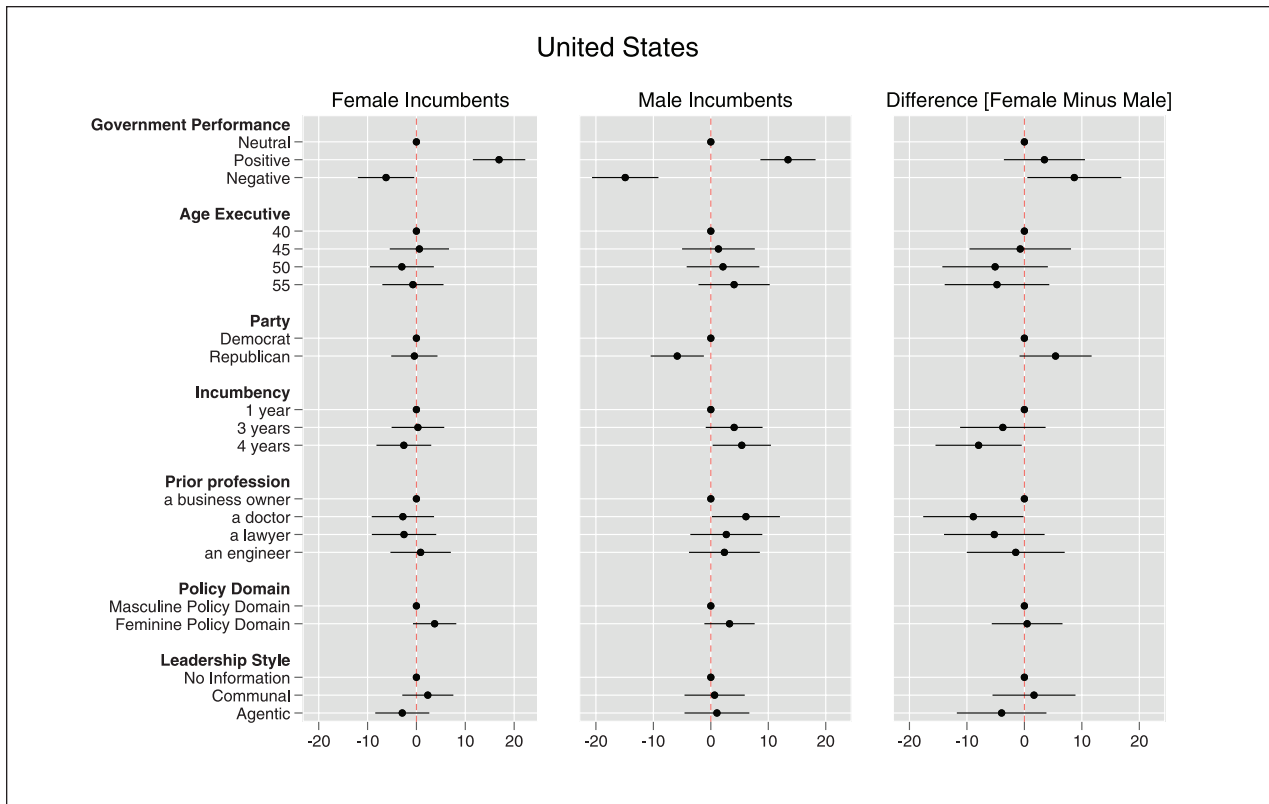


Figure 1. Average marginal component effects on job approval and their interaction with executive gender (U.S. study).

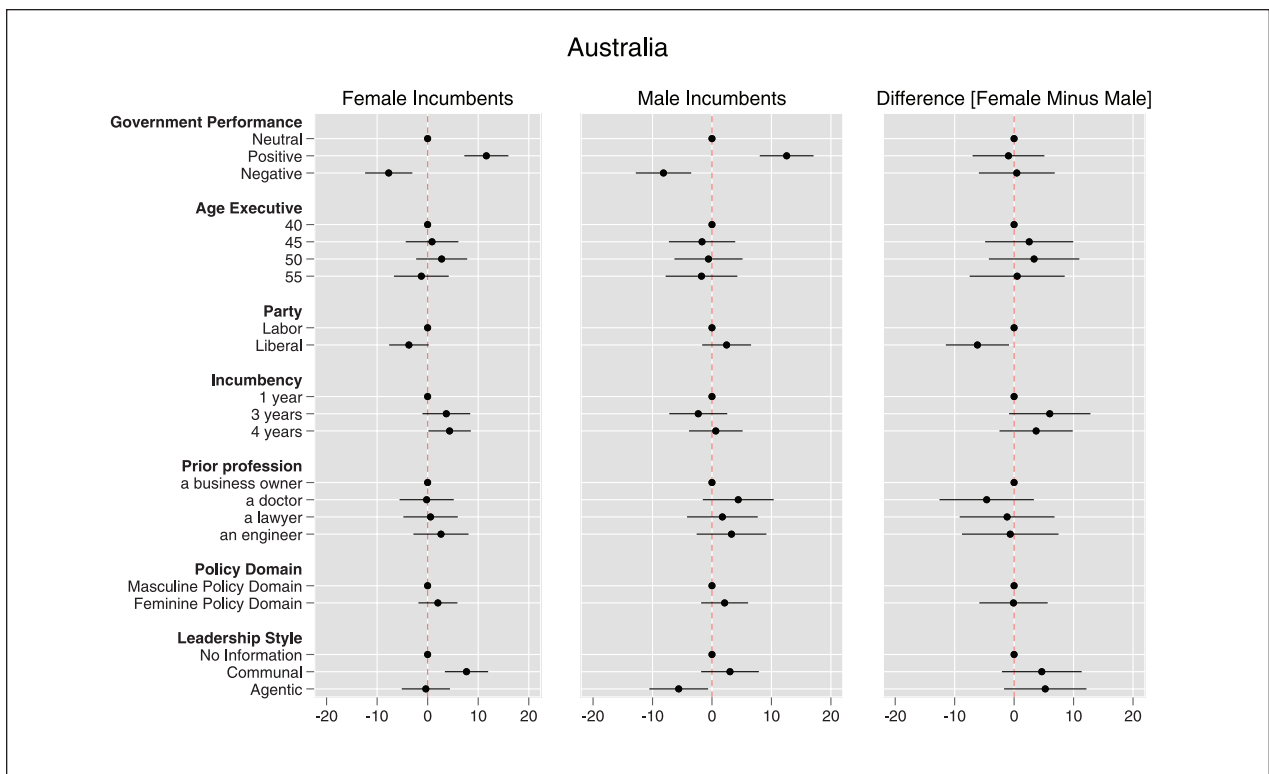


Figure 2. Average marginal component effects on job approval and their interaction with executive gender (Australia study).

We draw two important inferences from these preliminary findings. First, our experimental manipulation of the performance conditions had the desired effects: respondents attribute credit in light of good performance and blame in light of poor performance. This serves a kind of manipulation check on our key experimental treatment. Second, and more substantively, the fact that we find strong evidence of credit and reward for female executives in particular is noteworthy. This finding suggests that agentic stereotyping is not so strong as to wipe out retrospective judgments for female executives: women executives are rewarded for good performance and punished for bad performance.

To estimate the gender *differences* in the effects of performance—that is, the focus of H1 and H2—we turn to the right-most panels, beginning with the U.S. experiment in Figure 1. Here, we find no statistically significant gender difference in the effect of positive performance: male executives did not, as we hypothesized in H1, receive significantly greater credit than female executives for improving conditions. However, we do find a statistically significant difference in the effect of negative performance: the negative effect of worsening conditions on job approval is smaller for female executives than male executives. This is shown in the right-most panel of Figure 1 by the positive difference between female and male executives with respect to the effect of negative performance. In other words, the U.S. results suggest—consistent with our expectation in H2—that male executives receive stronger punishment. To put this in substantive terms, we use the estimates of the fully interacted model to generate predicted levels of job approval for male and female executives under the three performance conditions. For male executives, the difference in approval ratings between a neutral performance (e.g., crime rates stayed the same) and a negative performance (e.g., crime rates increased) is 14 points: a drop from 58 to 44.¹⁴ For female executives, the drop in approval is 6 points—from 57 to 51. This 8-point difference between male and female executives is significant at $p < .042$.

Next, we consider gender differences in the Australian experiment: the right-most panel of Figure 2. Do the U.S. results replicate? As in the U.S. experiment, we do not find that executive gender moderates the effect of positive performance on job approval: Australians—like their American counterparts—credited female and male executives similarly. However, we do not replicate the U.S. finding with respect to the gender difference in the effect of negative performance. In the Australian case, citizens punished the state premiers for worsening conditions—but this effect did not evidently differ between female and male executives.

Our experiments are designed to be able to detect reasonably large gender-based differences in performance effects (i.e., more than 10 points on a scale of 0–100). We lack conventional levels of statistical power to detect differences below this effect size. Thus, while we are confident in rejecting the possibility of large gender differences in the effect of positive performance in both the U.S. and Australian experiments, as well as a large gender difference in the effect of negative performance in the Australian case, we cannot entirely rule out of the possibility of small—and thus hard-to-detect—gender differences in retrospective voter judgments.¹⁵

As a robustness check, we estimated the effect of governing performance on a second dependent variable: namely, respondents' self-reported likelihood of voting for the executive. We find the same results as those reported above for job approval. In the United States, we find no significant gender differences in how citizens rewarded positive performance. However, we find again that poor performance has a more negative effect on the likelihood of voting for the male executive than the female executive. Once again, there were no gender-related differences in the Australian experiment with respect to the effects of either positive or negative performance. These estimates are available in the supplementary material.

As a final robustness check, we explore the effect of governing performance on perceived leadership skills of the executives. In addition to asking about job approval and hypothetical vote choice, we asked survey respondents whether the term “strong leader” was applicable to the executive about which they had read. The answer categories were either “yes” or “no.” Figures 3 and 4 provide the AMCEs of governing performance on the likelihood that respondents stated that the executive was a strong leader. We see that perceptions of strong leadership of both male and female executives are favorably affected by positive performance in both the United States and Australia, again validating the effectiveness of the experimental treatment.

In terms of gender differences, we find no significant difference in the effect of positive or negative performance on attribution of leadership skills in either country. We thus do not replicate the finding that male incumbents receive higher levels of blame in light of negative performance in the United States. In line with our main findings presented above, the effect sizes of gender and performance on evaluations of leadership are small, and hence we cannot distinguish between a null effect and a potentially small (smaller than 5 points), but undetectable, effect of gender stereotyping. We therefore conclude that there is no evidence of large (10 points or more) gender-related differences in leadership evaluations in light of governing performance.

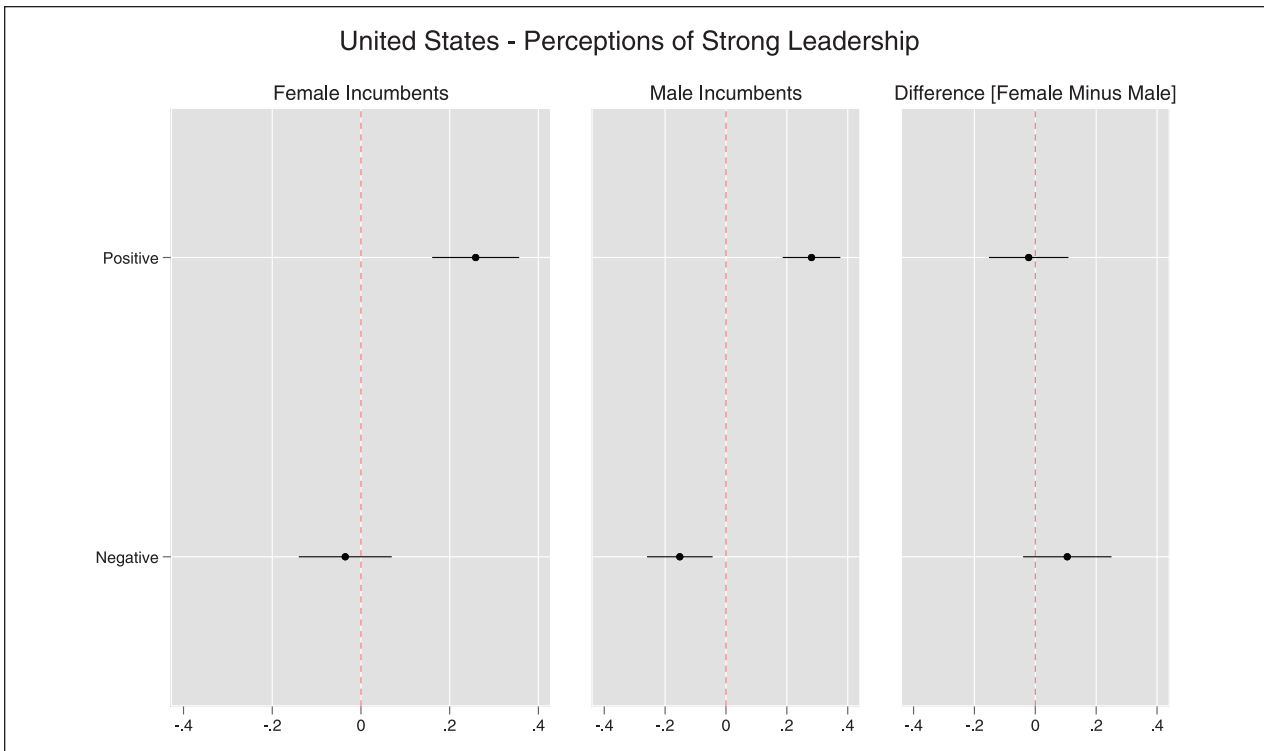


Figure 3. Average marginal component effects of governing performance on perceptions of strong leadership and their interaction with executive gender (U.S. study).

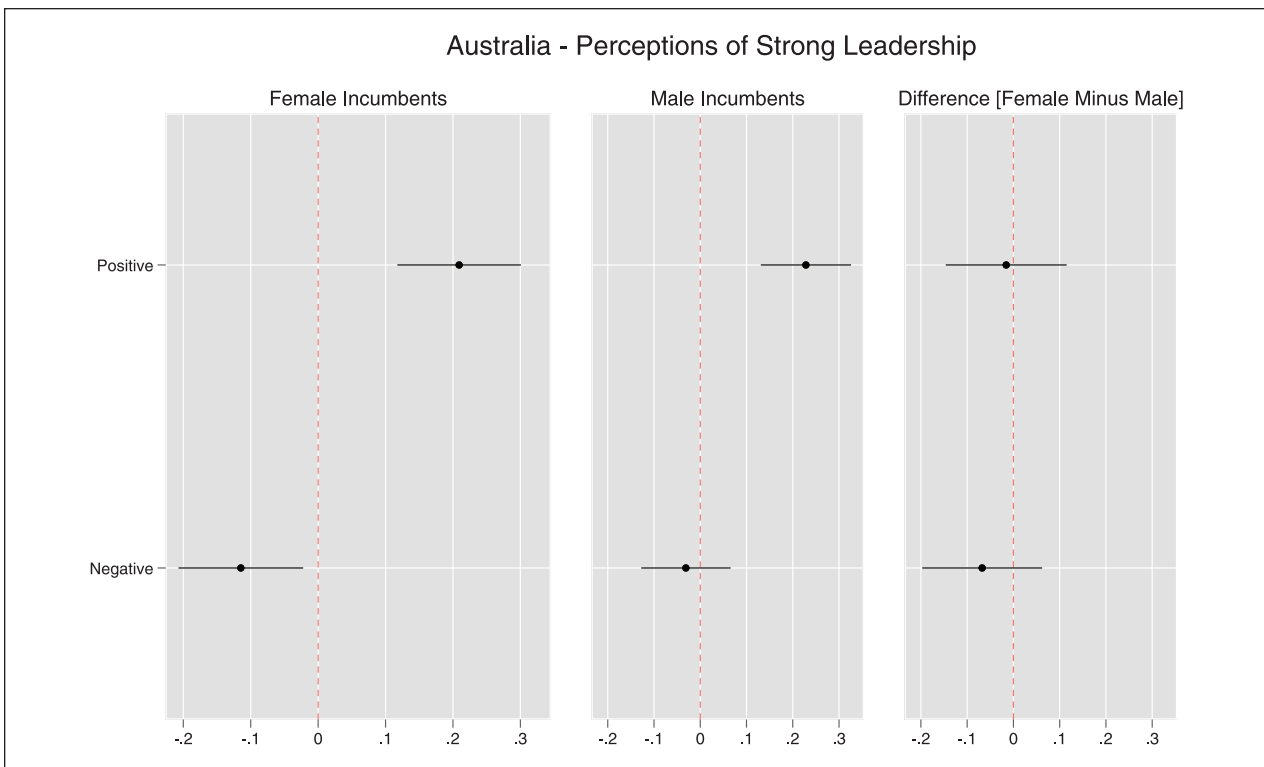


Figure 4. Average marginal component effects of governing performance on perceptions of strong leadership and their interaction with executive gender (Australia study).

Discussion

Studies of female political candidates indicate that gender stereotypes affect the experiences of women who run for office. Little is known, however, about the potential effects of gender-related stereotypes on voters' attribution of blame and reward for governing performance. We find only limited evidence that citizens evaluate the performance of female executives by different standards than those of their male counterparts. Specifically, we find no evidence of gender-related differences in either Australia or the United States when it comes to the attribution of credit for positive governing performance: improving conditions have broadly similar effects on voter judgments regardless of the executive gender. We do find that male leaders are punished more than female leaders for negative performance, a finding that is consistent with our theoretical expectations about agentic stereotyping, but this evidence is limited to the U.S. experiment only.

Our findings provide good news for female executives; they receive similar levels of credit for positive performance in office in both our studies. As such, this should open up opportunities for female executives to actively campaign on their performance record when seeking re-election. We find that voters are as responsive to positive performance signals under female compared to male executives, and female executives should thus be able to leverage a positive performance record in a similar fashion to their male counterparts. This suggests that female executives should be encouraged to make positive governing performance a key component of their re-election campaigns.

The fact that we find evidence in the United States to suggest that female executives are punished *less* severely for negative performance is suggestive of agentic stereotyping, but the fact that we find this gender gap only under conditions of negative performance is puzzling. One possible explanation for this might lie with a negativity bias. Research in the fields of psychology, politics, and communication suggests that people pay closer attention to negative information than positive information (Baumeister et al. 2001; Hibbing, Smith, and Alford 2014; Rozin and Royzman 2001; Soroka 2014; Soroka and McAdams 2015). It could be that the presentation of negative information in our experiments (e.g., crime rates have gone up, student test scores have gone down) activated gender and leadership stereotypes in a way that positive information did not (see also Hayes, Lawless, and Baitinger 2014). Further experimental work is needed to assess this explanation.

The fact that female executives in the U.S. experiment received lower levels of blame compared to male executives may further reflect a gender advantage for female

politicians. Barnes and Beaulieu (2014), for instance, have shown that female politicians in general are perceived to be more honest and less corrupt than their male counterparts. Work by Bruckmüller and Branscombe (2010) further suggests that women are seen as more suitable to lead organizations in times of crisis—broadly analogous to our negative performance condition—because they are perceived to have better interpersonal skills. These are instances in which gender stereotyping might favor women.

Yet, the fact that we find evidence of gender stereotypes in light of negative performance in the case of the United States but not Australia emphasizes the importance of replication of experiments across various contexts, as well as the importance of comparative work and the development of country-specific and comparative hypotheses. A potential explanation for the cross-national differences found in our study might be that U.S. respondents are more willing to give female executives the benefit of the doubt for negative performance because of the continued outsider status of women in U.S. politics. Research by Morgan and Buice (2013) in Latin America suggests that female politicians enjoy higher levels of trust among voters when they are seen as “outsiders” or novices to politics. Yet, the advantages associated with the “outsider” effect disappear when levels of female representation increase and female politicians lose their “novelty” status (Morgan and Buice 2013). This may help to explain why we do not find evidence of a gender difference with respect to the effect of negative performance in Australia—where voters are more familiar with female leaders.

Existing research on women's ascension to positions of executive leadership suggests that parties have a tendency to select women for top-level positions mostly under suboptimal conditions: when competition with other parties is fierce, electoral prospects are weak, and the economy is in decline (Funk, Hinojosa, and Piscopo 2017; O'Brien 2015). To some extent, this might be an effective strategy as our results suggest that female executives receive lower levels of blame in light of negative performance. However, our Australian results suggest that this particular advantage might disappear once women are no longer considered a novelty or outsider candidate. Another motivation for parties to select women for top-level positions under suboptimal conditions may be the fact that women's election prospects are often considered to be lower than those of men. Female politicians are then used as “sacrificial lambs,” only to make way for a male candidate to take their place once the party's prospects have improved (Thomas and Bodet 2013). Such a strategy seems based on assumptions about voter hostility to female politicians or beliefs that women politicians are less likely to reap the benefits of a positive record in governance. Our findings suggest that such beliefs are ill-founded.

It is perhaps surprising that we find so few differences in how citizens evaluate the performance of male and female executives. The high visibility of executive positions and the association of traditional masculine traits with this role (A. C. Alexander and Jalalzai 2018; Morgan and Buice 2013) might lead us to expect that judgments of political executives are especially prone to gender stereotyping. From a different point of view, however, we might expect women who have reached executive positions to be less affected by gender stereotyping. After all, these women have likely cleared many hurdles already in their electoral career and as a consequence may be less susceptible to gender stereotyping. The rather exceptional nature of women who reach executive office is reflected in the fact that it remains more difficult for women to obtain these positions compared to legislative positions (Hinojosa and Franceschet 2012). Perhaps therefore it is less surprising that once women reach positions of executive power, they are judged by voters as “leaders, not ladies” (Brooks 2013).

Our findings with respect to retrospective judgments are in line with an increasing set of studies that find no strong evidence of gender stereotyping by voters (Brooks 2013; Burrell 1992; Darcy, Welch, and Clark 1994; Dolan 2004, 2010, 2014a, 2014b; Dolan and Lynch 2016; Fox 2006; Fox and Oxley 2003; Hayes and Lawless 2015; Hayes, Lawless, and Baitinger 2014; Lawless and Pearson 2008; Sevi, Arel-Bundock, and Blais 2019; Smith and Fox 2001). In keeping with this literature, our study suggests that the advancement of women in politics might not be significantly hindered by the electorate. Rather, insights from studies of party selection mechanisms (Funk, Hinojosa, and Piscopo 2017) and the role of electoral institutions (Hinojosa and Franceschet 2012) suggest that a lack of active recruitment and promotion of women by political parties, as well as structural features of electoral systems, may present stronger barriers for the advancement of women in politics. In Australia, the major parties have adopted different strategies to improve female participation, and several of these have been proven to be effective. Yet, while the public are generally supportive of greater female representation, opinion is divided as to the best mechanism to achieve this, and men are more supportive of the status quo (Beauregard 2018).

Our conclusion in this study does not negate the fact that some political leaders—such as Abbott and Trump, quoted at the outset of this paper—make public statements that reinforce gender stereotypes about leadership. What is more, prior research shows that such stereotypes can be transmitted through gendered portrayals of female leaders in the media (Trimble 2017). Our findings do not suggest that these stereotypes are an unimportant feature of politics. Rather, they suggest that there are potential bounds on the distorting role of these stereotypes. When

citizens are presented with factual information about prevailing economic and social conditions—the kinds of information that often inform retrospective judgments of political leaders—they do not use this information in ways that the traditional stereotype of male leadership would lead us to expect.

We deliberately designed our survey experiments to present information about prevailing economic and social conditions in an unframed, factual manner. This is both an advantage and a limitation. The nature of the presentation allows us to conclude that it is not the information as such that drives agency stereotyping: when confronted with unframed performance information, voters in our experiment tended to judge male and female executives similarly. Yet, performance information is not always presented in a strictly matter-of-fact manner in the real world; rather, it can be framed by journalists in news reports and is subject to rhetoric from political rivals—as illustrated in Trump’s description of Clinton quoted at the start. As studies of political candidates show, the effects of gender stereotypes are conditional on the political environment (Bauer 2015a). Thus, it may be the case that the framing of performance information (e.g., through political campaigns or media reporting) may serve to activate agentic stereotyping related to gender (Cassese and Holman 2018). Indeed, it is easy to see how gender stereotypes could be used by political opponents to frame negative governing performance in gendered terms—priming stereotypes of women as weak leaders who lack agency. Future work should thus extend the present line of research to explore how gendered framing of performance information may influence voters’ retrospective judgments of female political leaders.

Finally, in this study, we focus specifically on the gender stereotype as it relates to agency and its link with the retrospective attribution of credit and blame. However, many other gender stereotypes exist and may affect retrospective judgments of governing performance of male and female executives.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Roosmarijn A. de Geus  <https://orcid.org/0000-0003-3053-2123>

John R. McAndrews  <https://orcid.org/0000-0002-3085-9479>

Notes

1. As quoted by Julia Gillard in her 2012 speech about misogyny in Australian politics, a transcript of which is here: <https://www.smh.com.au/politics/federal/transcript-of-julia-gillards-speech-20121010-27c36.html> (accessed August 21, 2019).
2. See Parker (2016).
3. See Galanes (2017).
4. Data taken from the Centre for American Women and Politics (2018).
5. We do not mean to imply that these changes *should* be attributed to the executive or that they are objectively indicative of their conduct in office. Instead, we stipulate only that voters often *do* interpret these changes as indicators of an executive's performance—rewarding and punishing accordingly.
6. Data from the most recent wave of the World Values Survey, for example, show that 21.9 percent of Australians agree or agree strongly with the statement that “on the whole men make better political leaders than women”—compared to 19.4 percent of respondents in the United States. Attitudes toward gender roles in society more broadly are also relatively similar in the two countries. Again, per the World Values Survey, 13.6 percent of Australians agree or agree strongly that “men make better business executives than women,” compared to 11.7 percent of Americans. Finally, 21.1 percent of Australians agree or agree strongly that “when a mother works, the children suffer,” compared to 24.9 percent of Americans. World Value Survey Wave 6, 2010–2014 (Inglehart et al. 2014). Available at: <http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>.
7. Parliament of Australia https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/FlagPost/2016/August/The_gender_composition_of_the_45th_parliament. Center for American Women and Politics <https://www.cawp.rutgers.edu/women-us-congress-2018>
8. As per April 2019.
9. We also registered hypotheses relating to respondent characteristics, namely, respondent gender and partisanship. However, in this paper, we focus only on the attributes that were randomized in the vignettes.
10. Given the representative nature of both our U.S. and Australian samples (see our supplementary materials for a comparison between our samples and the U.S. and Australian census), we have no reason to believe that the prevalence of gender stereotyping—extensively documented elsewhere in the literature (see Bauer 2019)—would differ between our sample and the general population. Supplementary material is available on journal website.
11. Specifically, for the U.S. vignettes, the mayoral profile either reported child poverty rates or crime rates, whereas the gubernatorial profile included information about either unemployment or student test scores. For the Australian vignettes, we provided information about either student test scores or unemployment rates. The aim was to provide performance information in a policy domain that was plausibly related to the executive's jurisdiction.
12. We conducted power analyses of our ability to detect the interaction effects set out in H1 and H2 (provided in the supplementary materials). The experiments are well powered to detect effect sizes in the range of 10 points (on a 100-point approval scale) at 80 percent power. For effect sizes in the 8- to 10-point range, our power is 60 percent. We have reduced power to effect sizes below 8 points, with power lower than 60 percent. We have no power to detect very small effect sizes; specifically our power is less than 20 percent to detect effects of 5 points or smaller.
13. In our registered preanalysis plan, we also set out a secondary set of expectations, namely we expected that leadership style might moderate the gender gap in the effects of performance. Specifically, as men are stereotypically seen to be more agentic than women, we expected that a woman with an agentic leadership style might be able to close some of the expected gender gap in retrospective judgments. We tested whether leadership style moderated the gender gap and found that it did not. We chose not to include this secondary set of expectations for space considerations.
14. Here, all other variables are kept at their assigned values.
15. Note that this effect is substantive at 8.7 points, and our power to detect effects of this size in the United States is ~60 percent.

Supplemental Material

Supplemental materials and replication materials for this article are available with the manuscript on the *Political Research Quarterly* (PRQ) website.

References

- Alexander, Amy C., and Farida Jalalzai. 2018. “Symbolic Empowerment and Female Heads of States and Government: A Global, Multilevel Analysis.” *Politics, Groups, and Identities*. Published electronically February 25. doi:10.1080/21565503.2018.1441034.
- Alexander, Deborah, and Kristi Andersen. 1993. “Gender as a Factor in the Attribution of Leadership Traits.” *Political Research Quarterly* 46 (3): 527–45.
- Andersen, David J., and Tessa Ditonto. 2018. “Information and Its Presentation: Treatment Effects in Low-Information vs. High-Information Experiments.” *Political Analysis* 26 (4): 379–98. doi:10.1017/pan.2018.21.
- Barnes, Tiffany D., and Emily Beaulieu. 2014. “Gender Stereotypes and Corruption: How Candidates Affect Perceptions of Election Fraud.” *Politics & Gender* 10 (3): 365–91. doi:10.1017/S1743923X14000221.
- Barnes, Tiffany D., and Diana Z. O'Brien. 2018. “Defending the Realm: The Appointment of Female Defense Ministers Worldwide.” *American Journal of Political Science* 62 (2): 355–68. doi:10.1111/ajps.12337.
- Bauer, Nichole M. 2015a. “Emotional, Sensitive, and Unfit for Office? Gender Stereotype Activation and Support Female Candidates.” *Political Psychology* 36 (6): 691–708. doi:10.1111/pops.12186.
- Bauer, Nichole M. 2015b. “Who Stereotypes Female Candidates? Identifying Individual Differences in Feminine

- Stereotype Reliance." *Politics, Groups, and Identities* 3 (1): 94–110. doi:10.1080/21565503.2014.992794.
- Bauer, Nichole M. 2017a. "Gender Stereotypes and Voter Evaluations of Female Candidates." In *The Political Psychology of Women in U.S. Politics*. Routledge Studies in Political Psychology, edited by Angela L. Bos and Monica C. Schneider, 167–84. New York: Routledge.
- Bauer, Nichole M. 2017b. "The Effects of Counterstereotypic Gender Strategies on Candidate Evaluations." *Political Psychology* 38 (2): 279–95. doi:10.1111/pops.12351.
- Bauer, Nichole M. 2018. "Untangling the Relationship between Partisanship, Gender Stereotypes, and Support for Female Candidates." *Journal of Women, Politics & Policy* 39 (1): 1–25. doi:10.1080/1554477X.2016.1268875.
- Bauer, Nichole M. 2019. "Gender Stereotyping in Political Decision Making." Oxford Research Encyclopedia of Politics. <http://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-772>.
- Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. "Bad Is Stronger than Good." *Review of General Psychology* 5 (4): 323–70.
- Bean, Clive. 1993. "The Electoral Influence of Party Leader Images in Australia and New Zealand." *Comparative Political Studies* 26:111–32.
- Bean, Clive, and Anthony Mughan. 1989. "Leadership Effects in Parliamentary Elections in Australia and Britain." *American Political Science Review* 83:1165–79.
- Beauregard, Katrine. 2018. "Partisanship and the Gender Gap: Support for Gender Quotas in Australia." *Australian Journal of Political Science* 53 (3): 290–319. doi:10.1080/10361146.2018.1449802.
- Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes Do not Harm Women Candidates*. Princeton: Princeton University Press.
- Bruckmüller, S., and Nyla R. Branscombe. 2010. "The Glass Cliff: When and Why Women Are Selected as Leaders in Crisis Contexts." *British Journal of Social Psychology* 49 (3): 433–51. doi:10.1348/014466609X466594.
- Burrell, Barbara. 1992. "Women Candidates in Open-Seat Primaries for the U.S. House: 1968–1990." *Legislative Studies Quarterly* 17 (4): 493–508. doi:10.2307/439863.
- Cassese, Erin C., and Mirya R. Holman. 2018. "Party and Gender Stereotypes in Campaign Attacks." *Political Behavior* 40 (3): 785–807. doi:10.1007/s11109-017-9423-7.
- Centre for American Women and Politics. 2018. "Women in Elective Office 2018." Eagleton Institute of Politics, Rutgers University. <http://www.cawp.rutgers.edu/women-elective-office-2018>.
- Darcy, R., Susan Welch, and Janet Clark. 1994. *Women, Elections and Representation*. Women & Politics. 2nd ed., Vol. 1. Lincoln: University of Nebraska Press.
- Devroe, Robin, and Bram Wauters. 2018. "Political Gender Stereotypes in a List-PR System with a High Share of Women MPs: Competent Men versus Leftist Women?" *Political Research Quarterly* 71 (4): 788–800. doi:10.1177/1065912918761009.
- Dittmar, Kelly. 2014. *Navigating Gendered Terrain: Stereotypes and Strategy in Political Campaigns*. Philadelphia: Temple University Press.
- Dolan, Kathleen. 2004. *Voting for Women: How the Public Evaluates Women Candidates*. Boulder: Westview Press.
- Dolan, Kathleen. 2010. "The Impact of Gender Stereotyped Evaluations on Support for Women Candidates." *Political Behavior* 32 (1): 69–88. doi:10.1007/s11109-009-9090-4.
- Dolan, Kathleen. 2014a. "Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates: What Really Matters?" *Political Research Quarterly* 67 (1): 96–107.
- Dolan, Kathleen. 2014b. *When Does Gender Matter? Women Candidates and Gender Stereotypes in American Elections*. New York: Oxford University Press.
- Dolan, Kathleen, and Timothy Lynch. 2016. "The Impact of Gender Stereotypes on Voting for Women Candidates by Level and Type of Office." *Politics & Gender* 12 (3): 573–95. doi:10.1017/S1743923X16000246.
- Eagly, Alice H., and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109 (3): 573–98. doi:10.1037/0033-295X.109.3.573.
- Eggers, Andrew C., Nick Vivyan, and Markus Wagner. 2018. "Corruption, Accountability, and Gender: Do Female Politicians Face Higher Standards in Public Life?" *The Journal of Politics* 80 (1): 321–26. doi:10.1086/694649.
- Fiorina, Morris. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.
- Fox, Richard L. 2006. "Congressional Elections: Where Are We on the Road to Gender Parity." In *Gender and Elections: Shaping the Future of American Politics*, edited by Susan J. Carroll and Richard L. Fox, 187–210. New York: Cambridge University Press.
- Fox, Richard L., and Zoe M. Oxley. 2003. "Gender Stereotyping in State Executive Elections: Candidate Selection and Success." *The Journal of Politics* 65 (3): 833–50. doi:10.1111/1468-2508.00214.
- Fridkin, Kim L., and Patrick J. Kenney. 2009. "The Role of Gender Stereotypes in U.S. Senate Campaigns." *Politics & Gender* 5 (3): 301–24. doi:10.1017/S1743923X09990158.
- Fridkin, Kim L., Patrick J. Kenney, and Gina Serignese Woodall. 2009. "Bad for Men, Better for Women: The Impact of Stereotypes during Negative Campaigns." *Political Behavior* 31 (1): 53–77. doi:10.1007/s11109-008-9065-x.
- Funk, Kendall D., Magda Hinojosa, and Jennifer M. Piscopo. 2017. "Still Left Behind: Gender, Political Parties, and Latin America's Pink Tide." *Social Politics* 24 (4): 399–424.
- Galanes, Philip. 2017. "Hillary Clinton and America Ferrera on Pain and Progress (and Hiking)." *The New York Times*, September 16. https://www.nytimes.com/2017/09/16/style/hillary-clinton-america-ferrera-table-for-three.html?_r=0.
- Goot, Murray. 2008. "Is the News on the Internet Different? Leaders, Frontbenchers and Other Candidates in the 2007 Australian Election." *Australian Journal of Political Science* 43 (1): 99–110. doi:10.1080/10361140701851939.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." *Proceedings of the National Academy of Sciences of the*

- United States of America* 112 (8): 2395–400. doi:10.1073/pnas.1416587112.
- Hainmueller, Jens, and Daniel J. Hopkins. 2015. “The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants.” *American Journal of Political Science* 59 (3): 529–48. doi:10.1111/ajps.12138.
- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis* 22 (1): 1–30. doi:10.1093/pan/mpt024.
- Hayes, Danny. 2011. “When Gender and Party Collide: Stereotyping in Candidate Trait Attribution.” *Politics & Gender* 7 (2): 133–65. doi:10.1017/S1743923X11000055.
- Hayes, Danny, and Jennifer L. Lawless. 2015. “A Non-Gendered Lens? Media, Voters, and Female Candidates in Contemporary Congressional Elections.” *Perspectives on Politics* 13 (1): 95–118. doi:10.1017/S1537592714003156.
- Hayes, Danny, Jennifer L. Lawless, and Gail Baitinger. 2014. “Who Cares What They Wear? Media, Gender, and the Influence of Candidate Appearance.” *Social Science Quarterly* 95 (5): 1194–212. doi:10.1111/ssqu.12113.
- Heilman, Madeline E., and Alice H. Eagly. 2008. “Gender Stereotypes Are Alive, Well, and Busy Producing Workplace Discrimination.” *Industrial and Organizational Psychology* 1 (4): 393–98. doi:10.1111/j.1754-9434.2008.00072.x.
- Heilman, Madeline E., and Elizabeth J. Parks-Stamm. 2007. “Gender Stereotypes in the Workplace: Obstacles to Women’s Career Progress.” In *Social Psychology of Gender*. Advances in Group Processes, vol. 24, 47–77. Bingley: Emerald. doi:10.1016/S0882-6145(07)24003-2.
- Herrick, Rebekah, and Almira Sapieva. 1998. “Perceptions of Women Politicians in Kazakhstan.” *Women & Politics* 18 (4): 27–40. doi:10.1300/J014v18n04_02.
- Hibbing, John R., Kevin B. Smith, and John R. Alford. 2014. “Differences in Negativity Bias Underlie Variations in Political Ideology.” *Behavioral and Brain Sciences* 37 (3): 297–307. doi:10.1017/S0140525X13001192.
- Hinojosa, Magda, and Susan Franceschet. 2012. “Separate but not Equal: The Effects of Municipal Electoral Reform on Female Representation in Chile.” *Political Research Quarterly* 65 (4): 758–70.
- Holman, Mirya R., Jennifer L. Merolla, and Elizabeth J. Zechmeister. 2011. “Sex, Stereotypes, and Security: A Study of the Effects of Terrorist Threat on Assessments of Female Leadership.” *Journal of Women, Politics & Policy* 32 (3): 173–92. doi:10.1080/1554477X.2011.589283.
- Holman, Mirya R., Jennifer L. Merolla, and Elizabeth J. Zechmeister. 2016. “Terrorist Threat, Male Stereotypes, and Candidate Evaluations.” *Political Research Quarterly* 69 (1): 134–47. doi:10.1177/1065912915624018.
- Holman, Mirya R., Jennifer L. Merolla, and Elizabeth J. Zechmeister. 2017. “Can Experience Overcome Stereotypes in Times of Terror Threat?” *Research & Politics*. Published electronically February 1. <https://doi.org/10.1177/2053168016688121>
- Huddy, Leonie, and Theresa Capelos. 2002. “Gender Stereotyping and Candidate Evaluation: Good News and Bad News for Women Politicians.” In *The Social Psychology of Politics*, edited by Victor C. Ottati, R. Scott Tindale, John Edwards, Fred B. Bryant, Linda Heath, Daniel C. O’Connell, Yolanda Suarez-Balcazar, and Emil J. Posavac, 29–53. New York: Kluwer Academic.
- Huddy, Leonie, and Nayda Terkildsen. 1993a. “Gender Stereotypes and the Perception of Male and Female Candidates.” *American Journal of Political Science* 37 (1): 119. doi:10.2307/2111526.
- Huddy, Leonie, and Nayda Terkildsen. 1993b. “The Consequences of Gender Stereotypes for Women Candidates at Different Levels and Types of Office.” *Political Research Quarterly* 46 (3): 503–25. doi:10.2307/448945.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, and M. Lagos, et al., eds. 2014. “World Values Survey: Round Six—Country-Pooled Datafile Version.” JD Systems Institute, Madrid. www.worldvaluessurvey.org/WVSDocumentationWV6.jsp.
- Kefford, Glenn. 2013. “The Presidentialisation of Australian Politics? Kevin Rudd’s Leadership of the Australian Labor Party.” *Australian Journal of Political Science* 48 (2): 135–46. doi:10.1080/10361146.2013.786676.
- Kelley, Jonathan, and Ian McAllister. 1983. “The Electoral Consequences of Gender in Australia.” *British Journal of Political Science* 13 (3): 365–77. doi:10.1017/S0007123400003306.
- Key, V. O. 1966. *The Responsible Electorate: Rationality in Presidential Voting 1936-1960*. Cambridge: Harvard University Press.
- King, Amy, and Andrew Leigh. 2010. “Bias at the Ballot Box? Testing Whether Candidates’ Gender Affects Their Vote.” *Social Science Quarterly* 91 (2): 324–43. doi:10.1111/j.1540-6237.2010.00695.x.
- King, David C., and Richard E. Matland. 2003. “Sex and the Grand Old Party: An Experimental Investigation of the Effect of Candidate Sex on Support for a Republican Candidate.” *American Politics Research* 31 (6): 595–612. doi:10.1177/1532673X03255286.
- Koch, Jeffrey W. 1999. “Candidate Gender and Assessments of Senate Candidates.” *Social Science Quarterly* 80 (1): 84–96.
- Koch, Jeffrey W. 2000. “Do Citizens Apply Gender Stereotypes to Infer Candidates’ Ideological Orientations?” *The Journal of Politics* 62 (2): 414–29. doi:10.1111/0022-3816.00019.
- Koenig, Anne M., Alice H. Eagly, Abigail A. Mitchell, and Tiina Ristikari. 2011. “Are Leader Stereotypes Masculine? A Meta-Analysis of Three Research Paradigms.” *Psychological Bulletin* 137 (4): 616–42. doi:10.1037/a0023557.
- Kramer, Gerald H. 1971. “Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964.” *The American Political Science Review* 65 (1): 131–43.
- Krook, Mona Lena, and Diana Z. O’Brien. 2012. “All the President’s Men? The Appointment of Female Cabinet Ministers Worldwide.” *The Journal of Politics* 74 (3): 840–55. doi:10.1017/S0022381612000382.
- Lawless, Jennifer L. 2004. “Women, War, and Winning Elections: Gender Stereotyping in the Post-September

- 11th Era." *Political Research Quarterly* 57 (3): 479–90. doi:10.2307/3219857.
- Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18 (1): 349–66. doi:10.1146/annurev-polisci-020614-094613.
- Lawless, Jennifer L., and Kathryn Pearson. 2008. "The Primary Reason for Women's Underrepresentation? Reevaluating the Conventional Wisdom." *The Journal of Politics* 70 (1): 67–82. doi:10.1017/S002238160708005X.
- Leeper, Mark Stephen. 1991. "The Impact of Prejudice on Female Candidates: An Experimental Look at Voter Inference." *American Politics Quarterly* 19 (2): 248–61. doi:10.1177/1532673X9101900206.
- Matland, Richard E. 1994. "Putting Scandinavian Equality to the Test: An Experimental Evaluation of Gender Stereotyping of Political Candidates in a Sample of Norwegian Voters." *British Journal of Political Science* 24 (2): 273–92.
- McAllister, Ian. 2011. *The Australian Voter: 50 Years of Change*. Sydney: UNSW Press.
- Mo, Cecilia Hyunjung. 2015. "The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters." *Political Behavior* 37 (2): 357–95. doi:10.1007/s11109-014-9274-4.
- Morgan, Jana, and Melissa Buice. 2013. "Latin American Attitudes toward Women in Politics: The Influence of Elite Cues, Female Advancement, and Individual Characteristics." *American Political Science Review* 107 (4): 644–62.
- Newman, Brian. 2013. "Polls and Elections: Decreasing the Economy's Impact on Evaluations of the President: An Experiment on Attribution Framing." *Presidential Studies Quarterly* 43 (4): 866–82.
- O'Brien, Diana Z. 2015. "Rising to the Top: Gender, Political Performance, and Party Leadership in Parliamentary Democracies." *American Journal of Political Science* 59 (4): 1022–39.
- Ono, Yoshikuni. 2018. "Do Voters Prefer Gender Stereotypic Candidates? Evidence from a Conjoint Survey Experiment in Japan." *Political Science Research and Methods*. Published electronically October 2. <https://doi.org/10.1017/psrm.2018.41>
- Parker, Ashley. 2016. "Donald Trump Says Hillary Clinton Doesn't Have 'A Presidential Look.'" *The New York Times*, September 6. <https://www.nytimes.com/2016/09/07/us/politics/donald-trump-says-hillary-clinton-doesnt-have-a-presidential-look.html>.
- Paul, David, and Jessi L. Smith. 2008. "Subtle Sexism? Examining Vote Preferences When Women Run against Men for the Presidency." *Journal of Women, Politics & Policy* 29 (4): 451–76. doi:10.1080/15544770802092576.
- Rosenwasser, Shirley M., and Norma G. Dean. 1989. "Gender Role and Political Office: Effects of Perceived Masculinity/Femininity of Candidate and Political Office." *Psychology of Women Quarterly* 13 (1): 77–85. doi:10.1111/j.1471-6402.1989.tb00986.x.
- Rosenwasser, Shirley M., and Jana Seale. 1988. "Attitudes toward a Hypothetical Male or Female Presidential Candidate: A Research Note." *Political Psychology* 9 (4): 591–98. doi:10.2307/3791529.
- Rozin, Paul, and Edward B. Royzman. 2001. "Negativity Bias, Negativity Dominance, and Contagion." *Personality and Social Psychology Review* 5 (4): 296–320. doi:10.1207/S15327957PSPR0504_2.
- Rudolph, Thomas J. 2006. "Triangulating Political Responsibility: The Motivated Formation of Responsibility Judgments." *Political Psychology* 27 (1): 99–122.
- Sanbonmatsu, Kira. 2002. "Gender Stereotypes and Vote Choice." *American Journal of Political Science* 46 (1): 20–34. doi:10.2307/3088412.
- Sanbonmatsu, Kira, and Kathleen Dolan. 2009. "Do Gender Stereotypes Transcend Party?" *Political Research Quarterly* 62 (3): 485–94. doi:10.1177/1065912908322416.
- Sapiro, Virginia. 1981. "If U.S. Senator Baker Were a Woman: An Experimental Study of Candidate Images." *Political Psychology* 3 (1/2): 61–83. doi:10.2307/3791285.
- Sawer, M. 2002. "The Representation of Women in Australia: Meaning and Make-Believe." *Parliamentary Affairs* 55 (1): 5–18. doi:10.1093/parlij/55.1.5.
- Schneider, Monica C., and Angela L. Bos. 2016. "The Interplay of Candidate Party and Gender in Evaluations of Political Candidates." *Journal of Women, Politics & Policy* 37(3):274–94. doi:10.1080/1554477X.2016.1188598.
- Schwindt-Bayer, Leslie A., and Catherine Reyes-Housholder. 2017. "Citizen Responses to Female Executives: Is It Sex, Novelty, or Both?" *Politics, Groups, and Identities* 5:373–98. doi:10.1080/21565503.2017.1283238.
- Sevi, Semra, Vincent Arel-Bundock, and André Blais. 2019. "Do Women Get Fewer Votes? No." *Canadian Journal of Political Science* 52:201–10. doi:10.31235/osf.io/sw8da.
- Smith, Eric, and Richard Fox. 2001. "The Electoral Fortunes of Women Candidates for Congress." *Political Research Quarterly* 54 (1): 205–21.
- Soroka, Stuart. 2014. *Negativity in Democratic Politics: Causes and Consequences*. Cambridge: Cambridge University Press.
- Soroka, Stuart, and Stephen McAdams. 2015. "News, Politics, and Negativity." *Political Communication* 32 (1): 1–22. doi:10.1080/10584609.2014.881942.
- Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112:525–41. doi:10.1017/S0003055418000217.
- Thomas, Melanee. 2018. "In Crisis or Decline? Selecting Women to Lead Provincial Parties in Government." *Canadian Journal of Political Science* 51 (2): 379–403.
- Thomas, Melanee, and Marc Bodet. 2013. "Sacrificial Lambs, women candidates, and district competitiveness in Canada." *Electoral Studies* 32:153–66.
- Trimble, Linda. 2017. *Ms. Prime Minister: Gender, Media, and Leadership*. Toronto: University of Toronto Press.
- Wellington, Sheila, Marcia B. Kropf, and Paulette R. Gerkovich. 2003. "What's Holding Women Back?" *Harvard Business Review*, June. <https://hbr.org/2003/06/whats-holding-women-back>.