BEYOND DOUBLE STANDARDS: HOW GENDER CONDITIONS POLITICAL RESPONSIBILITY

By

Georgia M. Anderson-Nilsson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Political Science

August 31, 2021

Nashville, Tennessee

Approved:

Cindy Kam, Ph.D.

Amanda Clayton, Ph.D.

Larry Bartels, Ph.D.

Cecilia Hyunjung Mo, Ph.D.

Kira Sanbonmatsu, Ph.D.

This dissertation is dedicated to my husband, Tyler, who moved across the country to support my dreams. He made me countless cups of coffee, cooked me hundreds of meals, reminded me to take breaks from writing, provided a sounding board for ideas, and reassured me of my abilities and my strength over and over again. Thank you, sincerely and eternally, for being there to support me every step of the way.

ACKNOWLEDGMENTS

This dissertation project would not have been possible without the consistent support and guidance of my advisers. To Cindy Kam, I am deeply grateful for your patience, for the many hours you put into meeting with me and reading my drafts, for your consistently thoughtful comments, for your close eye for detail, and for your immense skill in experimental methods. It is no easy task to balance encouragement with constructive criticism, but you manage it with grace. You have helped me become a more critical, careful thinker and researcher. To Amanda Clayton, thank you for your consistent support and understanding, for your attentive feedback, for the many opportunities you have provided me, and for lending your considerable expertise on the larger gender literature. You have been an excellent role model for me throughout graduate school in many different ways.

To Larry Bartels, Cecilia Hyunjung Mo, and Kira Sanbomatsu: I am enormously thankful for the time and energy each of you have dedicated to supporting this dissertation through to its completion. To Larry, your work is simply fundamental to my conception of American democracy; it informed much of this project and continues to shape how I think about just about every political question. It has been an honor to have you as a professor and committee member. To Cecilia, your mentorship and teaching during my first years of graduate school was instrumental in shaping how I think about political behavior and political psychology. During our various collaborations, you taught me a tremendous amount about the process of producing high-quality work; how to frame questions and actually get them on paper. My ability to finish this dissertation - and any other project, for that matter - is largely thanks to you. To Kira, your seminal research on gender in American politics ignited my interest in gender and inspired many of the questions I asked in this project. I was beside myself when you agreed to join my committee, and remain just as thrilled today.

I also want to acknowledge Claire Evans, who patiently helped me learn to code and analyze data. Claire, you answered late night and weekend texts and calls, and worked alongside me in coffee shops. You probably saved me hundreds of hours of troubleshooting, and I can't thank you enough for your generosity and kindness. To Meg Frost and Darrian Stacy, my fellow graduate students and friends - thank you for always lending an ear, for providing perspective, and for having my back. Last but certainly not least, to my parents, Julie and Eric: thank you for making me figure out how to explain my research clearly (and make it sound interesting), for reading my work without any formal methods or political science training, and for never once wavering in your support of this project or my pursuit of this degree. None of this would be realizable without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Both the 2018 and the 2020 U.S. election cycles earned nicknames for the unprecedented numbers of female candidates who ran for and won political office. The 2018 American midterm elections were described as having brought a 'pink wave' to American politics, with record numbers of Democratic female candidates entering races across the country. Pundits, activists, and journalists dubbed 2018 'another year of the woman,' (Kamarck, 2018; Yarvin, 2017). In 2020, a record number of Republican women candidates threw their proverbial hats into the electoral ring, earning that election cycle the moniker 'the year of the Republican woman' (Lips, 2020; Olohan, 2020). Additionally, as of July 2021, a record number of Republican women have already announced they plan to run for House seats in 2022 - doubling their numbers compared to this stage in the 2020 election cycle (Brewster, 2021). A record number of women also serve in state legislatures in 2021, with majority-female state legislative chambers in seven states (Dittmar and Hill, 2020).

What's more, once women reach elective office, they excel. On average, women sponsor more bills, bring more money back to their districts, and more closely represent their districts' ideological preferences (Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018). Bills sponsored by women survive longer in the legislative process, attract more media coverage, and are more likely to be regarded as important (Dokoupil, 2011; Volden et al., 2013), and women speak on the House floor at higher rates than men (Pearson and Dancey, 2011). Despite their disproportionately high policy activity, female legislators are also more likely to respond to constituent requests (Thomsen and Sanders, 2020). Women are also disproportionately effective when they are in the minority party, perhaps because they tend to be more collaborative and consensus-oriented, working across the partisan aisle more often (Volden et al., 2013; Jeydel and Taylor, 2003; Duerst-Lahti, 2002; Rosenthal, 1998; Holman and Mahoney, 2018).

Public opinion has also shifted in favor of women in political roles along a variety of dimensions. Most Americans now explicitly support gender equality, and a majority believe there are too few women in political office (Minkin, 2020; Pew Research Center, 2018). Large and growing majorities of American voters also say they would be willing to vote for a female president (Carnevale et al., 2019). Moreover, relative to men the public tends to view women in politics as more honest, ethical, trustworthy, and better at improving the quality of life for Americans (Pew Research Center, 2018; Barnes and Beaulieu, 2019; Cassese and Holman, 2017). Voters also tend to believe women are more compassionate than men, and are better at serving as role models and creating a tone of civility in politics (Pew Research Center, 2018).

Despite these recent trends, women remain vastly underrepresented in American government. In 2021,

women comprise just over a quarter of Congress, and just over a third of state legislatures (Center for American Women in Politics, 2021a). The proverbial glass ceiling also remains intact; women's descriptive underrepresentation increases as the level of office and scope of its authority increases (Carroll and Sanbonmatsu, 2013). Women are significantly less likely to occupy executive office compared to legislative office: there are only eight female governors, and just over a quarter of U.S. cities with populations greater than 30,000 have women mayors. (Center for American Women in Politics, 2021b). When women do occupy executive office, they are far more likely to run for positions that correspond with stereotypical feminine competencies, such as education or healthcare (Fox and Oxley, 2003a). In public agencies, men occupy a disproportionate share of middle- and upper-management positions (Anestaki et al., 2016; Choi, 2011; Clark et al., 2013; Crumpacker and Crumpacker, 2008).

The canonical view in the gender and politics literature tends to attribute women's continued underrepresentation to the gender gap in political ambition. Given that men and women are equally likely to win elections, and tend to garner similar electoral support, this perspective argues that women's descriptive underrepresentation is mainly due to the dearth of female candidates (Lawless, 2015). In other words, if only women ran at the same rates as men, we would achieve gender parity in government. This view has also trickled into popular media. As one journalist recently put it, "[g]ender parity has remained distant not because voters dislike female candidates - women win elections at the same rates as men - but because not enough women run in the first place" (Grethen, 2018). A 2016 New York Times editorial similarly contends, "[w]hen women run for political office, they are just as likely as men to be elected. The main reason they are so underrepresented is that they don't run in the first place" (Miller, 2016).

Women's increasing success in the political realm has been widely celebrated, and many of these recent changes have been heralded as signaling a sea change in American politics. However, this perspective belies the significant, unique, gender-based challenges women continue to face in American elections. For instance, although women sponsor significantly more legislation related to women's issues, women's issue bills are less likely to successfully reach the House floor and become law than other types of legislation; moreover, women's issue bills sponsored by men are significantly more likely to succeed than those sponsored by women (Volden et al., 2018). Women also face biased media coverage: coverage of elections that include female candidates is more likely to focus on candidate traits rather than policy issues, and news articles on female candidates provide significantly less information related to their political and professional qualifications compared to men (Dunaway et al., 2013; Bauer, 2020b).[1] The news media are also less likely to rely on information that comes directly from female candidates and officeholders, a tendency that can harm female candidates at least in part because candidate information that originates from outside sources is more likely

[1] But see Hayes and Lawless (2015); Hayes et al. (2014).

to activate harmful gender stereotypes for women (Fridkin and Kenney, 2015; Bos, 2011). Women also face more challengers at the primary stage, and experience systematically different barriers related to fundraising and recruitment processes (Thomsen, 2019; Shah et al., 2018; Thomsen and Swers, 2017; Ondercin and Welch, 2009; Pearson and McGhee, 2013).

Moreover, a variety of recent studies argue that far from signaling an even playing field, equivalent electoral performance by men and women is suggestive of gendered qualification standards. From this perspective, women should be *outperforming* men at the ballot box because they are higher quality candidates and more effective in office. Indeed, recent work shows that voters hold female candidates to systematically higher qualification standards (Bauer, 2020b,c; Fulton and Dhima, 2020a). Voters seek out significantly more competence-related information about female candidates, and information that calls candidates' qualifications into question is far more harmful for women than for men (Ditonto, 2017; Ditonto et al., 2014a). Various experimental studies have concluded that female candidates must indeed be more qualified than men in order to perform as well in elections (Mo, 2014; Bauer, 2020c). Scholars have also leveraged experimental techniques aimed at measuring social desirability bias to make the case that public opinion surveys - particularly those that ask about hypothetical female candidates - likely overestimate the public's willingness to support real female candidates (Krupnikov et al., 2016; Streb et al., 2008; Burden et al., 2016).

The ways these gendered standards translate to female officeholders remain largely unexplored. How does gender condition the way voters evaluate officeholder performance? Do voters credit women with their policy accomplishments, and how do they factor these assessments into their voting decisions? In this dissertation project, I explore these questions. I examine the ways in which officeholder gender conditions performance evaluation and the attribution of political responsibility. I approach this question from several different angles. In Part I, I take up the question of whether gender conditions political performance evaluation and the attribution of political responsibility. I conduct an original survey experiment that aims to understand whether there are systematic gender differences in the way voters evaluate and attribute responsibility for political performance. In Part II, I examine electoral returns for members of the U.S. House of Representatives, asking whether the rewards associated with legislative activity differ in systematic ways for male and female members. This chapter asks, do men and women get the same "bang for their buck"? Finally, in Part III, I ask whether men and women experience different evaluative consequences for directly claiming credit for policy accomplishments. I conduct an experiment that tests whether voters evaluate men and women who tout their policy accomplishments experience in different ways.

I hypothesize broadly that compared to men, women should experience lower evaluations and receive less credit for equivalent performance. In crafting my theoretical expectations, I draw primarily on role congruity theory, which holds that the incongruity between feminine and leadership stereotypes creates barriers to

leadership positions for women. This perspective contends that women are stereotyped as being more gentle, caring, nurturing, and sensitive, while men are stereotyped as being more aggressive and dominant (Eagly and Karau, 2002; Koenig et al., 2011) Feminine stereotypes align with communal, rather than agentic roles, so leadership is more closely and readily associated with masculine stereotypes. Hence, women are less likely to be seen as possessing the agentic traits necessary to perform leadership roles. Moreover, because individuals tend to react negatively when stereotypical expectations are violated, women also tend to suffer a penalty when they do occupy leadership positions stemming from the incongruity between gender and leadership stereotypes (Simon and Hoyt, 2008; Eagly and Karau, 2002). Stereotypical expectations can also generate low performance expectations that make voters less likely to recognize women's policy accomplishments, and less likely to credit women for those accomplishments when they do (Heilman and Haynes, 2005). On the flip side, stereotypical expectations work to men's advantage in a variety of ways; for instance, men are more likely to be viewed as strong leaders and tend to be attributed with more influence over decisions in mixed-gender groups (McClean et al., 2018; Karpowitz et al., 2012).

This dissertation contributes to the literature on gender in American politics in several ways. First, if women are held to different performance standards in office, or given less credit for their accomplishments, then this may help us understand not only why there are fewer women than men in politics, but also why women so often face the proverbial glass ceiling in their access to higher and more prestigious positions (Sabharwal, 2013; Bruckmüller et al., 2014; Escobar-Lemmon and Taylor-Robinson, 2009; Krook and O'Brien, 2012). Second, although there is strong evidence that women outperform men in office, it is not clear whether they actually need to perform at a higher level in order to secure reelection. Although voters hold female candidates to higher standards, this may not apply to officeholders; in other words, it is possible that higher-quality female candidates simply lead to higher-quality female representatives. It is also possible that women overperform because they feel they must. In fact, women in politics often express the feeling that they must do more to prove themselves (FiveThirtyEight, 2020; Astor, 2019). As Laura Sturgeon, State Senator from Delaware put it,

> "I just think we women are more likely to say, 'Is just being relatable enough? Don't I have to prove myself? Don't I have to have a certain gravitas in order to be seein in this role?' And maybe we do. But maybe it's just sort of an expectation we put on ourselves." (FiveThirtyEight, 2020)

This project also adds to a growing body of work that argues that studying vote choice alone is not sufficient to understand the true nature of attitudes toward women in politics among the American electorate (i.e., Costa, 2020; Anderson-Nilsson and Clayton, 2020; Bauer, 2020b; Krupnikov et al., 2016). Voters may express

explicit support for women in political roles, but react differently to the exercise of political power by a female leader. Elected officials engage in a wide variety of activities while in office that correspond with or violate feminine gender stereotypes to varying degrees, and past work suggests at least some of these activities are indeed judged differently for men and women (Costa, 2020; Vraga, 2017; Okimoto and Brescoll, 2010). This project adds new dimensions to our understanding of the ways voters react when politicians carry out the responsibilities of political office, and how gender conditions this relationship.

Finally, this project contributes in a broad sense to our understanding of democratic accountability in American politics. In an ideal democracy, voters should reward and punish incumbents similarly for similar representation. If voters hold men and women to different performance standards, this fundamentally jeopardizes the democratic principle that elections should enable voters to select the best candidates for the job, and corrodes the overall quality of democratic representation. Thus, this project also sheds additional light on the factors that condition the functioning of American democracy, writ large.

CHAPTER 2

Part I: Does Gender Condition Responsibility? An Experimental Test of Evaluation and Attribution Bias in American Politics

In the classic view of democracy, elections turn the public into a "rational god of vengeance and reward" (Key, 1964, 568). In this envisioning, citizens judge political representatives on the basis of their performance, keeping incumbents who have performed satisfactorily and "throwing out the rascals" who have not (Kramer, 1971; Stokes, 1963, 373). Conscientious voters pay particular attention to how well elected officials represent their interests; indeed, citizens are thought to keep an issue-based "running tally" of their representatives' performance, factoring in new information into their estimation of representatives' ability to govern and represent their interests effectively (Fiorina, 1981). In this world, elections ensure accountability by inducing re-election motivated representatives to produce good outcomes on behalf of their constituencies (Mayhew, 2004; Federalist 52, 1788; Ferejohn, 1986).

Performance evaluations have been shown to play an important role in both electoral choice and public opinion of political leaders (Feldman, 1981; Iyengar, 1990; Lenz, 2012; Iyengar, 1987; Achen and Bartels, 2015). Yet a variety of cognitive and informational limitations make it extremely difficult for the average voter to accurately ascribe credit and blame to political actors for outcomes and performance. Thus, in order to render political judgments, people tend to rely upon heuristics and engage in biased reasoning processes (Tversky and Kahneman, 1974; Bendor, 2002). Hence, the attribution of responsibility for outcomes and behaviors tends to be socially constructed and is often deeply flawed (Achen and Bartels, 2015; Taber et al., 2001). For example, incumbents have been found to suffer at the polls as a result of events as seemingly irrelevant as shark attacks and disappointing college basketball games (Achen and Bartels, 2015; Healy et al., 2010).

Voters are also likely to employ heuristics derived from candidates' and politicians' demographic traits to infer qualities, competencies, and issue positions (Popkin, 1991), which may also introduce bias into political evaluation. A politician's gender provides one such shortcut. Voters have been shown to employ gender stereotypes to infer candidate traits, competencies, and beliefs, which in turn affect their baseline preference for men or women candidates (Sanbonmatsu, 2002b). For example, women are expected to be more communal - or more nurturing, gentle, and sensitive - while men are expected to be more agentic, or more dominant and aggressive (Koenig et al., 2011; Huddy and Terkildsen, 1993b,a). Women are more likely to be viewed as competent in policy areas that are associated with such stereotypically feminine traits as compassion and care, including "topics like children, education, health, and women's affairs" (Fridkin and

Kenney, 2009; Fox and Oxley, 2003b; Dolan, 2010; Krook and O'Brien, 2012, 844), while men are more readily associated with issues such as terrorism, military or police issues, business and finance issues, and crime[1] (Dolan, 2010; Fox and Oxley, 2003a; Krook and O'Brien, 2012).

In this project, I specifically explore whether a politician's gender conditions performance evaluation and the attribution of political responsibility. Are men and women rewarded for positive political outcomes at equal rates? Conversely, when conditions are poor, are voters more apt to punish male or female leaders? The results of this experiment indicate that voters do evaluate equivalent performance by men and women in different ways, but these effects are highly conditional on respondent gender and partisanship. Across the entire sample of respondents, the only significant difference that emerges is that women are blamed significantly *less* than men for poor performance. Further analyses indicate that this finding is driven by women and Democrats. These findings suggest that officeholder gender alone does not necessarily drive differences in political evaluation; instead, voter characteristics interact with officeholder gender to condition the electoral connection.

2.1 Voter Evaluations of Women in Politics

A vast literature in political science has shown that men and women now generally fare equally well at the ballot box, uncovering little to no systematic evidence of electoral penalties against female candidates or officeholders on the basis of gender alone (Dolan, 2014; Hayes and Lawless, 2016; Brooks, 2011; Lawless and Pearson, 2008). Moreover, a variety of studies testing the impact of gender on the evaluation of candidates have found that voters often actually express a preference for female candidates (Schwarz et al., 2018). Moreover, the public tends to hold positive trait expectations of women in politics; for example, the public tends to believe women to be more honest and ethical, more collaborative, and better at working to improve quality of life in the U.S. (Pew Research Center, 2015; Barnes and Beaulieu, 2019; Ulbig, 2007; Watson and Moreland, 2014; Cassese and Holman, 2017; Diekman et al., 2002; Holman et al., 2016).

On the other hand, recent research has demonstrated that social desirability bias may lead studies to misrepresent true levels of gender bias toward women in politics in the American public (Krupnikov et al., 2016; Streb et al., 2008). Additionally, a bevy of recent studies have shown that women in politics tend to be systematically higher-quality than their male counterparts, outperforming men both as candidates and as officeholders (Fulton, 2011; Pearson and McGhee, 2013; Volden et al., 2013; Fulton, 2014; Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018). This quality differential may stem from differences in political

[1] However, Fridkin and Kenney (2009) analyzed survey data to find that women senators were actually seen as superior to men on both stereotypically female and stereotypically male traits and issue areas, and Lawless and Hayes (2015) found no sex differences in evaluations of candidates for the U.S. House in 2010. However, this may have something to do with the fact that women who do reach political office tend to outperform their male counterparts along a variety of dimensions (Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018).

ambition between men and women; for instance, women tend to doubt their qualifications to run for office more than men (Lawless, 2015). There is also evidence that voters seek out significantly more competence-related information for female candidates than for male candidates, and that information that casts doubt on candidates' competence affects female candidates more negatively than male candidates (Ditonto et al., 2014a; Ditonto, 2017). Moreover, there is some evidence that female candidates need to be more qualified than men in order to achieve the same level of electoral support (Bauer, 2020c; Mo, 2014).

2.1.1 Gendered Performance Standards

A wide body of research suggests that men and women tend to be evaluated differently when they hold equal or similar roles. Particularly in higher-status or leadership roles, women tend to elicit more negative evaluations than men, and their status is often devalued relative to men (Butler and Geis, 1990; Rudman and Glick, 2001; Koch, 2005). For instance, male instructors are also more likely to be referred to as "professor," while women are more likely to be called "teacher" (Mitchell and Martin, 2018). Women are also more likely to be be judged based on personality characteristics - particularly communal, interpersonal traits[2] - while men tend to be judged based on factors such as performance and qualifications (Mitchell and Martin, 2018; Biernat et al., 2012; Heilman and Chen, 2005). In business, men tend to be promoted based on beliefs surrounding their leadership potential, while women are more likely to be promoted based on leadership performance (Player et al., 2019). Moreover, successful female managers are often characterized as cold, manipulative, or selfish (Heilman and Martell, 1995; Heilman et al., 2004). For instance, in the tech industry, high-achieving women are more likely than their male colleagues to receive critical feedback in performance reviews, and that feedback is disproportionately likely to contain personality-related criticisms (Snyder, 2014). However, these effects can be attenuated or even reversed by the provision of personal information that makes women seem warmer (Rudman and Glick, 2001; Heilman and Haynes, 2005). In other words, individuals are likely to react more positively to a highly successful female manager *who is also a mother* than they are to a highly successful female manager about whom they have no additional information.[3]

Men and women also tend to be judged differently for similar performance. On average, even when their performance is objectively identical to men's, women are likely to receive more negative evaluations (Nieva and Gutek, 1980; Cejka and Eagly, 1994; Heilman, 2001, 1995; Heilman and Martell, 1995; Botelho and Abraham, 2017; Eagly et al., 1992). In education, female instructors tend to receive lower ratings than male instructors on competence, professionalism, effectiveness, and credibility, especially in larger courses (Martin, 2016; Boring, 2017; MacNell et al., 2014; Miller and Chamberlin, 2000; Bianchini et al., 2012). Gender

[2]For instance, women tend to be punished when they are perceived as lacking in favorable, communal interpersonal characteristics such as altruism, but not rewarded for possessing them; conversely, men tend to be rewarded for displaying such characteristics but not punished for lacking them (Heilman and Chen, 2005).

[3]The provision of such information does not impact evaluations of male managers.

has also been found to influence patterns of research funding and publication in favor of men (Witteman et al., 2019; Biernat et al., 2019; Teele and Thelen, 2017). In the field of economics, female-authored papers have been face significantly higher publication standards (Hengel, 2017). Women also tend to face higher standards for promotion across a variety of organizations (Inesi and Cable, 2015). In one study, Botelho and Abraham (2017) showed that the recommendations of male investment professionals were selected 33% more often than those of similarly performing female investment professionals. This disparity only disappeared at the highest performance levels.

There is also strong evidence that attribution for performance functions differently for men and women. Because women tend to be seen as less competent than men, they tend to receive less credit for accomplishments and more blame for mistakes, particularly in traditionally masculine realms. For example, Heilman and Haynes (2005) found that when respondents were asked to evaluate male-female teams that had produced objectively high-quality work, evaluations tended to give the male partner more credit, unless respondents were provided with additional information detailing the female partner's specific contributions and skills. In a similar vein, even when men and women read the same words from a script or speak up equally often in group settings, men are far more likely than women to be viewed as strong leaders and to be seen as having influenced outcomes (McClean et al., 2018; Karpowitz et al., 2012). Research in other fields has found that strong performance in mathematics is more often attributed to innate ability and logical reasoning skills for boys, but to exceptional effort for girls (Tiedemann, 2000; Dickhäuser, 2006). Male company executives also tend to be rewarded with larger bonuses relative to female executives for equally strong performance (Kulich et al., 2007, 2011).

Most empirical investigations of gendered evaluation bias have taken place outside of political science, in fields such as management, education, and psychology.[4] However, there is some work that suggests gendered evaluation bias may operate similarly within the realm of politics. For example, women have been found to be penalized more than men for corrupt behavior in political office (Eggers et al., 2018). Relative to men, women also tend to be disproportionately punished for engaging in negative campaigning and for displaying power-seeking intentions (Krupnikov and Bauer, 2014; Cassese and Holman, 2017; Okimoto and Brescoll, 2010). Constituents also reward and punish male and female representatives differently for district service and perceived friendliness (Costa and Schaffner, 2018; Costa, 2020).

[4]For additional examples, see Reuben et al. (2014); Nieva and Gutek (1980); Sekaquaptewa (2011); Romero and Garza (1986); Tosi and Einbender (1985); Heilman and Martell (1995)

2.2 Theory

Performance evaluation entails two main components: first, judgment regarding the quality of performance, and second, the attribution of causality for performance (Nieva and Gutek, 1980; Feldman, 1981; McGraw, 2001). Theories of attribution generally hold that performance may be attributed to four causes: task difficulty, effort, ability, and luck. These vary along two dimensions: internal versus external and stable versus unstable (Weiner and Rosenbaum, 1971; Nieva and Gutek, 1980). If performance is attributed to ability, it is both stable and internal, and thus seen as repeatable; conversely if it is attributed to luck, it is both unstable and external and thus not seen as repeatable (Nieva and Gutek, 1980; Heilman, 2001). These performance assessments in turn factor into larger beliefs regarding competence. For instance, positive performance that is attributed to internal causes such as ability and effort boosts perceptions of competence, while positive performance that is attributed to external factors such as task difficulty or luck does not (Ibid).

2.2.0.1 Gendered Evaluation & Attribution Bias

Gender systematically conditions and often biases both the processes of judging performance and attributing causality. Descriptive social norms tend to cast women as possessing communal characteristics and men as possessing more agentic traits. Thus, women tend to be stereotyped as being gentle, caring, nurturing, and sensitive, while men are stereotyped as being more aggressive and dominant (Eagly and Karau, 2002; Koenig et al., 2011). Because feminine stereotypes align with communal, rather than agentic roles, leadership is more closely and readily associated with masculine stereotypes, and women are less likely to be seen as possessing the agentic traits necessary to perform leadership roles effectively (Eagly and Karau, 2002). These expectations can lead to a perceived "lack of fit" between gendered characteristics and the requirements of leadership roles, which in turn can lead to lower performance expectations and ultimately biased evaluations (Heilman, 2001; Foschi, 2000). Moreover, when women do occupy leadership roles, this generates a psychological incongruity between descriptive and prescriptive expectations of women's attributes and behavior, and the agentic, masculine characteristics typically associated with success in leadership roles, resulting in a negative reaction (Eagly and Karau, 2002; Heilman and Haynes, 2005; Heilman, 2001; Simon and Hoyt, 2008).

Thus, judgments about the quality of women's performance are likely to be negatively biased because women are expected to be less competent in roles that are more readily associated with masculine characteristics (McClean et al., 2018; Ridgeway, 2001), and because women in these roles are likely to experience an evaluative penalty. This type of evaluative bias against women tends to be stronger in stereotypically masculine areas, and where there is greater ambiguity in assessing performance quality (Nieva and Gutek, 1980; Botelho and Abraham, 2017; Cejka and Eagly, 1994; Heilman and Martell, 1995).

An additional reason to expect that women will suffer an evaluative bias relates to the way humans process

negative and surprising information. Due to the limitations of human cognitive capacity, people must select what they pay attention to at any given time (Huang and Sherman, 2018; Bendor, 2002; Hogarth, 1987). Humans are highly adaptive in how they selectively direct their attention, primarily aiming to reduce uncertainty (Luque et al., 2017). This implies that attention is drawn to different degrees by different types of information (Huang and Sherman, 2018). The more attention a piece of information receives, the more likely it is to factor into an individual's evaluation (Huang and Sherman, 2018). In making social evaluations, people tend to pay greater attention to the behaviors and characteristics of other individuals that do not conform with their socially constructed expectations (Fiske et al., 1999; Hilton et al., 1991; Rudman and Glick, 2001; Belmore, 1987). In general, negative behaviors also attract more attention than positive behaviors, and negative information also tends to figure more heavily than positive information into social evaluations (Fiske, 1980; Belmore, 1987; Pratto and John, 1991; Skowronski and Carlston, 1987; Taylor, 1991).

Hence, if the tenets of role congruity theory hold, women in leadership positions are both (a) engaging in unexpected social behavior and (b) generating a negative reaction, both of which draw relatively high levels of limited human attention, and thus may also play a role in negatively biasing evaluations of women relative to men. The above discussion leads to the following testable expectations:

> **H1:** *Relative to men, women will receive lower evaluations for neutral performance in political office.*
>
> **H1a:** *Relative to men, women should receive a smaller evaluative boost for positive performance relative to neutral performance in political office.*
>
> **H1b:** *Relative to men, women should suffer a greater evaluative penalty for negative performance relative to neutral performance in political office.*

Attribution for performance should also operate differently for men and women. Descriptive and prescriptive expectations tend to cast women as unlikely to possess the characteristics required to perform the functions of leadership roles successfully, making it unlikely that women will receive as much credit for accomplishments as men (Nieva and Gutek, 1980; Deaux and Emswiller, 1974; Cash et al., 1976; Swim and Sanna, 1996). Moreover, there is a strong psychological propensity to support and maintain stereotypical expectations about women's performance in traditionally masculine areas (Heilman and Haynes, 2005; Heilman, 2001). Stereotyped expectations are highly functional in the sense that they help human beings to simplify a complex cognitive environment by placing individuals in social categories (Fiske, 1980). Because attention is directed adaptively, with the primary aim of reducing uncertainty, unexpected behavior can generate discomfort and tends to be rationalized in ways that comport with expectations (Luque et al., 2017; Deaux and Emswiller, 1974; Biernat et al., 2010). Thus in order to maintain consistency with social expectations, individuals may attribute women's successes in male domains to external factors rather than to women's abilities (Heilman and Haynes, 2005).

For these reasons, women's successes in leadership roles are more likely to be attributed to external factors such as luck or to internal but unstable factors such as effort, rather than to stable and internal factors like ability or skill. Conversely, failures are more likely to be blamed directly on women leaders' lack of ability than attributed to external factors. Hence, I test the following hypotheses:

> **H2:** *Relative to men, women's neutral performance in political office will be more likely to be attributed to external rather than internal factors.*
>
> **H2a:** *Relative to men, and relative to the neutral performance condition, women's positive performance in political office will be more likely to be attributed to external rather than internal factors.*
>
> **H2b:** *Relative to men, and relative to the neutral performance condition, women's negative performance in political office will be more likely to be attributed to internal rather than external factors.*

In sum, women may receive less credit for positive outcomes, and more blame for negative outcomes for several reasons. First, voters should be less likely to evaluate women in leadership roles favorably for their successes (and more likely to penalize them for their failures) (a) due to the incongruity between the characteristics required for success in leadership roles and the expectations associated with both prescriptive and descriptive feminine stereotypes, (b) because they have lower performance expectations of women as leaders to begin with and, (c) because of the human tendency to pay more attention to, and weight more heavily, negative and/or surprising information. Second, voters should be less likely to attribute women's positive policy performance to stable and/or internal causal factors such as skill or competence, instead attributing favorable outcomes to external factors such as luck or the work of others, or to internal but unstable factors such as effort. On the other hand, voters should be more likely to attribute negative performance information to women's lack of skill or effort; conversely, negative outcomes will be more likely to be attributed to external factors for men.

2.2.0.2 Heterogeneous Effects

Respondent gender may also moderate performance evaluations and the attribution of responsibility. Positive bias toward in-group members or negative bias toward out-group members can color evaluation (Huddy et al., 2015). In general, men (women) are more likely than women to prefer male (female) candidates (Sanbonmatsu, 2002b; Dolan, 1997). Thus, women (men) may reward women (men) more and penalize them less, and vice versa. For this reason, I also analyze evaluations by respondent gender.

> **H3:** *Evaluative penalties against women will be strongest, and rewards weakest, among men. Evaluative penalties against women will be weakest, and rewards strongest, among women.*

I also consider heterogeneous effects by party. There is reason to expect that partisanship should have consequences for the way women in politics are evaluated. Democratic and Republican women in politics

face different stereotypes, expectations, and incentive structures (Gimenez et al., 2016; Hayes, 2011; Holman et al., 2011, 2016; Ono and Burden, 2017; Sanbonmatsu, 2002b; Sanbonmatsu and Dolan, 2009; Teele et al., 2018). Women may be at a relative advantage among Democrats, as Democratic voters are more likely to care about gender equality in government (Sanbonmatsu and Dolan, 2009). Democratic donors also tend to favor more liberal women candidates, and women's PACs disproportionately favor Democratic women in the earlier stages of elections (Thomsen and Swers, 2017; Francia, 2001). Conversely, Republican voters tend to hold more traditional attitudes with regard to gender roles (Knuckey, 2005). Moreover, because women tend to be stereotyped as more liberal than they actually are, Republican women may be at a disadvantage among conservative voters, while Democratic women may enjoy an advantage among liberal voters (Sanbonmatsu and Dolan, 2009; King and Matland, 2003; Dolan, 2004; Koch, 2000). There is indeed evidence of this: for example, Burden and Ono (2018) show that Republicans are likely to exhibit a negative bias toward women when choosing between male and female Republican candidates. King and Matland (2003, 602) find that being female significantly reduces the likelihood that Republican voters describe a Republican candidate as a "strong leader," and that relative to an otherwise identical male candidate, Republicans are much less likely to support a female Republican candidate, even among women voters. Republican women are also more likely to be concerned with the negative impact of gender stereotypes on their electoral prospects (Sanbonmatsu and Dolan, 2009). Thus, I also examine heterogeneous effects by party, as well as the interactive effects of gender and party.[5]

> **H3a:** *Evaluative penalties against women will be strongest, and rewards weakest, among Republicans, especially Republican men. Evaluative penalties against women will be weakest, and rewards strongest, among Democrats, especially Democratic women.*

2.3 Empirical Approach

2.3.1 Survey Experiment

I conducted a survey experiment that provided information about incumbent performance to respondents and examined how that information impacted (a) evaluations of politicians and (b) causal attributions of responsibility for male and female politicians. I provided positive, negative, and neutral performance information about a male and a female officeholder. The information provided pertains to a hypothetical governor's[6] performance in a relatively gender-neutral policy area, the economy. The economy is an issue area that does not have specific gender connotations; Dolan (2010) showed a slight inclination among the public to view women as better able to handle economic matters than men. This is important because gendered policy areas

[5] In order to avoid priming respondents to think in gendered or partisan terms, I ask about respondent gender and party affiliation, after the experimental conditions.

[6] I provide information about the performance of men and women in executive rather than legislative roles because the attribution of responsibility and blame operates differently for executive and legislative office. It is easier to attribute performance to external factors for legislators than for executives, as legislators act as part of a larger body whereas executives more often act unilaterally.

might raise or lower expectations for women's performance and ability. Expectations of women have been shown to be lower in traditionally masculine policy areas, and higher in traditionally feminine contexts (McClean et al., 2018). Both men and women may also be evaluated more favorably in gender-congruent policy contexts (Nieva and Gutek, 1980). For women, a feminine gender cue can alleviate the negative reaction stemming from the incongruity between gender and leadership stereotypes (Heilman and Okimoto, 2007; Martin, 2016; Rudman and Glick, 2001).

The economic performance information provided[7] and hypothesized expectations for each condition are summarized in Table 1. The information vignettes follow previous studies (i.e., Huddy and Terkildsen (1993b,a)) in providing brief information about hypothetical officeholders. First, I provided a brief primer: "Imagine that Governor Susan/John Garrett is the governor of an American state. We are going to provide you with some brief information about that state's economy."

Table 1: Positive, Negative, & Neutral Performance Information

	Positive	Negative	Neutral
Female Governor	Since Governor Susan Garrett took office, unemployment rates have dropped, and wage growth is strong. Under her leadership, the economy has also grown at higher rates than expected, and new residential sales and construction are both up.	Since Governor Susan Garrett took office, unemployment rates have risen, and wage growth is weak. Under her leadership, the economy has also contracted, and new residential sales and construction are both down.	Since Governor Susan Garrett took office, unemployment rates and wage growth remains stable. Under her leadership, the economy, residential sales, and new construction are also steady.
Male Governor	Since Governor John Garrett took office, unemployment rates have dropped, and wage growth is strong. Under his leadership, the economy has also grown at higher rates than expected, and new residential sales and construction are both up.	Since Governor John Garrett took office, unemployment rates have risen, and wage growth is weak. Under his leadership, the economy has also contracted, and new residential sales and construction are both down.	Since Governor John Garrett took office, unemployment rates and wage growth remains stable. Under his leadership, the economy, residential sales, and new construction are also steady.

2.3.2 Outcome Measurement

The two dependent variables of interest are how respondents evaluate performance and how they assign responsibility for performance to officeholders. To measure performance evaluation, I adapt several questions from the ANES. I ask respondents:

[7]The economic indicators included are taken from the US Census Bureau's Economic Indicators (U.S. Census Bureau, 2018).

> Do you approve or disapprove of the way the Governor is handling [his/her] job as Governor? [7-point scale from Strongly approve to Strongly disapprove]
>
> Do you approve or disapprove of the way the Governor is handling the economy? [7-point scale from Strongly approve to Strongly disapprove][8]

To measure whether respondents attribute success and failure to internal or external factors, I follow methods employed by (Heilman and Haynes, 2005, 907) in using both a competence measure and an influence measure. According to attribution theory, performance that is attributed to skill and ability rather than to luck or effort is seen as repeatable (Weiner and Rosenbaum, 1971). Thus, in order to measure whether respondents attribute causality for performance to luck or skill, I ask respondents:

> "How well do you think the Governor will do on the economy in the future?" (Very well; well; Not very well; Not well; Don't know).

Importantly, I expect that there will be more uncertainty surrounding women's future positive performance, indicated by a greater portion of "Don't know" answers in the positive female governor condition. I assign this future performance variable scores ranging from 0 (not well at all) to 5 (extremely well), and then rescale the variable to range from 0-1, with 1 indicating the expectation that the governor will perform very well in the future. I create a separate variable to indicate respondents who answered "Don't know".[9]

For the influence measure, I ask two questions about the extent to which respondents attribute outcomes to the governor compared to other factors. From these two questions, I construct an index to create a single measure capturing internal or external attribution of responsibility, following the criteria in Table 2. I combine and average the two variables to create an internal-external influence measure ranging from 0-1, with 1 attributing the most influence to internal factors and 0 attributing the most influence to external factors.

Table 2: Influence Measure

Question	0	1	2	3	4	5	6
To what extent do you think the Governor was influential in these economic outcomes?	Not at all influential			Mid-point			Very influential
To what extent do you think other factors, such as advisers and state legislators, were influential in these economic outcomes?	Very influential			Mid-point			Not at all influential

[8] Questions adapted from ANES questions on presidential performance, page 148-9 of the Time Series codebook.

[9] These respondents are excluded from the main analysis of this variable, and discussed further in Appendix section 2.6.2.

2.4 Analysis and Results

The survey experiment was conducted on Qualtrics via the Lucid recruitment platform in May-June 2019.[10] A total of 1,748 respondents were recruited. The sample was 54% female, 39% Democrat, 32% Independent or something else, and 28% Republican, with an average age of 42 and an age range of 18-84. Table 3 depicts the average outcomes for John and Susan for each condition and dependent variable, the changes from neutral to positive or negative, and denotes whether the change is statistically significant. Overall, the positive and negative performance treatments impact evaluations in the expected direction: positive information raises approval and future performance expectations, while negative information damages these perceptions.[11] It is also worth noting that respondents credit the governor with significantly more influence in the positive condition and significantly less in the negative, only when the governor is male.

Table 3: Mean Values by Condition, All Dependent Variables

		Approval	Change	Economic Approval	Change	Future Performance	Change	Influence	Change
John	Neutral	0.591		0.589		0.534		0.458	
	Positive	0.691	+0.100***	0.689	+0.100***	0.628	+0.094***	0.476	+0.018**
	Negative	0.321	-0.270***	0.329	-0.260***	0.288	-0.246***	0.474	+0.016*
Susan	Neutral	0.612		0.593		0.537		0.472	
	Positive	0.678	+0.066***	0.688	+0.095***	0.628	+0.091***	0.476	+0.004
	Negative	0.303	-0.309***	0.311	-0.282***	0.308	-0.229***	0.461	-0.011

All tests reference comparison to neutral.

*** p<0.01, ** p<0.05, * p<0.10

I use two primary methods to test the hypotheses: difference of means tests and differences-in-differences regressions. The differences of means approach tests the null hypothesis that $\mu_{Male} = \mu_{Female}$ in the subsample of respondents in the neutral male condition and the subsample of respondents in the neutral female condition, where

$$\mu_{Male} = E[Approval \mid Male_{Neutral}] \text{ and } \mu_{Female} = E[Approval \mid Female_{Neutral}]$$

To perform differences-in-differences analyses, I use the model denoted in 2.1, substituting each dependent

[10] Lucid aggregates survey respondents from a variety of survey firms, using basic demographic information to recruit representative samples (Coppock and McClellan, 2019). Researchers set the rate of pay and then various firms can agree to field the survey or not and pay respondents the agreed upon rate. I set the rate of pay at $1.00 per respondent.

[11] I also performed a series of tests to determine whether the difference between the negative and neutral condition is significantly different from the difference between the positive and neutral condition, to ensure that the manipulations were operating as intended. I did this in part out of a concern that the use of 'stable' and 'steady' to describe wage growth, unemployment rates, and new construction could signal positive rather than neutral performance. These tests reveal that the shift from neutral to positive is significantly different than the shift from neutral to negative, for approval, economic approval, and future performance expectations. Moreover, the effects of positive performance are significant and positive, while the effects of negative performance are significant and negative. Thus, even if the neutral treatment is more positive than intended, the difference between the positive and neutral treatments are sufficient for the purposes of this analysis.

variable. I then test whether $\beta_1 = \beta_3 - \beta_2$, as β_1 yields the difference between neutral and positive male performance and β_3 - β_2 yields the difference between neutral and positive female performance. To test H1b, I test $\beta_4 = \beta_5 - \beta_2$, as β_4 yields the difference between neutral and negative male performance and β_5 - β_2 yields the difference in approval between neutral and negative female performance.

$$Y_{\text{Approval}} = \beta_0 + \beta_1 \text{Positive Male} + \beta_2 \text{Neutral Female} + \beta_3 \text{Positive Female} + \beta_4 \text{Negative Male} + \beta_5 \text{Negative Female} \tag{2.1}$$

This approach enables me recover the difference in each dependent variable between the subsample of respondents in the positive male condition relative to the subsample of respondents in the neutral male condition, and compare this difference to the difference in approval between the subsample of respondents in the positive female condition relative to the subsample of respondents in the neutral female condition. Essentially, these tests allow me to test whether the evaluative reward (penalty) for positive (negative) performance relative to neutral performance differs between men and women. Table 4 summarizes differences of means tests in the neutral condition for each dependent variable, while Table 5 presents the regression results used for the differences-in-differences tests.

Table 4: Differences of Means Tests, Neutral Performance

	Mean, Susan	Mean, John	Pr(T > t)
Approval	0.612	0.590	0.137
Economic Approval	0.592	0.589	0.423
Future Performance	0.537	0.533	0.429
Influence	0.471	0.458	0.070*
N	293	290	

*** p<0.01, ** p<0.05, * p<0.10

In the neutral condition, neither approval of Susan's performance nor estimations of Susan's handling of the economy are significantly different than John's. This that men and women receive similar evaluations for neutral performance, meaning H1 is not supported. Moreover, contrary to expectations, Susan is actually attributed with significantly more influence than John in the neutral condition, and expectations of future performance are not significantly different between the two. Thus, H2 is also not supported.

Table 5: Main Regression Results

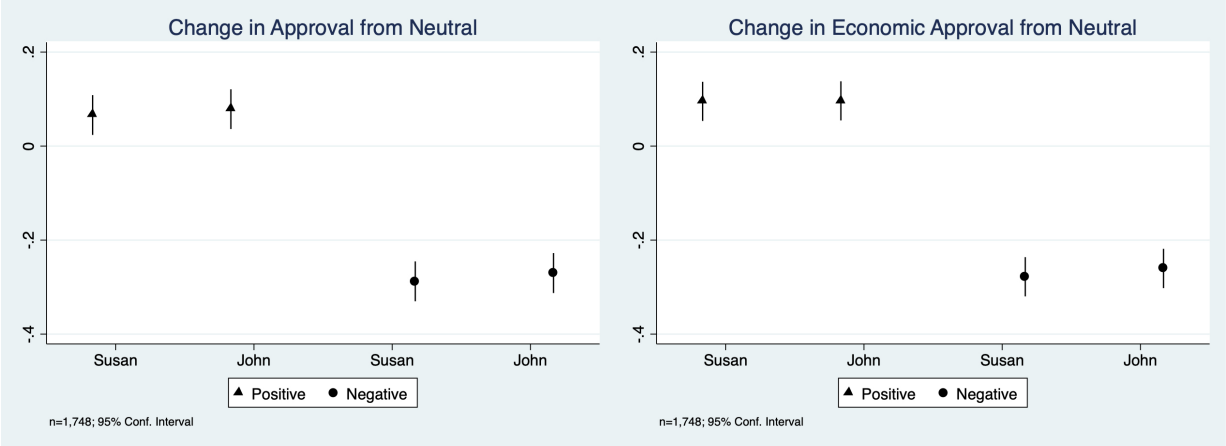
	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Susan Positive	0.087***	0.099***	0.094***	0.026***
	(0.022)	(0.021)	(0.021)	(0.009)
Susan Neutral	0.021	0.004	0.003	0.014
	(0.022)	(0.021)	(0.021)	(0.009)
Susan Negative	-0.288***	-0.278***	-0.226***	0.003
	(0.022)	(0.021)	(0.021)	(0.009)
John Positive	0.100***	0.100***	0.094***	0.018*
	(0.022)	(0.021)	(0.021)	(0.009)
John Negative	-0.270***	-0.260***	-0.246***	0.016*
	(0.022)	(0.021)	(0.021)	(0.009)
Constant	0.591***	0.589***	0.534***	0.458***
	(0.015)	(0.015)	(0.015)	(0.007)
Observations	1,748	1,748	1,565	1,748
R^2	0.274	0.270	0.258	0.006

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

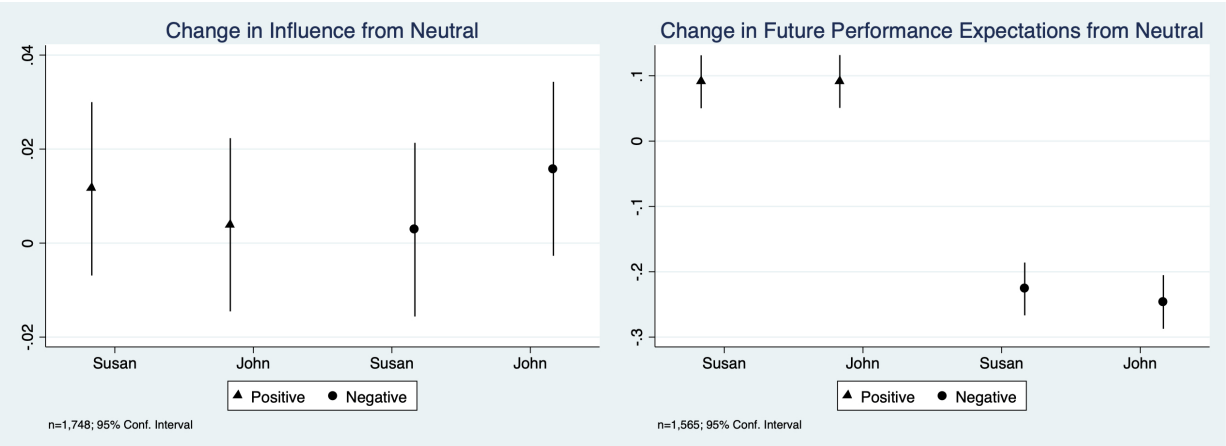
There is also limited support for H1a and H1b. While John does receive a slightly greater reward for positive performance and a lesser penalty for negative performance, the differences in differences analysis indicates that these differences are not statistically significant. The results of these tests are also depicted in Figure 4.1, which shows the changes in approval and economic approval for both John and Susan in the positive and negative conditions relative to the neutral condition.

Figure 2.1: Change in Approval and Economic Approval



The differences-in-differences analyses also indicate no significant differences in the amount of influence attributed to Susan and John for positive relative to neutral performance, as depicted in Figure 2.2. There are also no significant differences for any future performance estimations. Moreover, the differences in differences analysis indicates that Susan actually is attributed with significantly less influence than John in the negative performance condition ($Prob > F = 0.043$). This suggests that Susan is blamed significantly *less* for negative performance, relative to John, contrary to expectations. Thus, H2a and H2b are not supported.

Figure 2.2: Change in Influence and Future Performance Expectations



2.4.1 Gender Affinity

H3 posits a gender affinity effect, whereby evaluative penalties against women should be strongest among men and weakest among women, while rewards should be weakest among men and strongest among women. To test this, I performed the same series of differences of means and differences-and-differences analyses among men and women. Table 6 summarizes mean values for men and women by treatment condition. Dif-

ference columns indicate differences between men's and women's evaluations, while difference rows indicate differences in evaluations of John and Susan for men and women separately. Counter to expectations, there are no significant differences in the ways men and women evaluate John's performance compared to Susan's in the neutral condition.

Table 6: Mean Values by Condition and Respondent Gender

		Approval			Economic Approval			Future Performance			Influence		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Women	Men	Diff	Women	Men	Diff	Women	Men	Diff	Women	Men	Diff
Neutral	John	0.55	0.64	0.09***	0.55	0.63	0.08***	0.51	0.56	0.05**	0.45	0.46	0.01
	Susan	0.58	0.65	0.07***	0.56	0.64	0.08***	0.52	0.56	0.04**	0.47	0.47	0.00
	Diff	0.03	0.01		0.01	0.01		0.01	0.00		0.02	0.01	
Positive	John	0.69	0.69	0.00	0.68	0.70	0.02	0.62	0.64	0.02	0.48	0.48	0.00
	Susan	0.70	0.65	0.05*	0.70	0.67	0.03	0.63	0.62	0.01	0.49	0.48	0.01
	Diff	0.01	0.04		0.02	0.03		0.01	0.02		0.01	0.00	
Negative	John	0.29	0.36	0.07**	0.31	0.36	0.05*	0.25	0.34	0.09***	0.47	0.47	0.00
	Susan	0.27	0.34	0.07**	0.28	0.35	0.07**	0.26	0.36	0.10***	0.45	0.48	0.03**
	Diff	0.02	0.02		0.03	0.02		0.01	0.02		0.02**	0.01	

Column difference tests reference comparison between men and women; row difference tests reference comparison between male and female governor within groups.

*** p<0.01, ** p<0.05, * p<0.10

The results of the differences-in-differences regressions are summarized in Table 7 for men and Table 8 for women. Among men, there are no significant differences in the rewards and penalties Susan and John experience for any dependent variable. Among women, John and Susan receive similar shifts in terms of approval, economic approval, and future performance expectations for both positive and negative performance. But when it comes to influence, the picture changes: John is attributed with significantly more influence in the negative condition than in the neutral condition, and this effect is significantly different from the effect for Susan. Thus, the differences-in-differences analysis indicates that women are significantly less likely to blame Susan than John for negative performance ($Prob > F = 0.014$).[12]

[12] I also performed an additional analysis, testing whether the difference between the negative and positive conditions (rather than the neutral and positive conditions) is significantly different for John compared to Susan. Among women - but not men - the difference-in-difference is significant ($Prob > F = 0.0250$), providing further support for the notion that women blame the male governor significantly more for negative outcomes.

Table 7: Regression Results, Men

	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Susan Neutral	0.012	0.005	0.002	0.011
	(0.034)	(0.027)	(0.033)	(0.014)
John Positive	0.049	0.123***	0.077**	0.012
	(0.034)	(0.027)	(0.033)	(0.014)
Susan Positive	0.013	0.150***	0.059*	0.014
	(0.033)	(0.028)	(0.032)	(0.014)
John Negative	-0.280***	-0.248***	-0.220***	0.009
	(0.035)	(0.027)	(0.034)	(0.014)
Susan Negative	-0.296***	-0.277***	-0.203***	0.012
	(0.033)	(0.027)	(0.032)	(0.014)
Constant	0.641***	0.554***	0.562***	0.463***
	(0.024)	(0.019)	(0.023)	(0.010)
Observations	778	970	709	778
R^2	0.218	0.322	0.180	0.002

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 8: Regression Results, Women

	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Susan Neutral	0.033	0.006	0.009	0.017
	(0.027)	(0.033)	(0.026)	(0.013)
John Positive	0.144***	0.075**	0.112***	0.023*
	(0.027)	(0.034)	(0.026)	(0.013)
Susan Positive	0.152***	0.040	0.127***	0.036***
	(0.028)	(0.033)	(0.027)	(0.013)
John Negative	-0.254***	-0.268***	-0.258***	0.021*
	(0.027)	(0.034)	(0.026)	(0.012)
Susan Negative	-0.281***	-0.280***	-0.247***	-0.005
	(0.028)	(0.033)	(0.026)	(0.013)
Constant	0.548***	0.629***	0.508***	0.453***
	(0.020)	(0.023)	(0.019)	(0.009)
Observations	970	778	856	970
R^2	0.335	0.223	0.340	0.015

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

This exercise does not indicate whether the effect sizes are significantly different *between* men and women. In other words, it does not indicate for example whether the positive boost in approval John receives among men is significantly different than the boost he receives among women. To test whether these differences are significant, I performed an additional differences-in-differences analysis, comparing the changes from neutral to positive and negative for Susan and for John, between men and women. This analysis asks whether the rewards Susan and John receive among men are significantly different than the rewards they receive among women.

I performed the regressions denoted in models 2.2 and 2.3 for each dependent variable. I then tested $\beta_2 = \beta_3 - \beta_1$ (model 2.2) to determine whether Susan's reward among men was significantly different than her reward among women, and $\beta_4 = \beta_5 - \beta_1$ to determine if her penalty differed significantly between men and women. I also tested $\beta_8 = \beta_9 - \beta_1$ (model 2.3) to test the difference between John's reward among men and among women, and $\beta_{10} = \beta_{11} - \beta_1$ to test the difference in his penalty.

$$\mathrm{Y}_{\text{Approval}} = \beta_0 + \beta_1 \text{Susan Neutral Women} + \beta_2 \text{Susan Positive Men} + \beta_3 \text{Susan Positive Women} + \beta_4 \text{Susan Negative Men} + \beta_5 \text{Susan Negative Women} + \beta_6 \text{John Neutral Men} + \beta_7 \text{John Neutral Women} + \beta_8 \text{John Positive Men} + \beta_9 \text{John Positive Women} + \beta_{10} \text{John Negative Men} + \beta_{11} \text{John Negative Women} \tag{2.2}$$

$$\mathrm{Y}_{\text{Approval}} = \beta_0 + \beta_1 \text{John Neutral Women} + \beta_2 \text{Susan Positive Men} + \beta_3 \text{Susan Positive Women} + \beta_4 \text{Susan Negative Men} + \beta_5 \text{Susan Negative Women} + \beta_6 \text{Susan Neutral Men} + \beta_7 \text{Susan Neutral Women} + \beta_8 \text{John Positive Men} + \beta_9 \text{John Positive Women} + \beta_{10} \text{John Negative Men} + \beta_{11} \text{John Negative Women} \tag{2.3}$$

In terms of both approval and economic approval, this analysis indicates that the reward that Susan receives among men is significantly smaller than the reward she receives among women ($Prob > F = 0.006$, $Prob > F = 0.008$ respectively). However, the differences-in-differences analysis also indicates that the reward John receives is significantly greater among women than it is among women in terms of approval ($Prob > F = 0.029$), but not economic approval. Both Susan and John experience similar penalties among men and women. Moreover, neither Susan nor John experiences a significantly different change in evaluation among men or women in terms of future performance expectations or influence. Interestingly, this analysis indicates that the influence attributed to Susan in the negative condition does not differ significantly between men and women, although the earlier analysis showed that women were significantly more likely to blame John for negative outcomes. In other words, this means that women blame John significantly more than they blame Susan, but men and women blame Susan and John at roughly the same rate.

2.4.2 Partisanship

H3a posits that evaluative penalties against women should be strongest among Republicans, particularly Republican men, and weakest among Democrats, particularly Democratic women; conversely, rewards should strongest among Democrats, especially Democratic women, and weakest among Republicans, especially Republican men. To test this, I performed the same series of differences of means and differences-and-differences analyses among Democrats and Republicans, and then among as Democratic and Republican men and women. The results by party are summarized in Table 9.

The only significant result from the differences-in-differences analyses is that Democrats are significantly more likely to blame John compared to Susan for negative performance, as depicted by Figure 2.3. Moreover, difference of means tests indicate that there is no significant difference in the way either Democrats or Republicans evaluate John compared to Susan in the neutral condition across any of the outcome variables *except* for influence. Democrats attribute significantly more influence to Susan in the neutral condition.

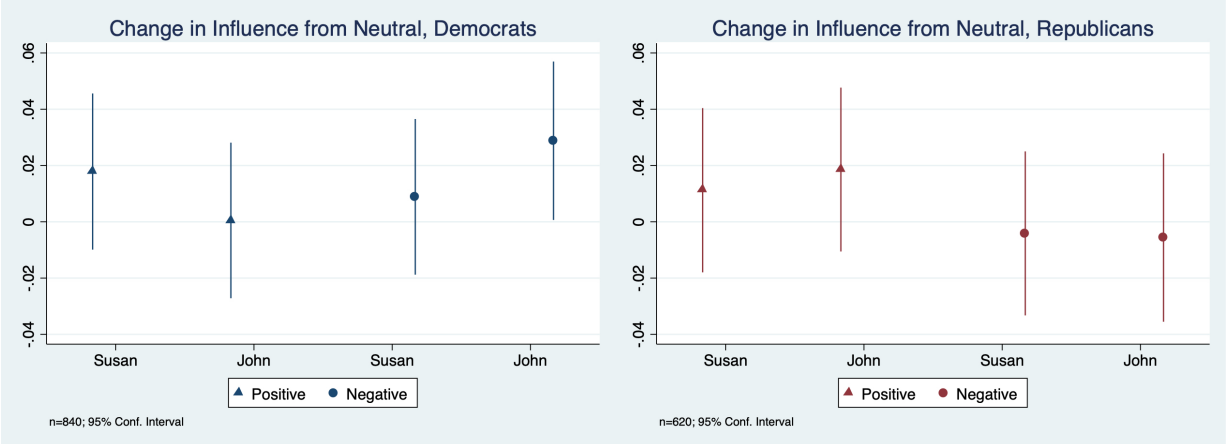
Table 9: Treatment Effects by Party

	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Democrats				
John Positive	0.068**	0.073**	0.068**	0.021
Susan Positive	0.078**	0.075**	0.066**	0.038***
Difference in Effect Size	0.010	0.002	0.002	0.017
John Negative	-0.305***	-0.299***	-0.255***	0.029**
Susan Negative	-0.300***	-0.279***	-0.202***	0.009
Difference in Effect Size	0.005	0.020	0.053	0.020*
Republicans				
John Positive	0.137***	0.151***	0.115***	0.010
Susan Positive	0.102***	0.156***	0.131***	0.003
Difference in Effect Size	0.035	0.005	0.016	0.007
John Negative	-0.291***	-0.255***	-0.273***	-0.006
Susan Negative	-0.314***	-0.302***	-0.273***	-0.004
Difference in Effect Size	0.023	0.047	0.000	0.002

Tests reference comparison to neutral for positive and negative conditions, and effect tests for *Difference in Effect Size*.

*** p<0.01, ** p<0.05, * p<0.10

Figure 2.3: Change in Influence by Party



The regression analyses by both party and gender are summarized in Tables 10 and 11. The only significant difference in effect sizes between John and Susan is among Democratic women: Susan is attributed with significantly less influence than John for negative performance ($Prob > F = 0.088$). Moreover, Democratic women credit Susan but not John with significantly more influence in the positive performance condition

relative to the neutral condition, although the difference in effect sizes is not significant. Finally, counter to H3a, Republican men show no obvious evaluative bias toward Susan.

Table 10: Treatment Effects by Party, Women

	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Democrats				
John Positive	0.101**	0.078**	0.094**	0.017
Susan Positive	0.132***	0.111***	0.098***	0.053***
John Negative	-0.281***	-0.291***	-0.259***	0.026
Susan Negative	-0.312***	-0.296***	-0.213***	0.007
N	488	488	440	488
Republicans				
John Positive	0.198***	0.196***	0.125***	0.019
Susan Positive	0.157***	0.206***	0.163***	0.002
John Negative	-0.291***	-0.239***	-0.283***	-0.001
Susan Negative	-0.288***	-0.282***	-0.318***	-0.012
N	309	309	274	309

*** p<0.01, ** p<0.05, * p<0.10

Tests reference comparison to neutral condition.

2.4.3 Exploratory Analysis: Gender Attitudes

The analyses above suggest that Democratic women punish the female governor significantly less for poor performance, while Democratic men, Republican men, and Republican women reward and punish both the male and female governor at similar rates. There are several reasons why Democratic women may evaluate female politicians more generously than male politicians. Democrats, particularly women, tend to be more supportive of female candidates in general. Democratic and Republican women also tend to express very different attitudes towards women in political roles, and towards gender equality and feminism more broadly. Democratic women tend to care more about gender equality in government and the election of women. Because the survey experiment does not provide information on the governor's partisanship, Democratic women may also assume the female governor is a Democrat (Roberts and Utych, 2020; Clifford, 2020; Sanbonmatsu and Dolan, 2009).

In order to examine whether attitudinal differences can help explain this result, I delve into an additional

Table 11: Treatment Effects by Party, Men

	(1) Approval	(2) Economic Approval	(3) Future Performance	(4) Influence
Democrats				
John Positive	0.030	0.071	0.043	0.027
Susan Positive	0.011	0.032	0.035	0.019
John Negative	-0.326***	-0.295***	-0.226***	0.037
Susan Negative	-0.273***	-0.246***	-0.173***	0.013
N	352	352	329	352
Republicans				
John Positive	0.076	0.106**	0.105*	0.000
Susan Positive	0.047	0.107**	0.103*	0.002
John Negative	-0.285***	-0.269***	-0.256***	-0.010
Susan Negative	-0.344***	-0.326***	-0.241***	0.002
N	311	311	283	311

*** p<0.01, ** p<0.05, * p<0.10

Tests reference comparison to neutral condition.

battery of questions included in the original survey intended to measure traditional gender attitudes. Specifically, I drew several questions from the Ambivalent Sexism Inventory, first developed by Glick and Fiske (1996).[13] I chose not to include these measures in the main analysis because their inclusion did not add any substantive effects. However, exploring gender attitudes across gender and partisanship may help glean some insight as to what might be causing the evaluative differences I observe.

Specifically, I asked respondents for their degree of agreement with four statements, listed below. I then use respondents' answers to these questions to create the variables *Exaggerate Problems*, *Too Easily Offended*, *Cherished & Protected*, and *Superior Moral Sensibility*. In Table 12, I summarize the mean values for these variables for men and women in each party. I also perform a series of differences of means tests comparing Democratic women to the rest of the respondents.

> How strongly do you agree or disagree with the following statements? *(0 = Strongly disagree; 6 = Strongly agree.)*
>
> Women tend to exaggerate problems at work.
> Women are too easily offended.
> Women should be cherished and protected by men.
> Women have a superior moral sensibility.

[13] See Appendix section 2.6.1 for additional discussion of the Ambivalent Sexism Inventory and my selection criteria for questions.

Several patterns are noteworthy. Compared to all other respondents, Democratic women express significantly lower levels of agreement with the two statements meant to capture overt, hostile sexism: *Exaggerate Problems* and *Too Easily Offended*. Democratic women also express significantly lower levels of agreement with *Cherished & Protected*, but significantly higher levels of agreement with *Superior Moral Sensibility*.

Table 12: Gender Attitude Questions by Party & Gender (Mean Values)

	(1) Exaggerate Problems	(2) Too Easily Offended	(3) Cherished & Protected	(4) Superior Moral Sensibility
Democratic Men	0.337	0.385	0.708	0.530
Democratic Women	0.239	0.280	0.610	0.564
Republican Men	0.513	0.596	0.739	0.510
Republican Women	0.410	0.466	0.696	0.521

It may be the case that a stronger belief in women's superior moral sensibility makes Democratic women more likely than any other group to give the female governor the benefit of the doubt given negative performance information. These attitudes may also only work in one direction, or function differently across different groups. For instance, Republican men are essentially the mirror image of Democratic women, exhibiting significantly higher levels of agreement than any other group with *Exaggerate Problems*, *Too Easily Offended* and *Cherished & Protected*, and significantly lower levels of agreement with *Superior Moral Sensibility*. Yet, the analysis indicated no gender bias in Republican men's evaluations of the female governor's performance. This brief analysis cannot fully answer these questions, but may serve to provide some direction for future research.

2.5 Discussion

If men and women in political office are held to different standards, or assigned credit and blame in systematically different ways, this may differentially impact their ability to advance in organizational ranks, to affect policy outcomes, and their chances of reelection. Studies outside of political science have demonstrated that gender conditions performance evaluation and the attribution of responsibility. In this experiment, I aimed to test whether such patterns carry over to the political sphere. 1,748 respondents completed a survey experiment in which they were exposed to neutral, positive, or negative performance information about a hypothetical male or female governor, and then asked a series of questions to gauge both evaluation and assignment of responsibility for the governor's performance. I hypothesized that the female governor would receive greater evaluative penalties and more blame for negative performance, as well as weaker rewards and less credit for positive performance.

Several noteworthy results emerged from this study. First, the hypothesized effects of gender on performance evaluations were generally not borne out by the statistical analyses for the entire group of respondents. In fact, contrary to expectations, the only statistically significant difference between the male and female governor was that the female governor was actually blamed significantly *less* for negative performance. However, it is worth nothing that across the majority of tests, the male governor receives a greater reward for positive performance and never receives a lesser reward. Although these differences were not statistically significant, the pattern is consistent across nearly every dependent variable, particularly for positive performance, though effect sizes are quite small.

Next, the results of this study point to several interesting heterogeneous effects by gender and party. First, women and Democrats blamed the female governor significantly less than the male governor for negative performance. Contrary to expectations, men and Republicans rewarded and punished the male and female governor more or less equally, and Republican men did not exhibit any particular gender bias against the female governor. Further analyses revealed that Democratic women were the group driving the evaluative differences in blame. A brief exploration of Democratic women's answers to a series of questions measuring gender role attitudes suggests these underlying attitudes may have something to do with this result. Previous research also suggests Democratic women care about gender equality in government, have more positive attitudes about feminism, and support female candidates at relatively high rates. Moreover, particularly given the salience of gender in the recent political climate, it is possible that Democratic women feel a sense of protection towards or solidarity with women in political office.[14] This is an avenue for future research to explore further.

This analysis also indicates that men and women reward and punish political performance in systematically different ways. Compared to men, women show a significantly higher proclivity to reward positive performance by the governor, while men and women penalize the governor for poor performance at similar rates. Why might women reward positive performance more than men, and does this translate across different policy issues? This unexpected finding prompts a variety of future research questions, and suggests there may be additional differences in how men and women assign political credit and blame.

These results may indeed indicate that gendered performance evaluation is not an underlying cause of the gendered performance premium in American politics, nor of gender disparities in elective office more broadly. However, this experiment tested these questions in a gender-neutral policy area. It is possible that variation in policy context may shift the way voters evaluate performance. Policy context may raise or lower the salience of gender-stereotypic strengths and issue areas, thereby influencing the way gender

[14]Recent work indicates raising the salience of gender increases levels of mistrust between Democratic and Republican women (Klar, 2018). Democratic and Republican women are also more likely to evaluate female politicians differently when a specifically gendered issue is made salient (Anderson-Nilsson and Clayton, 2021).

stereotypes affect evaluations (Kahn, 1996; Bauer, 2015a; Holman et al., 2016). Because voters are likely to view women as better-suited to leadership roles that align with the strengths and issue areas stereotypically associated with women, a more feminine policy context may alleviate the incongruity between feminine stereotypes and leadership roles. For instance, where the major themes of a campaign context align with women's stereotypical strengths and issue areas, women candidates tend to be viewed more positively by voters and perform significantly better electorally(Kahn, 1993, 1996; Fox and Oxley, 2003b; Holman et al., 2016). The reverse is true for masculine contexts, in which men tend to perform better (Kahn, 1996; Holman et al., 2016). It is also possible that results would be different for legislative rather than executive, as executive offices tend to be associated with masculine stereotypes. These are questions for future research.

It should also be noted that this experiment included hypothetical American governors without any party affiliation information. Partisanship and policy area may have interactive effects on evaluations of male or female leaders. For example, Holman et al. (2016, 2017) examined preferences for male or female leadership under conditions of terrorist threat and found a strong preference for male, Republican leaders within the context of terrorist threat, followed by female Republicans, then male Democrats, with female Democrats being the least preferred. Partisanship is also consistently the most important factor in American politics in terms of predicting electoral support, and can strongly bias reasoning. Moreover, past work has uncovered clear differences in the way Republican and Democratic voters evaluate and support female candidates (see e.g. Burden and Ono, 2018; Knuckey, 2005; Sanbonmatsu and Dolan, 2009; King and Matland, 2003). For these reasons, future work might also make incumbents' partisanship explicit in order to fully examine how party conditions the effects of gender on performance evaluation and the attribution of political responsibility. Finally, it is possible that the use of hypothetical governors rather than real officeholders does not adequately capture the way stereotypes and biases operate in real elections. Although the use of real officeholders introduces a variety of potentially confounding factors, a growing number of studies call into question whether survey experiments using hypothetical candidates can replicate real world results (see e.g. Krupnikov et al., 2016; Streb et al., 2008; Clayton et al., 2020). Thus, future research may also benefit by studying the ways voters assign political responsibility to real men and women in political office.

2.6 Appendix

2.6.1 Ambivalent Sexism Inventory Questions

The Ambivalent Sexism Inventory has been shown to capture both hostile and benevolent sexist attitudes and to perform as well as or better than similar measures, such as the Modern Sexism Scale (Conn et al., 1999). The ASI is constructed by asking respondents the degree to which they agree with a series of statements. Eleven of these statements are designed to capture hostile sexism, and eleven aim to capture benevolent sexism. This includes attitudes such as protective paternalism, complementary gender differentiation, and heterosexual intimacy. I choose four questions from the ASI (two from the hostile and two from the benevolent category) and use these to measure respondents' gender attitudes in my survey. I select these questions for two reasons. First, three of the four questions load most heavily in the the factor loadings analysis for the ASI. The question that does not load as heavily, "Men should sacrifice to provide for women[,]" captures traditional gender role attitudes better than the other questions (Glick and Fiske, 1996, 500). I choose two questions from the benevolent sexism category and two from the hostile sexism category because these capture different aspects of gender role attitudes. The first two capture more directly sexist attitudes, and most respondents are not likely to answer that they agree with these statements, whether due to social desirability bias or to true disagreement. However, respondents who do not express explicitly sexist attitudes may still hold implicit, gender-based biases (Mo, 2014). Benevolent sexism encapsulates attitudes that are sexist in the sense that they view women "stereotypically and in restricted roles, but that are subjectively positive in feeling tone" (Glick and Fiske, 1996, 491). Thus, the questions drawn from the benevolent sexism category aim to capture this more subtle form of sexism. Moreover, posing the first two hostile-category statements to respondents before the second two may give respondents an 'out,' so to speak, as they may satisfy social desirability concerns by expressing nonsexist attitudes, thus mitigating social desirability bias concerns in answering the second two questions.

2.6.2 "Don't Know" Future Performance Answers

I initially expected that respondents would be more likely to express uncertainty about the female governor's future performance. I examined the proportion of respondents who answered "don't know" when asked about future performance expectations for the male and female governors in each condition, summarized in Table 13. In the neutral and positive conditions, virtually the same proportion of respondents expressed uncertainty about the governor's future performance regardless of governor gender. In the negative condition, a larger proportion of respondents indicate uncertainty about the governor's future performance. However, respondents are no more or less likely to express uncertainty about either governor's future performance when given positive or negative information.

It is also interesting that women were more likely than men to express uncertainty about the male governor's future performance, while for the female governor men and women were equally likely to express uncertainty in all but the negative condition. While I cannot fully explain these patterns with the data here, it may be that gender affinity is also at play here. Men are less likely than women to express uncertainty about the male governor's future performance prospects, while women express somewhat more uncertainty about the male governor than the female governor.

Table 13: Number of Respondents Answering "Don't Know" per Condition

Governor Gender	Neutral Condition	Positive Condition	Negative Condition
Female Governor	30	31	23
	(16 women; 14 men)	*(15 women, 16 men)*	*(15 women; 8 men)*
Male Governor	32	29	38
	(23 women, 9 men)	*(21 women, 8 men)*	*(24 women; 14 men)*

CHAPTER 3

Part II: Gendered Returns on Electoral Behavior? How Legislative Activity Impacts House Members' Electoral Fortunes

Classical conceptions of democracy posit that voters should evaluate, reward, and punish their elected representatives according to the quality of representation they provide. In this rational conception, elections induce re-election motivated representatives to produce good outcomes on behalf of their constituencies (Mayhew, 2004; Ferejohn, 1986). Yet, a wide body of research shows that American voters are quite far from being "rational god[s] of vengeance and reward" (Key, 1964, 568). Not only does the public often lack the information necessary to hold representatives accountable for their performance, voters also tend to use the information they do have in ways that contradict classical notions of democratic accountability. On average, voters do not evaluate political representatives or political realities objectively: instead, partisan loyalties, group identities, and a profusion of reasoning biases cloud the connection between political representation and voting behavior (Achen and Bartels, 2015).

An underexplored source of bias in political accountability is representative gender. Myriad studies in management, education, and sociology indicate that men and women who occupy similar roles and perform equally well are evaluated and rewarded in systematically different ways. Many women in political office report the belief that they are held to different performance standards than their male colleagues (Lazarus and Steigerwalt, 2018). As presidential contender Amy Klobuchar put it in November of 2019, "[w]omen are held to a higher standard... We have to work harder, and that's a fact" (Astor, 2019). Public opinion surveys show the public tends to agree, with majorities of Americans reporting the belief that women in politics are held to higher standards (Pew Research Center, 2018; Ballard, 2018). Yet, very little work in political science has asked directly whether gender does indeed condition the democratic connection.

In this paper, I investigate whether men and women experience different consequences for their electoral behavior. Specifically, I ask whether legislative activity impacts subsequent electoral performance differently for men and women overall, as well as within the two major political parties. I examine roll call records, bill sponsorship patterns, and electoral returns for members of the U.S. House of Representatives from 1973-2017 to compare whether the impact of legislative activity is different for male and female members. I examine two main aspects of legislative activity: bill sponsorship and district policy congruence. Overall, I find that gender does condition the impact of legislative behavior on electoral support. Although women sponsor significantly more bills and more closely represent their districts' policy preferences in their roll-call voting behavior, these behaviors are not associated with any significant electoral payoff. Conversely, policy congruence has

significant electoral consequences for men, and bill sponsorship incurs electoral rewards for Republican men. I further find that partisanship and gender interact to condition the electoral rewards and penalties men and women experience.

Taken together, these results suggest that equivalent electoral behavior by male and female legislators translates to different electoral returns within the two parties. My findings have several broad implications. First, gendered patterns of political reward and punishment can help explain why men and women tend to perform equally well in elections even though women's performance records are consistently stronger than those of their male colleagues (Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018; Thomsen and Sanders, 2020). Second, these findings can also help us to better understand continued gender disparities in American government, both in terms of raw numbers and in terms of women's access to higher and more powerful offices. Finally, these results have implications for the quality of American democracy. If voters do not hold representatives equally accountable for similar performance in office, it is likely elections do not select the best representatives, corroding the quality of democratic representation.

3.1 Democratic Accountability

Rational models of democratic accountability posit that citizens exert democratic control by electing and removing public leaders such that the actions of elected officials are more or less consistent with the preferences and best interest of their constituents. Within this tradition, models of voter behavior posit that voters respond make voting decisions on the basis of incumbent performance and policy preferences (i.e., Key, 1964; Kramer, 1971; Stokes, 1963; Fiorina, 1981). Elections occasion political accountability by incentivizing representatives to produce good outcomes on behalf of their constituencies, because "if voters do not like what the members are doing, they can vote them out of office" (Mayhew, 2004; Ferejohn, 1986; Jacobson and Carson, 2020, 260). A critical assumption of this classical conception of democratic accountability is that citizens acquire relevant political information about their representatives' performance, and factor this information rationally into their voting choices, rewarding good - and punishing poor - political representation.

However, the political information citizens acquire and the ways they use it to make voting decisions is affected by a wide variety of factors beyond the performance of their representatives. Scholars have produced ample evidence of widely uninformed and often irrational, myopic, and fickle electorate - as likely to punish incumbents for shark attacks or disappointing college basketball games as for poor policy representation (Healy et al., 2010; Achen and Bartels, 2015). In particular, voters are likely to engage in motivated reasoning, acquiring and processing information in ways that comport with their predispositions (Slothuus and de Vreese, 2010; Druckman et al., 2013; Leeper and Slothuus, 2014). For instance, citizens who view a rep-

resentative favorably are likely to to ignore or deflect blame for negative information about her performance or character, and to give her the benefit of the doubt in the absence of information (Wilson and Gronke, 2000; Bisgaard, 2015). Partisan motivated reasoning has been particularly well-documented, with partisan identity strongly shaping voters' perceptions and interpretations of political realities in favor of their preferred party (Bartels, 2002; Leeper and Slothuus, 2014; Gaines et al., 2007; Bisgaard, 2015). In other words, the attribution of political responsibility tends to be socially constructed, which departs significantly from what the classical view of democratic accountability envisions (Achen and Bartels, 2015).

Yet, there is evidence that what members of Congress do in office does yield electoral dividends. One of the primary electorally relevant aspects of representation is representatives' legislative activity. Legislative activity encompasses members' roll-call voting behavior, the volume of bills introduced, sponsored, and co-sponsored, their legislative effectiveness, speeches on the house floor, and various other forms of policy entrepreneurship (Sulkin et al., 2015a). Perhaps the most electorally important aspect of legislative activity is policy congruence, or the extent to which a constituency's policy views are reflected in the policy stances of their elected officials, most often measured through their roll-call voting patterns (Jacobson and Carson, 2020). Particularly in Congressional elections, greater policy congruence is on average associated with greater electoral success, while higher levels of policy incongruity are associated with electoral penalties (Bernhard and Sulkin, 2018; Jacobson and Carson, 2020). There is substantial evidence that this is because voters evaluate how well representatives' policy preferences overlap with their own policy preferences, and factor these evaluations into their voting decisions[1] (Griffin et al., 2019; Jacobson and Carson, 2020; Jones, 2011; Ansolabehere and Jones, 2010; Nyhan et al., 2012). Thus, on average, "members who stray too far from home suffer at the polls" (Jacobson and Carson, 2020, 262). For example, members who are more ideologically distant from their districts tend to garner lower vote shares and are more likely to lose reelection (Canes-Wrone et al., 2002; Bernhard and Sulkin, 2018, 136). Members of Congress appear to recognize this dynamic and respond strategically: for example, U.S. House members whose records are more out of step with their constituency's preferences are more likely to avoid calling attention to their voting records in communications with the public compared to their more congruent colleagues (Taylor, 2017).

The evidence that other aspects of legislative activity impact representatives' electoral fortunes is more mixed. In terms of the overall volume of legislative activity, some studies find a positive impact of higher levels of bill sponsorship on vote shares (e.g., Box-Steffensmeier et al., 2003), while others uncover no electoral effects of bill sponsorship (e.g., Ragsdale and Cook, 1987). Scholars have argued that sponsoring

[1] Importantly, this dynamic does not require that all voters pay close attention to representatives' legislative activities. Voters are able to form relatively accurate impressions of representatives' voting patterns and policy positions *in the aggregate*. (Jacobson and Carson, 2020; Jones, 2011; Ansolabehere and Jones, 2010; Nyhan et al., 2012). Indeed, the impact of legislative activities on approval and voting behavior varies substantially across voters' levels of partisanship, information, political sophistication, and policy preferences (Sulkin et al., 2015b; Dancey and Sheagley, 2016; Hollibaugh et al., 2013; Guisinger, 2009).

a large number of bills may be electorally beneficial because it can draw media attention, improve name recognition, and generate positive impressions of members' policy expertise (Box-Steffensmeier et al., 2003; Parker and Goodman, 2009). Additionally, legislative effectiveness has been associated with greater electoral success, improved fundraising, and higher approval ratings (Miquel and Snyder, 2006; Sulkin, 2005; Box-Steffensmeier and Grant, 1999).[2] Members also experience evaluative consequences among voters for factors such as district attention, missed votes, and party loyalty (Sulkin et al., 2015b). Moreover, the direction and magnitude of these effects depends on citizens' co-partisanship: for instance, citizens represented by a co-partisan representative are more likely to react positively to high levels of party loyalty, while out-partisans evaluate party loyalty more negatively (Sulkin et al., 2015b).

Taken together, this body of evidence suggests citizens respond in rational ways, at least to some degree, to various aspects of their representatives' electoral behavior. But just as voters do not reward and punish electoral behavior equally across representatives of different parties, representatives' demographic characteristics also condition voter responses to electoral behavior. In elections, voters have been found to rely on gender and racial stereotypes to make assumptions about candidates' policy stances (Mcdermott, 1998), and there is considerable evidence that representatives' demographic traits can condition the way citizens acquire and process information about them (Tate, 2003; Lerman and Sadin, 2016). For instance, citizens who live in districts represented by black representatives acquire significantly more information about their representatives than citizens who are represented by white representatives (Griffin and Flavin, 2007). In a similar vein, constituents who are represented by female senators have significantly more knowledge about their senators' policy positions than female constituents with male senators (Jones, 2014). These findings suggest political accountability can function differently for representatives of different backgrounds.

3.1.1 Gender, Performance, and Candidate Evaluation

Several areas of past research suggest reasons to expect that political accountability may operate differently for men and women. First, a bevy of recent studies have have identified a 'performance premium' faced by female candidates (Fulton and Dhima, 2020b, 6). Studies in this area show that although men and women generally perform equally well in elections, female candidates are systematically higher-quality than male candidates (Fulton, 2011, 2014; Pearson and McGhee, 2013; Fulton and Dhima, 2020b). Moreover, this is at least in part because voters hold female candidates to higher qualification standards (Bauer, 2020c; Mo, 2014; Branton et al., 2018; Carey and Lizotte, 2019). Information that casts doubt on candidates' competence is also more harmful for women than for men (Ditonto, 2017), and quality challengers damage women's electoral

[2] For additional evidence of various forms of legislative activity and performance in office impacting electoral outcomes, see Berry and Howell (2007); Gasper and Reeves (2011); Healy and Malhotra (2009); Karol and Miguel (2007).

prospects more than those of similarly positioned men (Barnes et al., 2017).

Given that women face higher qualification hurdles as candidates, it follows that they may also face higher performance standards as officeholders. Indeed, studies investigating differences in how men and women perform their roles in Congress show that women outperform their male counterparts on a variety of dimensions, particularly in terms of bill sponsorship (Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018; Atkinson and Windett, 2019).[3] Women in both the House and the Senate are also less likely than men to deviate from their constituencies' ideological preferences in their roll-call voting behavior (Lazarus and Steigerwalt, 2018, 169). Considering men and women generally perform equally well in elections, it makes sense that the electoral payoffs for these activities may be different for men and women. Indeed, Atkinson and Windett (2019) show that women must sponsor twice as much legislation as men in order to reduce the probability of facing a quality challenger to the same level as men.

Another line of research has uncovered differences in how voters evaluate similar characteristics and behaviors of men and women in office. For example, compared to men, women experience greater penalties for engaging in corrupt behavior, for displaying power-seeking intentions, and for employing negative campaign strategies[4] (Eggers et al., 2018; Krupnikov and Bauer, 2014; Cassese and Holman, 2017; Okimoto and Brescoll, 2010). Hence, it stands to reason that voter evaluations of other electorally relevant behaviors may also be gendered. On the other hand, it should also be noted that not all behaviors and characteristics of politicians are evaluated in gendered ways, and the extent to which evaluations are gendered may vary across groups. For instance, Brooks (2011) shows that voters penalize male and female candidates similarly for displays of anger, while women punish female candidates far more than male candidates for crying.

An officeholder or candidate's gender can also condition the type and volume of information the public receives about her. This may condition political accountability because information about politicians is crucial to voters' ability to evaluate them. Thus, media coverage that focuses on a male politician's policy accomplishments while devoting more attention to a female politician's background makes it more difficult for voters to reward the female politician for her achievements in office, relative to the male. A wide array of studies examining media coverage of individual female candidates have found evidence of gender bias in both the volume and nature of media coverage, particularly for female presidential and gubernatorial contenders (Bystrom and Dimitrova, 2014; Fowler and Lawless, 2009).[5] For example, media coverage of elections that

[3]Women also direct more discretionary spending to their districts; bring home more federal outlays; and introduce, sponsor, and cosponsor more resolutions - even in ostensibly safe seats (Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018).

[4]It should also be noted that negative campaign ads aimed *at* candidates negatively impact evaluations of targeted female candidates less than targeted male candidates (Fridkin et al., 2009), suggesting that women experience evaluative penalties when they are the sources rather than the targets of negative campaigning.

[5]It should be noted that gender bias in media coverage is not a universal finding. For instance, Hayes and Lawless (2015) find no evidence of gender differences in media coverage of House elections, and Hayes et al. (2014) find no gender differences in the effects of media coverage of candidate appearance on voter evaluations.

include female candidates is more likely to focus on candidate traits rather than policy issues, compared to races that include only men (Dunaway et al., 2013). Along similar lines, although female Senate candidates in 2016 provided a significantly higher volume information about their qualifications, female candidates received significantly less news coverage of their political qualifications and significantly more coverage about their husbands, children, and family lives in the months directly before the election (Bauer, 2020b).

3.2 Gender and Political Accountability

In this paper, I ask: do men and women see different returns for their electoral behavior? I draw primarily from role congruity theory in developing my theoretical expectations. Role congruity theory proposes that stereotypically masculine traits (ambition; assertiveness; dominance) align more closely with leadership roles, while stereotypically feminine traits (compassion; warmth; kindness) align more closely with communal roles (Eagly and Karau, 2002). Women are often assumed to lack the agentic traits necessary to perform leadership roles effectively (Foschi, 2000). Moreover, the mismatch between feminine and leadership stereotypes is likely to generate evaluative penalties for women who occupy leadership roles, engage in stereotypically masculine behaviors, or display stereotypically masculine traits. This stems from a negative (often unconscious) reaction against counter-stereotypic behavior (Eagly and Karau, 2002; Koenig et al., 2011; Simon and Hoyt, 2008). Role incongruity can be exacerbated or alleviated by factors that heighten or reduce perceived lack of fit.[6] Hence, evaluative penalties tend to be most severe for women in highly stereotypically masculine areas and roles. Role incongruity can also be exacerbated when women perform successfully in leadership roles, because success in these roles implies a deficit in (stereotypically feminine) communal attributes, further exacerbating role incongruity (Heilman and Haynes, 2005; Heilman and Okimoto, 2007; Heilman, 2012). Put another way, elected women who perform their roles successfully may experience a particularly strong evaluative backlash among voters.

Even in the absence of a negative affective reaction, role incongruity can make it difficult for individuals to earn recognition for their performance in counter-stereotypical roles. Perceived lack of fit can generate negative performance expectations, which can in turn negatively impact assessments of women's performance in and suitability for leadership roles (Heilman, 2012; McClean et al., 2018; Ridgeway, 2001). In other words, even if voters do not penalize women for legislative behavior, they may be less likely to give women credit for their activities in political office. This discussion further suggests that information that casts doubt on an individual's suitability for a role is likely to be more detrimental where lack of fit is already high.[7] Thus, electorally detrimental factors may be more deleterious for women than for men because they can fuel

[6] For instance, leadership is even more likely to be construed as masculine with regard to higher-status roles and executive office (Koenig et al., 2011; Huddy and Terkildsen, 1993b,a)

[7] For instance, women who work in masculine domains are likely receive disproportionate blame for joint failures (Caleo, 2010).

beliefs about women's unsuitability for political leadership roles (Ditonto, 2017). In other words, in light of poor policy representation, voters may give male representatives the benefit of the doubt but punish female representatives. Overall, I hypothesize that men and women representatives are likely to earn differential electoral rewards for similar performance.

> **H1:** *The electoral impact of legislative productivity and policy congruence is more positive for male representatives than for female representatives.*

3.2.1 Gender, Partisanship, and Candidate Evaluation

Gender and partisan stereotypes are intertwined, with feminine traits, stereotypes, and language more closely associated with the Democratic party and liberal politics; and masculine stereotypes, traits, and language more closely associated with conservative ideology and the Republican party (Roberts and Utych, 2020; Clifford, 2020). Female politicians tend to be stereotyped as both more liberal than men and more liberal than they actually are (Sanbonmatsu and Dolan, 2009; King and Matland, 2003). Past studies suggest that representative gender and party affiliation interact to condition the ways voters form views and expectations about representatives' issue competencies and traits (Sanbonmatsu and Dolan, 2009; Schneider and Bos, 2016). For instance, voters tend to stereotype Republicans and men as better able to handle issues surrounding crime and security, putting Democratic women at a disadvantage in this policy area (Sanbonmatsu and Dolan, 2009; Holman et al., 2017). Conversely, Democrats and women are perceived as better able to handle education (Sanbonmatsu and Dolan, 2009), uniquely advantaging Democratic women on this issue. Party and gender stereotypes are especially likely to impact Democratic and Republican women differently under high levels of polarization. Democratic women have been shown to benefit from the stereotype that they are more liberal, as Democratic voters and donors alike tend to favor more liberal candidates, more so as polarization increases (Sanbonmatsu and Dolan, 2009; Kitchens and Swers, 2016; Thomsen and Swers, 2017). On the other hand, the perception that Republican women are more moderate may work to their disadvantage as they face increasingly conservative electorates (Kitchens and Swers, 2016).

There are also significant differences in gender role attitudes between Democratic and Republican voters. Republican voters tend to hold more traditional views on gender roles, particularly more conservative voters (Knuckey, 2005; Shames, 2018), while Democratic voters are more likely to care about gender equality in government and to hold more favorable stereotypes about women politicians (Sanbonmatsu and Dolan, 2009). Republican voters are significantly less likely than Democrats and Independents to describe female candidates as "strong leaders," and also evaluate policy arguments made by women more negatively than Democratic voters, even in the absence of explicit party cues (Anderson-Nilsson and Clayton, 2020; King and Matland, 2003, 602). Moreover, while Democrats and Independents often show a bias in favor of women

candidates in primary elections, Republicans exhibit a negative bias toward women when choosing between male and female primary candidates (Burden and Ono, 2018; King and Matland, 2003).

Role incongruity and its negative effects may be reduced or even eliminated for Democratic women, for two main reasons: (1) because Democratic voters tend to hold less traditional views on gender roles, and (2) because feminine stereotypes are closely associated with liberal and Democratic party stereotypes. Conversely, the effects of role incongruity are likely to be stronger for Republican women, (1) because Republican voters hold more traditional gender role attitudes and (2) because the association between feminine and liberal stereotypes may constitute a barrier rather than a boon for Republican women. For these reasons, I expect gender differences in the electoral consequences of legislative behavior and policy incongruity to be more (less) pronounced within the Republican (Democratic) party.

> **H2:** *Gender differences in the electoral impact of legislative productivity and policy congruence will be more pronounced within the Republican party, and less pronounced within the Democratic party.*

I further expect that ideological direction of policy incongruity will impact electoral performance for men and women differently within the two parties. I refer to representatives who vote more conservatively (liberally) than their districts as engaging in conservative (liberal) policy incongruity. Democratic voters are likely to have prescriptive or descriptive expectations that women's voting behavior should reflect a more liberal ideology. Because of the association between liberal and feminine stereotypes, these expectations should be stronger for Democratic women relative to Democratic men. Therefore, conservative policy incongruity by women is likely to violate these expectations and generate a backlash. For this reason, I expect to observe greater electoral penalties associated with conservative policy incongruity for Democratic women compared to Democratic men. On the other hand, liberal policy incongruity by women is likely to align more closely with expectations compared to similar behavior by men. Thus, I expect the effect of liberal policy incongruity to be more positive for Democratic women than for Democratic men.

> **H3a:** *The electoral impact of conservative policy incongruity is more negative for female Democratic representatives than for male Democratic representatives.*
>
> **H3b:** *The electoral impact of liberal policy incongruity is less negative for female Democratic representatives than for male Democratic representatives.*

There is also reason to expect that the consequences of conservative and liberal policy incongruity will be different for Republican men and women. Republican women may struggle to establish their credentials as ideological conservatives, with their gender sending a liberal cue and their party sending a conservative cue (Sanbonmatsu and Dolan, 2009; King and Matland, 2003; Dolan, 2004; Koch, 2002). Thus, Republican women may need to do more than men to establish credibility as conservatives (Dittmar, 2015, 73), but doing

so can backfire electorally because of the association between conservative and masculine stereotypes. Because Republican voters tend to hold more traditional gender role attitudes, these efforts can violate prescriptive gender role expectations surrounding femininity, exacerbating role incongruity. Past literature suggests that when Republican women highlight masculine traits in order to establish their conservative credentials, they can suffer an electoral penalty (Bauer, 2017; Cassese and Holman, 2017; Herrnson et al., 2003). Thus, there is reason to expect that Republican women will experience a greater electoral penalty than Republican men for conservative policy incongruity.

On the other hand, conservative policy incongruity could incur an electoral benefit for Republican women. First, this behavior may indeed help them overcome the liberal stereotype associated with their gender. Second, Republican women may emphasize more feminine aspects of conservatism, such as tradition and family, rather than more masculine aspects associated with policy issues and agentic characteristics. For instance, many Republican women running for office in 2020 emphasized their roles as mothers in their campaign materials (Dittmar, 2020). Drawing attention to this traditional role may soften Republican women's image, assuaging the gender role mismatch for conservative voters by providing a feminine signal in a traditionally masculine realm (e.g., Heilman and Okimoto, 2007). Moreover, past research has also found that the impact of ideological conservatism on the likelihood of winning House elections is virtually identical for men and women (Thomsen, 2020).[8] Thus, there is also reason to believe that Republican women may experience a greater or equal electoral reward than Republican men for conservative policy incongruity. For these reasons, I test the dual hypotheses that the electoral consequences of conservative policy incongruity are more negative or more positive for Republican women.

Finally, due to the association between liberal and feminine stereotypes, I expect the electoral consequences of liberal policy incongruity to be more negative for Republican men compared to Republican women. Because voters are likely to have an expectation that Republican women will be more liberal, and because of the alignment between liberal and feminine stereotypes, I expect Republican men will experience a greater electoral penalty than Republican women associated with liberal policy incongruity.

> **H4a:** *The electoral impact of conservative policy incongruity is more negative (positive) for female Republican representatives than for male Republican representatives.*
>
> **H4b:** *The electoral impact of liberal policy incongruity is more positive for female Republican representatives than for male Republican representatives.*

[8] However, this study examined House candidates rather than officeholders. Voters may respond differently to a conservative female candidate than to a woman actually exercising her duties in public office.

3.3 Data & Methods

I examine electoral returns for members of the U.S. House of Representatives from 1973-2017, focusing on two major areas of legislative activity: policy congruence and legislative productivity. The data includes 756 Democratic women, 4,735 Democratic men, 333 Republican women and 4,305 Republican men.[9] The main dependent variable is the representative's vote share in the subsequent election. Across the entire sample, men garner a significantly higher vote share than women (68.2% versus 67.5%, significant at $p = 0.08$). This difference is larger for Republican men and women (65.9% versus 63.4%, $p = 0.00$) than for Democratic men and women (70.2% versus 69.3%, $p = 0.05$).

The main independent variable is legislative activity, as captured through both legislative productivity and policy congruence. To measure legislative productivity, I examine the total number of bills each member sponsored during a Congressional session, which generates the variable *Bills*.[10] To capture policy congruence, I begin by constructing a variable which captures the extent of each member's deviation from their district's ideology for each session of Congress. To do this, I regress members' DW-NOMINATE score on each district's most recent Democratic presidential vote share to create a set of predicted DW-NOMINATE values for each district-year.[11] I then take the difference between each member's actual and predicted DW-NOMINATE scores to create a deviation score for each member. This captures how differently the member voted than what their district's ideology would predict, with positive values indicating deviation in a conservative direction and negative values indicating deviation in a liberal direction.[12] I then create the variable *Policy Congruence* by simply subtracting deviation scores from 1, such that *Policy Congruence* measures how closely members' voting records align with their districts' preferences. I also use these deviation scores to create *Conservative Policy Incongruity*, which captures deviation in a conservative direction; and *Liberal Policy Incongruity*, which captures deviation in a liberal direction. All variables are rescaled to run from 0-1.

Table 1 presents summary statistics for all independent variables by gender for the entire chamber and within each party, as well as for members' ideological voting behavior (measured through DW-NOMINATE scores) and district ideology, (measured through districts' democratic presidential vote share). It also includes summary statistics for the dependent variable, vote share. The table also indicates whether gender differences are significant for each variable within each group by bolded mean values. There are several key takeaways.

[9]The unit of observation is legislator-year, so the total number of individual legislators in each group is smaller: 156 Democratic women, 838 Democratic men, 73 Repulibcan women, and 901 Republican men.

[10]Bill sponsorship data from Volden et al. (2018); electoral data from Volden et al. (2018), The Cook Political Report (2020), and CQ Press (2018). I choose to use bill sponsorship rather than variables such as legislative effectiveness because of the clear gender difference between men and women in terms of bill sponsorship patterns.

[11]District presidential vote share data from Volden et al. (2018), The Cook Political Report (2020), and CQ Press (2018). I use members' first dimension DW-NOMINATE scores, as the first dimension captures members' placement along the traditional liberal-conservative dimension (Voteview, 2020).

[12]This follows the method employed by Lazarus and Steigerwalt (2018, 167-168). Additional details on this method in Appendix section 3.5.1.

First, women's voting behavior is generally more liberal than men's within both parties.[13] However, women - particularly Democratic women - are also elected from significantly more liberal districts. Consistent with past literature, women sponsor significantly more legislation annually on average than their male co-partisans, and also hew significantly closer to their districts' ideological preferences than men, as the measure of *Policy Congruence* indicates. This is also true for liberal and conservative policy incongruity: women deviate significantly less in both directions.

Within the parties, the picture changes somewhat. Interestingly, male Democrats are more likely than female Democrats to deviate in a liberal direction, with 86% of male Democrats and 77% of female Democrats exhibiting positive *Liberal Policy Incongruity* scores. Among Republicans, women have significantly higher *Policy Congruence* scores, while men have significantly higher *Conservative Policy Incongruity* scores, but there are no significant differences in liberal deviation.

The electoral fortunes of elected officials are tied to a wide variety of factors. Thus, I include several electorally relevant control variables in the analysis. First, I include several institutional and campaign-related control variables that have been shown to impact the electoral fortunes of House members, including incumbent and challenger expenditures, seniority, and majority party status (Jacobson and Carson, 2020). To control for district ideology, I use data on each district's Democratic presidential vote share to construct a variable that captures the extent to which each member's district voted for his or her co-partisan presidential candidate in the most recent election, called *District Co-Partisanship*. I also include party-year fixed effects in the model. This enables me to account for factors that may vary in importance over time and within the two parties, such as presidential approval and fluctuations in economic conditions, which have been shown to impact electoral performance for members of the president's party differently than out-party members (Jacobson and Carson, 2020; McGarrity, 2008; Grier and McGarrity, 2002; Bartels, 2008; Achen and Bartels, 2004). This strategy also enables me to account for the potential effects of shifts in public opinion with regard to gender role attitudes since 1973, which may lead to reduced gender bias against women in politics over time.

[13]This difference has remained relatively constant for Democrats since 1973, while for Republicans it has steadily grown smaller but remains significant. See Figure 3.7 in the Appendix. Table 6 in the Appendix also summarizes correlations between DW-NOMINATE scores and liberal/conservative policy incongruity within the parties.

Table 1: Summary Statistics

	Women				Men			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Policy Congruence	**0.785**	0.140	-0.193	1.000	**0.755**	0.157	-0.716	1.000
Conservative Policy Incongruity	**0.189**	0.117	0.000	0.587	**0.242**	0.159	0.000	1.716
Liberal Policy Incongruity	**0.237**	0.153	0.001	1.193	**0.248**	0.154	0.000	1.006
Avg Bills Sponsored Annually	**18.588**	9.225	10.667	45.889	**16.621**	7.809	9.699	40.026
District Democratic Presidential Vote Share	**57.508**	15.387	21.820	91.000	**47.822**	13.681	13.170	96.700
DW-NOMINATE Score	**-0.172**	0.419	-0.765	0.967	**0.036**	0.434	-0.916	1.633
Vote Share	**67.512**	12.779	44.000	100.000	**68.269**	13.850	36.000	100.000
N		1,086				9,037		
Democrats	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Policy Congruence	**0.776**	0.146	-0.193	1.000	**0.752**	0.156	-0.218	1.000
Conservative Policy Incongruity	**0.153**	0.102	0.000	0.482	**0.176**	0.159	0.001	1.217
Liberal Policy Incongruity	**0.251**	0.151	0.001	1.193	**0.265**	0.150	0.000	0.998
Avg Bills Sponsored Annually	**19.288**	9.794	10.064	48.214	**17.387**	8.765	8.384	45.240
District Democratic Presidential Vote Share	**64.169**	12.957	26.600	91.000	**54.388**	14.206	16.310	96.700
DW-NOMINATE Score	**-0.427**	0.141	-0.765	-0.017	**-0.332**	0.183	-0.916	0.912
Vote Share	**69.361**	13.041	44.000	100.000	**70.301**	14.758	36.000	100.000
N		*753*				*4,724*		
Republicans	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Policy Congruence	**0.804**	0.124	0.413	0.996	**0.759**	0.157	-0.716	1.000
Conservative Policy Incongruity	**0.216**	0.120	0.004	0.587	**0.257**	0.154	0.000	1.716
Liberal Policy Incongruity	0.102	0.092	0.010	0.502	0.121	0.114	0.000	1.005
Avg Bills Sponsored Annually	**17.074**	7.808	11.000	39.500	**15.499**	6.892	9.899	33.871
District Democratic Presidential Vote Share	**42.446**	8.080	21.82	62.47	**40.639**	8.499	13.170	75.000
DW-NOMINATE Score	**0.403**	0.223	-0.230	0.967	**0.440**	0.216	-0.524	1.633
Vote Share	**63.402**	11.143	44.000	100.000	**66.077**	12.431	37.000	100.000
N		*333*				*4,301*		

Gender differences significant at $p \leq 0.10$ based on one-tailed t-tests indicated by **bolded** mean values.

Liberal policy incongruity: N = 545 Democratic women; 3,821 Democratic men; 58 Republican women; 527 Republican men.

Conservative policy incongruity: N = 208 Democratic women; 903 Democratic men; ; 275 Republican women; 3,774 Republican men.

3.3.1 Analysis & Results

I begin by examining the impact of legislative activity on electoral performance for all House members, for men, and for women. To estimate these effects, I use the interactive regression model denoted in equation 3.1, with fixed effects for party and year and standard errors clustered by member:

$$\begin{aligned} V_{it+1} = {} & \beta_1 \text{Bills}_{it} + \beta_2 \text{PolicyCongruence}_{it} + \beta_3 \text{Seniority}_{it} + \beta_4 \text{MajParty}_{it} + \beta_5 \text{DistrictPresVote}_{it} + \beta_6 \text{OpponentSpending}_{it} + \\ & \beta_7 \text{IncumbentSpending}_{it} + \beta_8 \text{Bills*Female}_{it} + \beta_9 \text{PolicyCongruence*Female}_{it} + \beta_{10} \text{Seniority*Female}_{it} + \beta_{11} \text{MajParty*Female}_{it} + \\ & \beta_{12} \text{DistrictPresVote*Female}_{it} + \beta_{13} \text{OpponentSpending*Female}_{it} + \beta_{14} \text{IncumbentSpending*Female}_{it} + u_{PartyYear} + \varepsilon_{it} \end{aligned} \tag{3.1}$$

Table 2 presents the results of this model.[14] The effect of bill sponsorship on subsequent vote share is statistically indistinguishable for men and women. Although the effect of bill sponsorship is positive for men (0.006) and negative for women (-0.029), both of the individual coefficients and the gender difference are insignificant. However, the effect of *Policy Congruence* on subsequent vote share is significantly more positive for men than for women, as indicated by the significant interaction coefficient between *Policy Congruence*Female*. The coefficient on *Policy Congruence* for men is 7.916 ($p = 0.000$, s.e.= 1.557), while the coefficient for women is 1.235 ($p = 0.800$, s.e.= 4.862). Thus, there is mixed support for H1: although legislative productivity does not have significantly different electoral effects for men and women, policy congruence appears to benefit men significantly more than women. This is particularly interesting considering that women on average stay significantly closer to their districts' policy preferences than men.

[14]To recover marginal effects for women, I ran the same regression model with a dummy variable for male rather than female, summarized in Appendix Table 7. I also ran a pooled regression model for the full chamber to recover marginal effects for all members, summarized in Appendix Table 8.

Table 2: Interactive Model, Bill Sponsorship and Policy Congruence

	(1)
Variables	Vote %
Bills	0.006
	(0.014)
Policy Congruence	7.916***
	(1.557)
Seniority	0.138***
	(0.053)
Majority Party Member	10.309***
	(1.005)
District Co-Partisanship	0.401***
	(0.029)
Opponent Spending	-0.051***
	(0.013)
Incumbent Spending	-0.016**
	(0.008)
Bills*Female	-0.035
	(0.032)
Policy Congruence*Female	-10.269*
	(5.290)
Seniority*Female	0.001
	(0.168)
Majority Party Member*Female	-7.999
	(7.204)
District Co-Partisanship*Female	0.212***
	(0.056)
Opponent Spending*Female	0.028**
	(0.014)
Incumbent Spending*Female	-0.001
	(0.009)
Female	0.781
	(8.337)
Constant	26.303***
	(1.958)
Observations	8,157
R^2	0.313

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

3.3.2 Party Effects

To explore how legislative activity impacts vote share for men and women within the two parties, I ran the same fully interactive model within each party. Table 3 presents the results of both interactive models for the main independent variables.[15] There is some support for the hypothesis that gender differences in the electoral impact of legislative activity are more pronounced in the Republican party than in the Democratic party (H2). Consistent with expectations, the gender difference in the effect of bill sponsorship is larger within the Republican party compared to the Democratic party. Moreover, the effect of bill sponsorship is significantly more positive for Republican men compared to Republican women, as the significant coefficient on *Bills*Female* indicates. Bill sponsorship is associated with an electoral boost for Republican men but has no significant effect for Republican women.[16] For Democrats, there are no significant differences in the electoral effect of bill sponsorship between men and women. Figure 3.1 also plots predicted vote share values within each party across different levels of bill sponsorship.[17]

Conversely, the gender difference in the effect of policy congruence is significant within the Democratic party but not the Republican party. In fact, the effect of policy congruence is positive for Democratic men and negative for Democratic women.[18] Although the effect of policy congruence is significant for Republican men but not for Republican women, [19] the difference in effect sizes is not statistically significant (as indicated by the insignificant coefficient on *Policy Congruence*Female* for Republicans). This finding does not comport with the expectations of H2.

[15]The full results including all control variables are presented in Appendix Tables 9 and 10 for men. Appendix Tables 11 and 12 present the results of the same interactions with a dummy variable for male rather than female to recover marginal effects for women.

[16]The coefficient on *Bills* for Republican women is 0.034, $p = 0.226$. The gender difference in effect sizes is only significant in the first model, most likely due to the low number of female Republicans in models 2 and 3.

[17]I plot predicted vote share values at the mean number of bills sponsored within each party (μ), one standard deviation below the mean ($\mu - 2\sigma$), one standard deviation above the mean, and two standard deviations above the mean ($\mu + 2\sigma$).

[18]The coefficient for men is 10.160, $p = 0.000$, s.e.= 2.360; compared to a coefficient of -5.830, $p = 0.343$, s.e.= 6.145 for women.

[19]The coefficient is 4.879, $p = 0.010$, s.e.= 1.189 for men compared to a coefficient of 9.108, $p = 0.142$, s.e.= 6.198 for women.

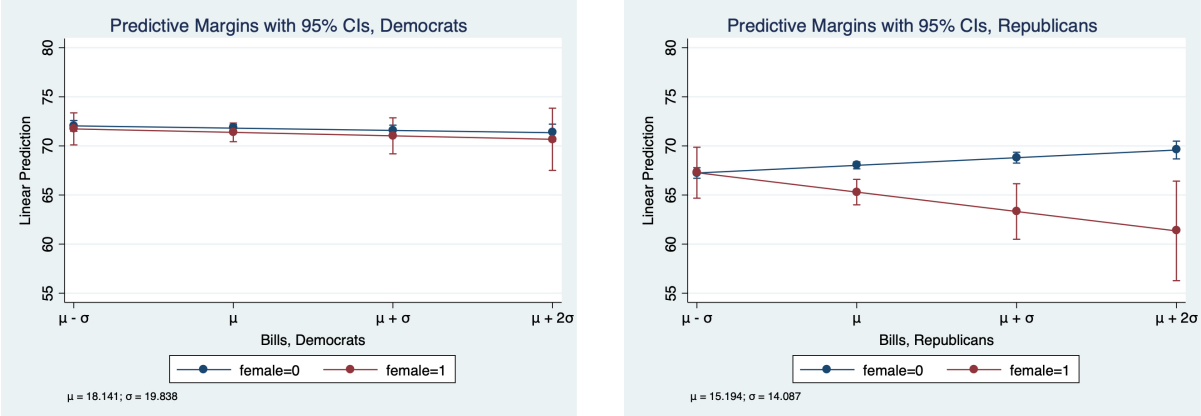
Table 3: Effects of Bill Sponsorship & Policy Congruence by Party

	(1)	(2)
Variables	Vote %, Democrats	Vote %, Republicans
Bills	-0.012	0.055**
	(0.016)	(0.015)
Bills*Female	-0.006	-0.196**
	(0.036)	(0.088)
Policy Congruence	10.160***	4.879**
	(2.360)	(1.342)
Policy Congruence*Female	-16.237**	4.229
	(6.585)	(6.481)
Female	-7.308	14.115
	(6.909)	(10.898)
Constant	44.213***	26.616***
	(2.582)	(3.371)
Observations	4,486	3,667
R^2	0.331	0.270

Robust standard errors in parentheses. Additional control variables not shown.

*** p<0.01, ** p<0.05, * p<0.10

Figure 3.1: Linear Prediction, Bill Sponsorship



3.3.2.1 Liberal and Conservative Policy Incongruity

To test the impact of liberal and conservative policy incongruity, I substituted *Conservative Policy Incongruity* and *Liberal Policy Incongruity* for *Policy Congruence*. Recall that a higher value on *Conservative (Liberal) Policy Incongruity* signifies that a member's roll call voting record is more conservative (liberal) than his or her district would predict. The results for of these models are presented in Table 4. As indicated by the insignificant coefficient on the interaction term *Conservative Policy Incongruity*Female*, the electoral effect of conservative policy incongruity is not significantly different for men and women within either party, although the penalty associated with conservative deviation is larger for Republican women (the coefficient is -11.812, $p = 0.058$, s.e.$= 6.210$ for women compared to -6.306, $p = 0.001$, s.e.$= 1.902$ for men.). The results thus suggest no substantive support for the expectation that deviation in a conservative direction impacts women more negatively within either party (H3a and H4a).

Table 4: Effects of Bill Sponsorship & Directional Policy Incongruity by Party

	(1)	(2)
Variables	Vote %, Democrats	Vote %, Republicans
Bills	0.009	0.055**
	(0.015)	(0.025)
Liberal Policy Incongruity	-22.426***	15.623***
	(2.637)	(6.033)
Conservative Policy Incongruity	20.010***	-6.306***
	(6.082)	(1.902)
Bills*Female	-0.026	-0.156*
	(0.033)	(0.090)
Liberal Policy Incongruity*Female	23.456***	8.097
	(7.695)	(19.461)
Conservative Policy Incongruity*Female	-2.103	-5.507
Female	-30.666***	20.205*
	(7.833)	(12.213)
Constant	70.557***	36.782***
	(3.666)	(4.000)
Observations	4,486	3,667
R^2	0.370	0.278

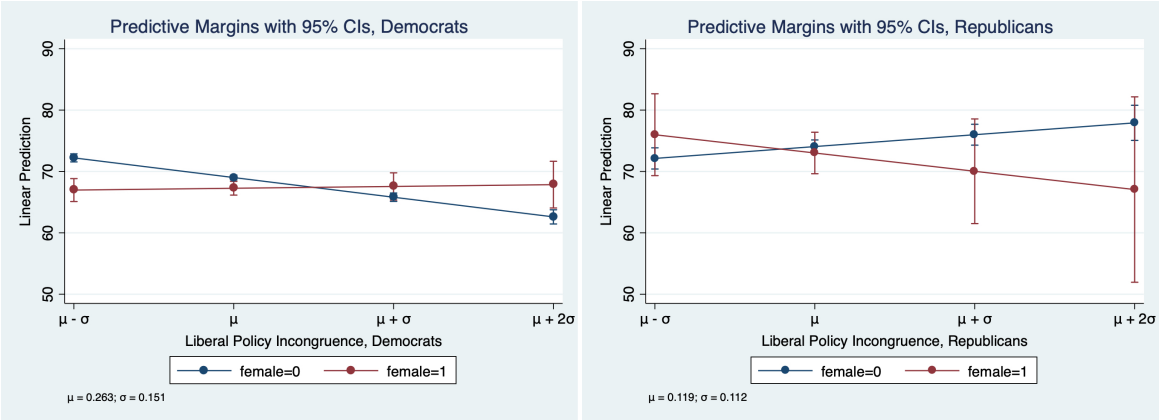
Robust standard errors in parentheses. Additional control variables not shown.

*** p<0.01, ** p<0.05, * p<0.10

On the other hand, there is support for the hypothesis that *Liberal Policy Incongruity* affects Democratic women more positively than Democratic men (H3b), although the results indicate that it is not so much that the effect is more *positive* for women but rather more *negative* for men. As the coefficient on *Liberal Policy Incongruity*Female* indicates, the effect of liberal policy incongruity is significantly more negative

for Democratic men than for Democratic women.[20] Conversely, H4b is not supported: the effect of liberal incongruity is not significantly different for Republican men and women. This may be due to the very low number of Republican women who engage in liberal policy deviation; there are only 58 total observations, which represent a total of only 20 individual Republican women. These results are also visualized in Figure 3.2.[21]

Figure 3.2: Linear Prediction, Liberal Policy Incongruity



3.3.3 Alternative Specifications

In order to better understand these results, I perform a series of additional tests and alternative specifications. First, it is possible that the effect of policy congruence is non-linear; in other words, that the effect of policy congruence changes as policy congruence increases. To test this, I use a quadratic model to discern the marginal effect of policy congruence.[22] This test shows that the effect of policy congruence is non-monotonic for both men and women. For women, the effect of policy congruence becomes significantly more positive as policy congruence increases. For men, the effect of policy congruence changes significantly across different levels of policy congruence, but does not increase in a linear fashion. To visualize this finding, I plot the effect of *Policy Congruence* for men and women who fall within different percentiles of *Policy Congruence* in Figure 3.3.

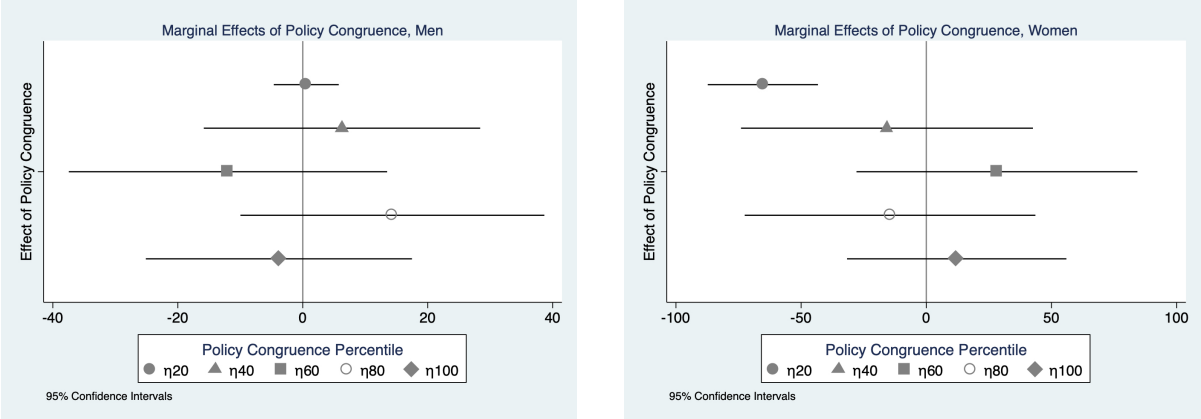
[20]The coefficient is -21.220, $p = 0.000$, s.e.= 2.858 for men compared to 1.751, $p = 0.831$, s.e.= 8.222 for women. The significant difference in the effect of *Liberal Policy Incongruity* for Democratic men and women may be because among the subsample of men and women deviate in a liberal direction, women come from districts that are significantly more Democratic.For men with positive liberal policy incongruity scores, mean co-partisanship is 50.81%; for women, it is 58.52%; difference significant at $p < 0.000$. However, the overall impact of *Liberal Policy Incongruity* on vote share is actually more negative for Democrats in strongly Democratic districts (those with a Democratic presidential vote share above 60%), compared to weakly Democratic districts (those with a Democratic presidential vote share between 50% and 60%), so this does not seem plausible.

[21]I do not plot predictive margins for *Policy Incongruity* because as I discuss below, additional tests indicate that its effects are non-linear.

[22]More detail and discussion on these tests in Appendix Section 3.5.3.1.

This analysis indicates that the effect of policy congruence is more positive for women who hew more closely to their districts' policy preferences, and more negative for women who do not. This suggests the gender differences I find between men and women may be driven by outliers, either in terms of vote share or policy incongruity. In order to discern whether this is the case, I run a series of alternative specifications. First, I run a specification using *Vote %* < 100 as the dependent variable, which excludes members elected with a vote share of 100% (column 1). This excludes members who ran unchallenged. Next, I run an alternative specification that excludes the most extreme cases of policy congruence, those with scores above the 95th percentile of *Policy Congruence*, using the variable *Policy Congruence* $< \eta 95$ (column 2). Finally, I run a model with both *Vote %* < 100 and *Policy Congruence* $< \eta 95$ (column 3).

Figure 3.3: Marginal Effects of Policy Incongruity, Men and Women



The results of the alternative specifications are summarized in in Table 5.[23] As the coefficients on the interaction variables indicate, the exclusion of the *Policy Congruence* and vote share outliers eliminates the significant gender difference in the effect of *Policy Congruence* for the entire chamber. However, the effect of greater congruence remains positive and significant for men in all three specifications, and insignificant for women. The gender difference in the effect of *Policy Congruence* also remains significant for Democrats in all three specifications. However, the coefficient on *Bills*Female* is no longer statistically significant for Republicans in two of the three alternative specifications; yet the effect of *Bills* remains significant and positive for Republican men, and is negative and insignificant for Republican women.

[23] Tables 15-19 in the Appendix present the full results. Coefficients for women were obtained by running the model with a dummy variable for male.

Table 5: Alternative Specifications

Variables	Vote % < 100	Vote Share	Vote % < 100
		Chamber	
Bills	0.021*	0.001	0.016
	(0.011)	(0.014)	(0.011)
Bills*Female	-0.017	-0.019	-0.015
	(0.028)	(0.032)	(0.028)
Policy Congruence	4.738***		
	(1.252)		
Policy Congruence*Female	-4.688		
	(3.618)		
Policy Congruence $< \eta 95$		8.261***	4.647***
		(1.625)	(1.308)
Policy Congruence $< \eta 95$*Female		-12.252**	-3.603
		(5.618)	(3.803)
Female	-0.546	2.345	-0.529
	(7.075)	(8.453)	(7.069)
Observations	7,478	7,679	7,049
R^2	0.424	0.317	0.429
		Democrats	
Bills	0.007	-0.014	0.006
	(0.013)	(0.017)	(0.014)
Bills*Female	-0.006	0.003	-0.008
	(0.032)	(0.036)	(0.033)
Policy Congruence	4.437**		
	(1.898)		
Policy Congruence*Female	-7.920*		
	(4.448)		
Policy Congruence $< \eta 95$		10.278***	4.415**
		(2.431)	(2.009)
Policy Congruence $< \eta 95$*Female		-18.249***	-7.545
		(6.873)	(4.712)
Female	-7.636	-6.115	-7.462
	(4.272)	(7.118)	(4.422)
Observations	4,055	4,234	3,837
R^2	0.458	0.336	0.463
		Republicans	
Bills	0.052***	0.043*	0.038**
	(0.019)	(0.026)	(0.019)
Bills*Female	-0.067	-0.145*	-0.050
	(0.065)	(0.080)	(0.065)
Policy Congruence	5.491***		
	(1.412)		
Policy Congruence*Female	5.653		
	(4.659)		
Policy Congruence $< \eta 95$		5.398***	5.552***
		(2.005)	(1.462)
Policy Congruence $< \eta 95$*Female		1.998	7.595
		(7.623)	(4.927)
Female	12.216	14.730	9.737
	(7.901)	(10.621)	(7.600)
Observations	3,408	3,441	3,197
R^2	0.358	0.272	0.358

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

It is also worth noting that in the alternative specifications that exclude vote shares of 100%, the effect of *Policy Congruence* and *Policy Congruence* $< \eta 95$ is positive and significant for Republican women[24] In other words, when considering all elections, *Policy Congruence* does not significantly benefit Republican women;[25] but when elections without challengers are excluded, Republican women benefit significantly - and at similar levels to Republican men - from hewing closer to their districts' policy preferences. This may be because a total of 249 observations of Republican men in the entire sample have vote shares of 100%, compared to only 13 Republican women. Moreover, *Policy Congruence* is significantly higher for Republican men who have vote shares of 100%, compared to Republican men who do not. This is not the case for Republican women.

I also perform the alternative specifications within the parties using *Liberal Policy Incongruity* and *Conservative Policy Incongruity*, with scores that fall above the 95th percentile excluded.[26] The effects of liberal policy incongruity remain indistinguishable for Republican men and women in the alternative specifications, while the effect of conservative policy incongruity is significantly more negative for women in one specification. The effect of *Liberal Policy Incongruity* remains significantly more negative for Democratic men than for Democratic women in one of the three alternative specifications.

3.4 Discussion

The results of this analysis suggest that male and female members of the U.S. House of Representatives earn differential rewards and penalties at the ballot box for similar legislative behavior, but these effects are highly conditional on party. As a whole, although female members sponsor more bills and more closely reflect their districts' policy preferences in their roll call votes, they experience no significant electoral consequences for these behaviors. Overall, the electoral effect of bill sponsorship is virtually identical for men and women, while the electoral effects of greater policy congruence are significantly more positive for men than for women.[27]

My findings further suggest gender and party interact to condition the electoral effects of different aspects of legislative activity. First, despite sponsoring significantly fewer bills than their female co-partisans, bill sponsorship is associated with a significant electoral reward for Republican men but not for Republican women. On the other hand, Republican men and women are equally impacted by policy congruence. Republican men and women also incur similar consequences for deviation away from district preferences, whether this deviation is liberal or conservative. The reverse is true for Democrats: while bill sponsorship impacts men and women equally, policy congruence benefits men significantly more than women. Moreover,

[24] Marginal effects for women were recovered by running the model with a dummy variable for male rather than female. The results of this model are summarized in Appendix Table 19.

[25] The coefficient is 9.108, see Table 11.

[26] Full results for these alternative specifications within the party presented in Appendix Tables 20 and 21.

[27] Although alternative specifications indicate that this significant gender difference is driven by outliers, the effect of policy congruence remains significant and positive for men, and insignificant for women in all three alternative specifications.

although Democratic women do not incur a significant reward for more closely representing their districts' preferences, they also are not punished as severely compared to Democratic men for deviating from their district's preferences in a liberal direction.

For the most part, these results do not comport with my theoretical expectations. There are a variety of possible reasons for this. First, my theoretical expectations with regard to the effects of counter-stereotypical behavior may be flawed. It may be the case that the stereotypes associated with political leadership roles act as a counterweight to feminine stereotypes for female officeholders, thereby reducing rather than exacerbating the negative impacts of counter-stereotypic behavior for female officeholders. Voters tend to expect politicians to be tough, ambitious, aggressive, assertive, and strategic (Huddy and Capelos, 2002). Although these traits are at odds with feminine trait stereotypes, it is possible that rather than working against women, they could work to their advantage by creating different expectations for women in political roles than for women in general. For instance, although successful female professionals tend to be seen as lacking warmth, they also tend to be seen as competent (Fiske, 2012).

Second, it may be the case that counter-stereotypical behavior by women does not impact voter evaluations in the aggregate so much as in specific cases. For example, DW-NOMINATE scores that reflect more conservative or liberal voting records *as a whole* may be less important than position-taking on specific issues, particularly if those issues are strongly associated with gender stereotypes. The data analyzed here is limited in the sense that I cannot distinguish bill sponsorship according to policy area, and I cannot distinguish which bills are highly salient in terms of public opinion. Different issue areas can alleviate or exacerbate the effects of gender, party, and leadership stereotypes, and the types of legislation sponsored by men and women is systematically different, both in terms of the policy issues it pertains to and its likelihood of success (Volden et al., 2018; Swers, 2013). Thus, women may experience significantly larger electoral rewards relative to men for sponsoring more bills that in areas that align with feminine stereotypes, like education or healthcare; conversely, they may incur penalties for sponsoring bills related to security or gun control. These effects may be further conditioned by the amount of attention different legislation receives.

Finally, it is possible that the null results with regard to Republican women specifically are attributable to the low number of Republican women in the sample. A wide body of research indicates convincingly that Republican women face systematically greater challenges than Democratic women, stemming from the mismatch between conservative and feminine stereotypes, as well as from the greater propensity of Republican voters to hold more traditional views with regard to gender roles. Future work studying these questions might isolate the impact of different forms of legislative activity on Republican women by employing experimental methods. For instance, do Republican voters react differently to a Republican female legislator who announces her support for a conservative bill related to abortion or healthcare, compared to a conservative

bill related to gun rights or crime?

These findings add new dimensions to our understanding of the role of legislative voting and accountability in American politics. Past studies on the electoral consequences of roll-call behavior have found an electoral penalty associated with party loyalty and ideological extremism (Canes-Wrone et al., 2002; Carson et al., 2010; Ansolabehere et al., 2001). Even accounting for district ideology, Democrats who vote more liberally and Republicans who vote more conservatively lose electoral support (Canes-Wrone et al., 2002). When considering the entire sample of House members, my results are consistent with this body of work.[28] However, my findings suggest that this effect is *conditional* on both gender and party. Specifically, Democratic women who vote more liberally than their districts experience no electoral consequences for this behavior, while Democratic men incur significant penalties. Along similar lines, while past work on the electoral consequences of the volume of bill sponsorship is scant, scholars have produced mixed findings. My findings suggest this may be because the consequences of bill sponsorship also depend in part on party and gender.

These findings indicate a promising avenue for future work that asks whether the effects of other canonical electorally relevant behaviors are also gendered. Moreover, these questions are not limited to House members or legislators, and indeed, there is reason to expect that gender differences in returns on electoral behavior may be more pronounced for executive offices, given the additional hurdles women face in these elections (Huddy and Terkildsen, 1993a; Dolan, 1997; Shaffner, 2005; Rose, 2013; Sanbonmatsu and Rogers, 2020). Future work can also take up this mantle.

[28]The effect of liberal policy incongruity for all Democratic House members is -19.107***, while the effect of conservative policy incongruity is -6.739*** for all Republican members.

3.5 Appendix

3.5.1 Additional Details: Policy Congruence Measurement

In order to calculate members' deviation scores, I regress district presidential vote share on DW-NOMINATE scores for each year, and then use these regression coefficients to create a set of predicted DW-NOMINATE scores for each district-year. These scores represent the DW-NOMINATE score each district's ideology (captured by the Democratic presidential vote share in the most recent election) would predict, given the relationship that year between district ideology and DW-NOMINATE scores. I then take the residuals between actual and predicted DW-NOMINATE scores to create deviation scores for each member, in each year. I then take the absolute value of the deviation score, and create the variable *Policy Congruence* by subtracting this value from 1 for each member. I calculate the directional variables by dividing the deviation variable in two, with values above zero reflecting conservative deviation and the absolute value of values below zero capturing liberal deviation. I then take the absolute value of the values below zero, so that liberal deviation scores are positive.

Figures 3.4, 3.5 and 3.6 depict the average annual *Policy Incongruity*, *Liberal Policy Incongruity*, and *Conservative Policy Incongruity* values for men and women in each party over time.

Figure 3.4: Average Annual Policy Incongruity

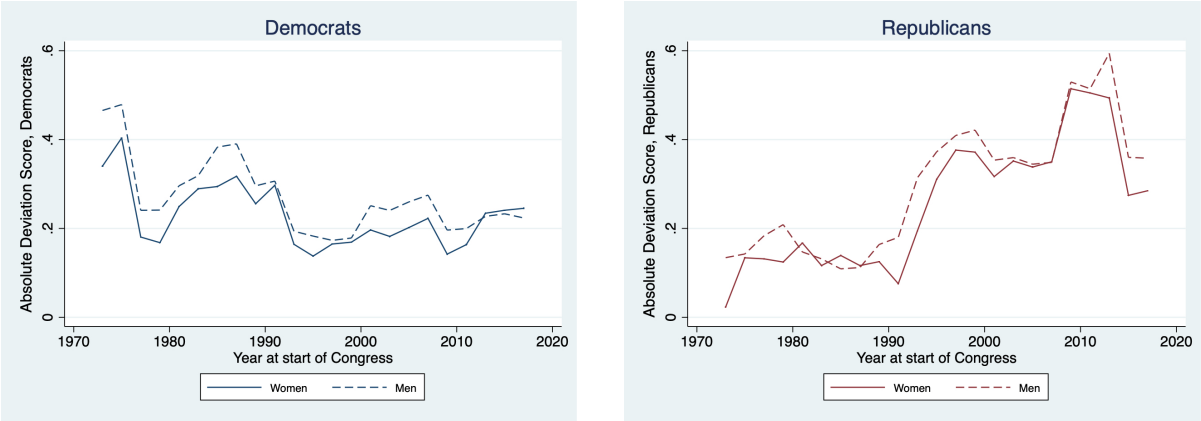


Figure 3.5: Average Annual Liberal Policy Incongruity

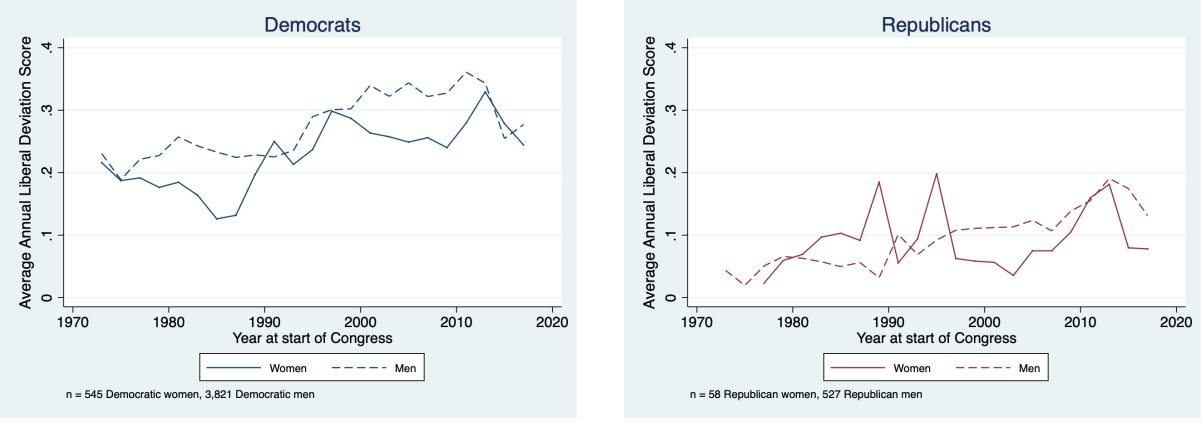


Figure 3.6: Average Annual Conservative Policy Incongruity

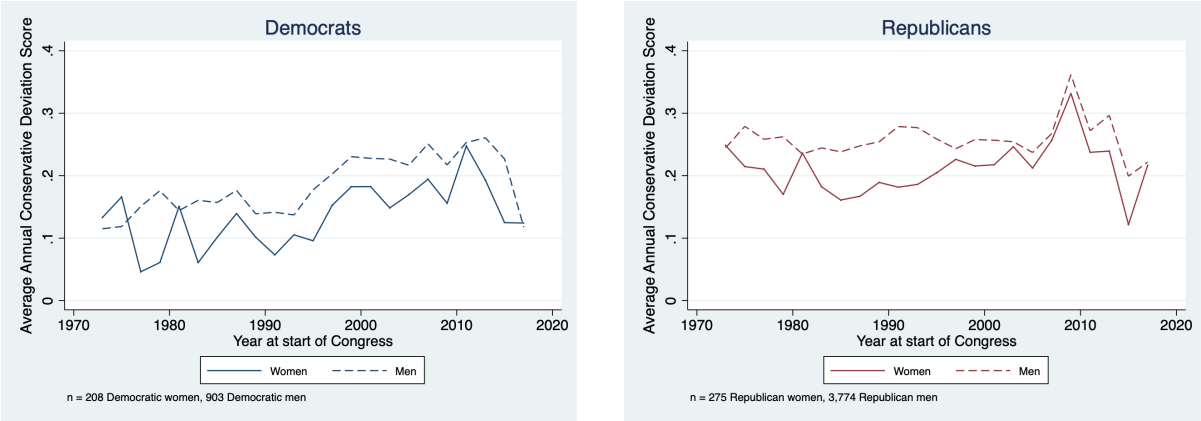
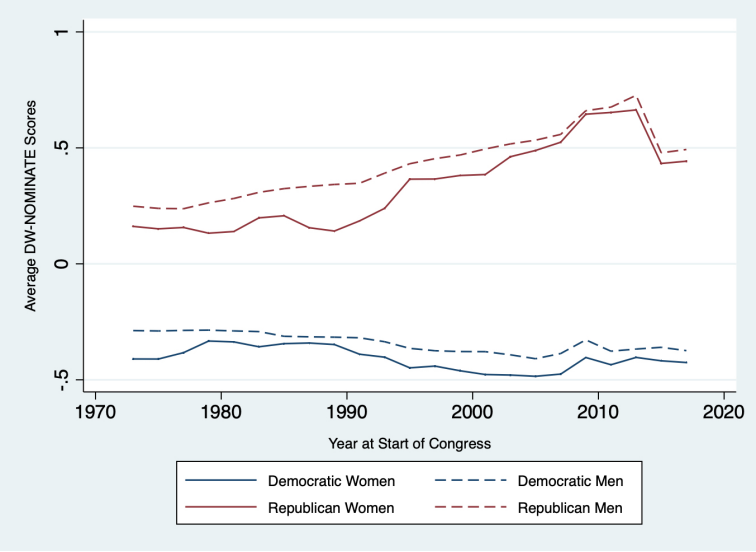


Table 6: Correlations: DW-NOMINATE and Directional Policy Incongruity

	DW-NOMINATE Democrats	DW-NOMINATE Republicans
Liberal Policy Incongruity	-0.140	0.033
Conservative Policy Incongruity	0.078	0.421

Figure 3.7: Average Annual DW-NOMINATE Scores



3.5.2 Full OLS Results

Table 7: Interactive Model, Bill Sponsorship and Policy Incongruity with Male Gender Dummy

	(1)
Variables	Vote %
Bills	-0.029
	(0.029)
Policy Congruence	-2.288
	(5.062)
Seniority	0.136**
	(0.053)
Majority Party Member	2.268
	(7.267)
District Co-Partisanship	0.615***
	(0.048)
Bills*Male	0.034
	(0.032)
Policy Congruence*Male	10.168*
	(5.297)
Seniority*Male	0.003
	(0.167)
Majority Party*Male	8.011
	(7.225)
District Co-Partisanship*Male	-0.215***
	(0.056)
Opponent Spending*Male	-0.033**
	(0.014)
Incumbent Spending*Male	0.001
	(0.010)
Constant	26.891***
	(8.120)
Observations	8,157
R^2	0.314

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Full OLS Results, Chamber

Variables	(1) Vote %
Bills	0.001
	(0.013)
Policy Congruence	6.945***
	(1.478)
Seniority	0.161***
	(0.050)
Majority Party Member	10.515***
	(0.941)
District Co-Partisanship	0.415***
	(0.026)
Opponent Spending	-0.046***
	(0.009)
Incumbent Spending	-0.017**
	(0.007)
Constant	26.239***
	(1.780)
Observations	8,157
R^2	0.305

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Full Results by Party, Policy Incongruity

Variables	(1) Vote %, Democrats	(2) Vote %, Republicans
Bills	-0.012	0.055**
	(0.016)	(0.025)
Policy Congruence	10.160***	4.879**
	(2.360)	(1.892)
Seniority	0.160**	0.107
	(0.066)	(0.094)
Majority Party Member	9.576***	10.828***
	(2.060)	(1.307)
District Co-Partisanship	0.393***	0.408***
	(0.036)	(0.051)
Opponent Spending	-0.034**	-0.067***
	(0.015)	(0.009)
Incumbent Spending	-0.043***	-0.005
	(0.007)	(0.006)
Bills*Female	-0.006	-0.196**
	(0.036)	(0.093)
Policy Congruence*Female	-16.237**	4.229
	(6.585)	(6.481)
Seniority*Female	-0.168	0.482
	(0.186)	(0.335)
Majority Party Member*Female	-1.610	-14.499**
	(4.443)	(5.861)
District Co-Partisanship*Female	0.276***	-0.112
	(0.063)	(0.153)
Opponent Spending*Female	0.014	0.040***
	(0.018)	(0.012)
Incumbent Spending*Female	0.018	-0.008
	(0.012)	(0.008)
Constant	44.213***	26.616***
	(2.582)	(3.371)
Observations	4,486	3,667
R^2	0.331	0.270

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10: Full Results by Party, Directional Policy Incongruity

Variables	(1) Vote %, Democrats	(2) Vote %, Democrats	(3) Vote %, Republicans	(4) Vote %, Republicans
Bills	-0.018	0.018	0.060**	-0.036
	(0.048)	(0.014)	(0.025)	(0.092)
Conservative Policy Incongruity	14.076*		-6.779***	
	(8.338)		(1.985)	
Seniority	0.068	0.238***	0.077	0.013
	(0.120)	(0.065)	(0.092)	(0.202)
Majority Party Member	7.448*	3.208	8.623***	1.082
	(4.046)	(2.364)	(1.376)	(6.256)
District Co-Partisanship	0.173**	0.121**	0.313***	0.360**
	(0.078)	(0.055)	(0.055)	(0.141)
Opponent Spending	-0.157**	-0.029**	-0.062***	-0.091***
	(0.065)	(0.014)	(0.008)	(0.033)
Incumbent Spending	-0.025**	-0.045***	-0.012**	0.015***
	(0.010)	(0.007)	(0.005)	(0.003)
Bills*Female	-0.010	-0.046	-0.150**	-0.242
	(0.061)	(0.053)	(0.076)	(0.378)
Conservative Policy Incongruity*Female	0.094		-3.121	
	(11.914)		(7.075)	
Seniority*Female	-0.257	-0.154	0.133	-1.208
	(0.265)	(0.230)	(0.319)	(0.818)
Majority Party Member*Female	-8.694	8.675	-13.126***	-30.728**
	(6.487)	(5.443)	(4.618)	(14.397)
District Co-Partisanship*Female	0.447***	0.408***	-0.236	1.302**
	(0.149)	(0.129)	(0.155)	(0.571)
Opponent Spending*Female	0.093	0.012	0.037***	-0.113
	(0.070)	(0.017)	(0.011)	(0.094)
Incumbent Spending*Female	0.008	0.017	-0.002	0.024
	(0.025)	(0.014)	(0.007)	(0.053)
Liberal Policy Incongruity		-21.220***		17.107**
		(2.858)		(8.548)
Liberal Policy Incongruity*Female		23.173***		-43.481
		(8.701)		(30.412)
Constant	69.509***	71.673***	37.573***	47.581***
	(6.162)	(4.114)	(4.053)	(12.736)
Observations	920	3,566	3,229	438
R^2	0.253	0.272	0.263	0.336

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Full Results by Party, Policy Congruence with Male Dummy

Variables	(1) Vote %, Democrats	(2) Vote %, Republicans
Bills	-0.017	-0.140
	(0.032)	(0.089)
Policy Congruence	-5.830	9.108
	(6.145)	(6.198)
Seniority	0.159**	0.107
	(0.066)	(0.094)
Majority Party Member	7.898**	-3.671
	(3.946)	(5.902)
District Co-Partisanship	0.671***	0.297**
	(0.051)	(0.145)
Opponent Spending	-0.015*	-0.027***
	(0.008)	(0.008)
Incumbent Spending	-0.027***	-0.013***
	(0.010)	(0.004)
Bills*Male	0.006	0.196**
	(0.036)	(0.093)
Policy Congruence*Male	15.946**	-4.229
	(6.584)	(6.481)
Seniority*Male	-0.165	0.482
	(0.185)	(0.335)
Majority Party*Male	1.661	14.499**
	(4.451)	(5.861)
District Co-Partisanship*Male	-0.280***	0.112
	(0.063)	(0.153)
Opponent Spending*Male	-0.022	-0.040***
	(0.019)	(0.012)
Incumbent Spending*Male	-0.016	0.008
	(0.012)	(0.008)
Constant	36.601***	40.732***
	(6.414)	(10.364)
Observations	4,486	3,667
R^2	0.332	0.270

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 12: Full Results by Party, Directional Policy Incongruity with Male Dummy

Variables	(1) Vote %, Democrats	(2) Vote %, Democrats	(3) Vote %, Republicans	(4) Vote %, Republicans
Bills	-0.029	-0.029	-0.090	-0.278
	(0.038)	(0.051)	(0.072)	(0.367)
Conservative Policy Incongruity	14.170*		-9.900	
	(8.510)		(6.791)	
Seniority	0.068	0.236***	0.077	0.013
	(0.120)	(0.066)	(0.092)	(0.202)
Majority Party Member	-1.246	11.832**	-4.504	-29.646*
	(5.071)	(4.926)	(4.673)	(15.321)
District Co-Partisanship	0.620***	0.534***	0.077	1.662***
	(0.127)	(0.116)	(0.145)	(0.553)
Opponent Spending	-0.064**	-0.013*	-0.024***	-0.204**
	(0.027)	(0.007)	(0.008)	(0.088)
Incumbent Spending	-0.017	-0.029**	-0.013***	0.039
	(0.023)	(0.012)	(0.004)	(0.053)
Bills*Male	0.010	0.046	0.150**	0.242
	(0.061)	(0.053)	(0.076)	(0.378)
Conservative Policy Incongruity*Male	-0.094		3.121	
	(11.914)		(7.075)	
Seniority*Male	-0.257	-0.151	0.133	-1.208
	(0.265)	(0.230)	(0.319)	(0.818)
Majority Party*Male	8.694	-8.639	13.126***	30.728**
	(6.487)	(5.465)	(4.618)	(14.397)
District Co-Partisanship*Male	-0.447***	-0.414***	0.236	-1.302**
	(0.149)	(0.129)	(0.155)	(0.571)
Opponent Spending*Male	-0.093	-0.019	-0.037***	0.113
	(0.070)	(0.018)	(0.011)	(0.094)
Incumbent Spending*Male	-0.008	-0.015	0.002	-0.024
	(0.025)	(0.014)	(0.007)	(0.053)
Liberal Policy Incongruity		1.754		-26.373
		(8.223)		(29.186)
Liberal Policy Incongruity*Male		-22.936***		43.481
		(8.706)		(30.412)
Constant	37.362***	38.853***	62.191***	-7.824
	(10.834)	(8.824)	(11.388)	(30.396)
Observations	920	3,566	3,229	438
R^2	0.253	0.273	0.263	0.336

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

3.5.3 Alternative Specifications

One additional concern is that various variables included in the model can impact campaign spending by attracting donations, or by deterring or encouraging quality challengers. Campaign spending in turn impacts vote share. However, this should only impact the substantive results if these variables *only* impact vote share through their effect on campaign spending. To ensure this is not the case, I ran a model with each independent and control variable interacted with *Opponent Spending* and *Incumbent Spending* within both parties to discern whether spending significantly changes the effects of these variables on vote share. These results, summarized in Table 13.

For Democrats, the inclusion of spending significantly weakens the impacts of *Bills* and *Policy Congruence*. The significant coefficient on *Bills*Opponent Spending* suggests that for Democrats, bill sponsorship impacts vote share through its impact on opponent spending. However, this does not affect the substantive results, as the finding with regard to bill sponsorship for Democrats is null.

For Republicans, the impact of *Bills* is only significant when the spending variables are included; however, the insignificant coefficients on *Bills*Opponent Spending* and *Bills*Incumbent Spending* suggest *Bills* exerts an impact on vote share for Republicans, independent of spending. Thus, the effects of campaign spending also does not impact the substantive results for Republicans.

Table 13: Interactive Model, Spending

Variables	(1) Vote Share, Democrats	(2) Vote Share, Republicans
Bills	0.018*	-0.025
	(0.011)	(0.016)
Policy Congruence	17.896***	11.042***
	(1.832)	(1.752)
Seniority	0.220***	0.180**
	(0.057)	(0.071)
Majority Party Member	0.542	0.654
	(0.664)	(0.532)
District Co-Partisanship	0.212***	0.079**
	(0.019)	(0.033)
Bills*Opponent Spending	-0.002***	0.000
	(0.001)	(0.000)
Policy Congruence*Opponent Spending	-0.157***	0.035
	(0.036)	(0.029)
Seniority*Opponent Spending	-0.006***	-0.005***
	(0.001)	(0.001)
Majority Party*Opponent Spending	0.040***	0.034***
	(0.009)	(0.011)
District Co-Partisanship*Opponent Spending	0.001	-0.006***
	(0.001)	(0.001)
Opponent Spending	0.049	0.221***
	(0.031)	(0.037)
Bills*Incumbent Spending	-0.000	-0.000
	(0.000)	(0.000)
Policy Congruence*Incumbent Spending	-0.136***	-0.069***
	(0.022)	(0.016)
Seniority*Incumbent Spending	0.000	0.001**
	(0.001)	(0.001)
Majority Party*Incumbent Spending	-0.016**	-0.002
	(0.007)	(0.006)
District Co-Partisanship*Incumbent Spending	0.002***	0.003***
	(0.000)	(0.000)
Incumbent Spending	-0.053***	-0.144***
	(0.019)	(0.020)
Constant	47.997***	56.358***
	(1.633)	(2.032)
Observations	4,486	3,667
R^2	0.266	0.200

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

3.5.3.1 Moderation Analysis

In order to test whether the effect of policy incongruity is non-linear, I use a moderation analysis, using the model denoted in equation 3.2. I run this model separately for men and women to test whether the effect of policy incongruity varies significantly across different values of *Policy Congruence*. These results are summarized in Table 14. Indeed, the marginal effect of policy incongruity varies significantly as it increases, indicating that it is non-monotonic for both men and women.[29]

$$V_{it+1} = \beta_1 \text{Bills}_{it} + \beta_2 \text{PolicyCongruence}_{it} + \beta_3 \text{PolicyCongruence*PolicyCongruence}_{it} + \beta_4 \text{Seniority}_{it} + \beta_5 \text{MajParty}_{it} + \beta_6 \text{DistrictPresVote}_{it} + \beta_7 \text{OpponentSpending}_{it} + \beta_8 \text{IncumbentSpending}_{it} + u_{PartyYear} + \varepsilon_{it} \tag{3.2}$$

Table 14: Marginal Effect of Policy Congruence

	(1)	(2)
Variables	Vote %, Women	Vote %, Men
Bills	-0.027	0.005
	(0.030)	(0.014)
Policy Congruence	-22.405**	-13.888***
	(9.371)	(3.542)
Policy Congruence*Policy Congruence	45.928**	9.927*
	(17.825)	(5.071)
Seniority	0.134	0.135**
	(0.168)	(0.053)
Majority Party Member	5.268	10.444***
	(7.169)	(1.008)
District Co-Partisanship	0.626***	0.399***
	(0.052)	(0.029)
Opponent Spending	-0.023***	-0.051***
	(0.006)	(0.013)
Incumbent Spending	-0.015***	-0.016**
	(0.004)	(0.008)
Constant	27.161***	35.001***
	(7.934)	(2.364)
Observations	856	7,301
R^2	0.510	0.296

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

[29] I also performed a similar test for *Bills*, which indicated that the effect of bills is monotonic for both men and women.

Table 15: Alternative Specifications, Chamber

Variables	(1) Vote % < 100	(2) Vote Share	(3) Vote % < 100
Bills	0.021*	0.001	0.016
	(0.011)	(0.014)	(0.011)
Policy Congruence	4.738***		
	(1.252)		
Seniority	0.092**	0.161***	0.112***
	(0.038)	(0.055)	(0.038)
Majority Party Member	13.366***	10.296***	13.460***
	(0.819)	(1.040)	(0.845)
District Co-Partisanship	0.437***	0.406***	0.438***
	(0.026)	(0.029)	(0.027)
Opponent Spending	-0.038***	-0.050***	-0.038***
	(0.009)	(0.013)	(0.009)
Incumbent Spending	-0.013***	-0.016*	-0.014***
	(0.004)	(0.008)	(0.005)
Bills*Female	-0.017	-0.019	-0.015
	(0.028)	(0.032)	(0.028)
Policy Congruence*Female	-4.688		
	(3.618)		
Seniority*Female	0.017	-0.008	0.009
	(0.118)	(0.158)	(0.118)
Majority Party Member*Female	-5.012	-7.602	-5.340
	(6.481)	(7.191)	(6.488)
District Co-Partisanship*Female	0.151***	0.210***	0.146***
	(0.055)	(0.056)	(0.055)
Opponent Spending*Female	0.018*	0.028*	0.017
	(0.011)	(0.014)	(0.011)
Incumbent Spending*Female	0.001	0.000	0.002
	(0.005)	(0.010)	(0.006)
Policy Congruence $< \eta 95$		8.261***	4.647***
		(1.625)	(1.308)
Policy Congruence $< \eta 95$*Female		-12.252**	-3.603
		(5.618)	(3.803)
Constant	25.635***	25.321***	25.203***
	(1.676)	(1.994)	(1.702)
Observations	7,478	7,679	7,049
R^2	0.424	0.317	0.429

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 16: Alternative Specifications, Chamber with Male Dummy

Variables	(1) Vote % < 100	(2) Vote %	(3) Vote % < 100
Bills	0.004	-0.018	0.002
	(0.026)	(0.028)	(0.026)
Bills*Male	0.017	0.019	0.014
	(0.028)	(0.032)	(0.028)
Policy Congruence	0.119		
	(3.397)		
Policy Congruence*Male	4.594		
	(3.621)		
Seniority	0.091**	0.160***	0.111***
	(0.038)	(0.055)	(0.038)
Majority Party Member	8.300	2.660	8.074
	(6.518)	(7.255)	(6.527)
District Co-Partisanship	0.589***	0.618***	0.585***
	(0.048)	(0.048)	(0.048)
Opponent Spending	-0.018***	-0.020***	-0.018***
	(0.005)	(0.005)	(0.005)
Incumbent Spending	-0.012***	-0.017***	-0.012***
	(0.003)	(0.005)	(0.003)
Seniority*Male	0.016	-0.007	0.008
	(0.117)	(0.157)	(0.118)
Majority Party*Male	5.043	7.606	5.364
	(6.497)	(7.209)	(6.503)
District Co-Partisanship*Male	-0.153***	-0.213***	-0.148***
	(0.055)	(0.057)	(0.055)
Opponent Spending*Male	-0.022**	-0.032**	-0.021*
	(0.011)	(0.015)	(0.011)
Incumbent Spending*Male	-0.001	0.001	-0.002
	(0.006)	(0.010)	(0.006)
Policy Congruence $< \eta 95$		-3.923	1.126
		(5.382)	(3.574)
Policy Congruence $< \eta 95$*Male		12.147**	3.496
		(5.623)	(3.806)
Constant	25.361***	27.472***	24.892***
	(6.922)	(8.231)	(6.912)
Observations	7,478	7,679	7,049
R^2	0.425	0.318	0.430

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 17: Alternative Specifications, Democrats, Policy Congruence

Variables	(1) Vote % < 100	(2) Vote Share	(3) Vote % < 100
Bills	0.007	-0.014	0.006
	(0.013)	(0.017)	(0.014)
Policy Congruence	4.437**		
	(1.898)		
Seniority	0.103**	0.160**	0.110**
	(0.047)	(0.068)	(0.049)
Majority Party Member	10.638***	10.035***	10.897***
	(1.455)	(2.090)	(1.477)
District Co-Partisanship	0.459***	0.398***	0.459***
	(0.031)	(0.036)	(0.032)
Opponent Spending	-0.027**	-0.033**	-0.027**
	(0.012)	(0.015)	(0.012)
Incumbent Spending	-0.031***	-0.044***	-0.031***
	(0.005)	(0.007)	(0.006)
Bills*Female	-0.006	0.003	-0.008
	(0.032)	(0.036)	(0.033)
Policy Congruence*Female	-7.920*		
	(4.448)		
Seniority*Female	-0.052	-0.139	-0.043
	(0.144)	(0.184)	(0.146)
Majority Party Member*Female	2.137	-0.620	3.041
	(3.615)	(4.580)	(3.737)
District Co-Partisanship*Female	0.193***	0.272***	0.188***
	(0.059)	(0.063)	(0.060)
Opponent Spending*Female	0.008	0.013	0.008
	(0.015)	(0.019)	(0.015)
Incumbent Spending*Female	0.014	0.019	0.015
	(0.010)	(0.013)	(0.010)
Policy Congruence $< \eta 95$		10.278***	4.415**
		(2.431)	(2.009)
Policy Congruence $< \eta 95$*Female		-18.249***	-7.545
		(6.873)	(4.712)
Constant	37.866***	43.541***	37.801***
	(1.985)	(2.608)	(2.012)
Observations	4,055	4,234	3,837
R^2	0.458	0.336	0.463

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 18: Alternative Specifications, Republicans, Policy Congruence

Variables	(1) Vote % < 100	(2) Vote Share	(3) Vote % < 100
Bills	0.052***	0.043***	0.038***
	(0.011)	(0.016)	(0.011)
Bills*Female	-0.067	-0.145	-0.050
	(0.063)	(0.090)	(0.064)
Policy Congruence	5.491***		
	(0.967)		
Policy Congruence*Female	5.653		
	(4.904)		
Policy Congruence $< \eta 95$		5.398***	5.309***
		(1.438)	(1.036)
Policy Congruence $< \eta 95$*Female		1.998	7.839
		(7.371)	(5.225)
Seniority	0.051	0.174***	0.097**
	(0.039)	(0.057)	(0.041)
Majority Party Member	11.800***	10.925*	11.995***
	(4.312)	(6.212)	(4.325)
District Co-Partisanship	0.351***	0.417***	0.359***
	(0.021)	(0.030)	(0.022)
Opponent Spending	-0.052***	-0.065***	-0.050***
	(0.003)	(0.005)	(0.003)
Incumbent Spending	-0.005**	-0.005**	-0.006***
	(0.002)	(0.003)	(0.002)
Seniority*Female	0.064	0.332	-0.002
	(0.226)	(0.318)	(0.231)
Majority Party Member*Female	-6.922	-13.059	-7.032
	(6.014)	(8.714)	(6.091)
District Co-Partisanship*Female	-0.175**	-0.111	-0.161*
	(0.084)	(0.117)	(0.087)
Opponent Spending*Female	0.028***	0.040***	0.026***
	(0.006)	(0.009)	(0.006)
Incumbent Spending*Female	-0.007	-0.008	-0.006
	(0.004)	(0.006)	(0.004)
Constant	29.867***	25.367***	29.291***
	(1.525)	(2.194)	(1.586)
Observations	3,419	3,441	3,208
R^2	0.348	0.272	0.348

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 19: Alternative Specifications, Republicans, Policy Congruence with Male Dummy

Variables	(1) Vote % < 100	(2) Vote %	(3) Vote % < 100
Bills	0.052***	0.043*	0.038**
	(0.019)	(0.026)	(0.019)
Policy Congruence	5.491***		
	(1.412)		
Seniority	0.051	0.174*	0.097
	(0.062)	(0.097)	(0.061)
Majority Party Member	11.800***	10.925***	11.995***
	(1.040)	(1.338)	(1.051)
District Co-Partisanship	0.351***	0.417***	0.359***
	(0.041)	(0.052)	(0.041)
Opponent Spending	-0.052***	-0.065***	-0.050***
	(0.006)	(0.009)	(0.006)
Incumbent Spending	-0.005	-0.005	-0.006
	(0.004)	(0.007)	(0.004)
Bills*Female	-0.067	-0.145*	-0.050
	(0.065)	(0.080)	(0.065)
Policy Congruence*Female	5.653		
	(4.659)		
Seniority*Female	0.064	0.332	-0.002
	(0.213)	(0.264)	(0.210)
Majority Party Member*Female	-6.922*	-13.059**	-7.032*
	(3.696)	(5.577)	(3.862)
District Co-Partisanship*Female	-0.175	-0.111	-0.161
	(0.120)	(0.150)	(0.115)
Opponent Spending*Female	0.028***	0.040***	0.026***
	(0.010)	(0.012)	(0.010)
Incumbent Spending*Female	-0.007	-0.008	-0.006
	(0.004)	(0.008)	(0.005)
Policy Congruence $< \eta 95$		5.398***	5.309***
		(2.005)	(1.434)
Policy Congruence $< \eta 95$*Female		1.998	7.839
		(7.623)	(4.918)
Constant	29.867***	25.367***	29.291***
	(2.658)	(3.478)	(2.735)
Observations	3,419	3,441	3,208
R^2	0.348	0.272	0.348

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 20: Alternative Specifications, Democrats, Directional Policy Incongruity

Variables	(1) Vote % < 100	(2) Vote Share	(3) Vote % < 100
Bills	0.021	0.002	0.018
	(0.013)	(0.015)	(0.013)
Liberal Policy Incongruity	-12.538***		
	(1.931)		
Conservative Policy Incongruity	16.393***		
	(5.326)		
Seniority	0.135***	0.164***	0.115***
	(0.042)	(0.060)	(0.043)
Majority Party Member	7.261***	6.719***	8.375***
	(1.494)	(2.002)	(1.371)
District Co-Partisanship	0.287***	0.230***	0.327***
	(0.041)	(0.040)	(0.031)
Opponent Spending	-0.026**	-0.034**	-0.027**
	(0.011)	(0.015)	(0.012)
Incumbent Spending	-0.029***	-0.043***	-0.030***
	(0.005)	(0.006)	(0.005)
Bills*Female	-0.016	-0.022	-0.013
	(0.030)	(0.033)	(0.030)
Liberal Policy Incongruity*Female	11.439**		
	(5.611)		
Conservative Policy Incongruity*Female	-1.464		
	(7.813)		
Seniority*Female	-0.081	-0.198	-0.071
	(0.135)	(0.179)	(0.137)
Majority Party Member*Female	3.241	-1.382	1.895
	(3.459)	(4.406)	(3.396)
District Co-Partisanship*Female	0.250***	0.224**	0.182**
	(0.094)	(0.103)	(0.076)
Opponent Spending*Female	0.009	0.016	0.010
	(0.015)	(0.019)	(0.015)
Incumbent Spending*Female	0.007	0.018	0.009
	(0.009)	(0.012)	(0.009)
Liberal Policy Incongruity $< \eta 95$		-16.480***	-10.696***
		(2.251)	(1.570)
Conservative Policy Incongruity $< \eta 95$		27.197***	21.834***
		(4.479)	(3.529)
Liberal Policy Incongruity $< \eta 95$*Female		4.727	5.103
		(5.899)	(4.222)
Conservative Policy Incongruity $< \eta 95$*Female		-12.306	-8.276
		(8.200)	(6.496)
Constant	53.513***	64.057***	50.499***
	(2.855)	(3.085)	(2.296)
Observations	4,038	4,486	4,038
R^2	0.492	0.364	0.494

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 21: Alternative Specifications, Republicans, Directional Policy Incongruity

Variables	(1) Vote % < 100	(2) Vote Share	(3) Vote % < 100
Bills	0.052***	0.055**	0.053***
	(0.019)	(0.025)	(0.019)
Liberal Policy Incongruity	14.973***		
	(4.734)		
Conservative Policy Incongruity	-7.006***		
	(1.448)		
Seniority	0.012	0.098	0.047
	(0.061)	(0.094)	(0.062)
Majority Party Member	10.377***	10.269***	11.372***
	(1.026)	(1.315)	(0.976)
District Co-Partisanship	0.271***	0.376***	0.317***
	(0.039)	(0.050)	(0.036)
Opponent Spending	-0.050***	-0.067***	-0.051***
	(0.006)	(0.009)	(0.006)
Incumbent Spending	-0.006**	-0.005	-0.006**
	(0.003)	(0.006)	(0.003)
Bills*Female	-0.047	-0.156*	-0.047
	(0.063)	(0.090)	(0.063)
Liberal Policy Incongruity*Female	-8.958		
	(12.436)		
Conservative Policy Incongruity*Female	-5.389		
	(4.787)		
Seniority*Female	-0.018	0.258	-0.049
	(0.209)	(0.372)	(0.208)
Majority Party Member*Female	-5.749*	-13.346**	-6.048*
	(3.386)	(5.267)	(3.513)
District Co-Partisanship*Female	-0.150	-0.183	-0.171
	(0.128)	(0.149)	(0.125)
Opponent Spending*Female	0.026***	0.040***	0.027***
	(0.010)	(0.012)	(0.010)
Incumbent Spending*Female	-0.005	-0.008	-0.005
	(0.004)	(0.008)	(0.004)
Liberal Policy Incongruity $< \eta 95$		20.202***	20.538***
		(6.021)	(4.892)
Conservative Policy Incongruity $< \eta 95$		-2.516	-2.672*
		(1.912)	(1.458)
Liberal Policy Incongruity $< \eta 95$*Female		3.860	-13.813
		(19.562)	(12.723)
Conservative Policy Incongruity $< \eta 95$*Female		-7.981	-7.974*
		(6.263)	(4.734)
Constant	41.158***	33.077***	36.980***
	(3.066)	(3.772)	(2.811)
Observations	3,408	3,667	3,408
R^2	0.370	0.274	0.362

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

CHAPTER 4

Part III: Gender and Political Credit Claiming

Women in American politics enjoy a variety of apparent electoral advantages over their male colleagues. As candidates, women are systematically higher-quality, and as officeholders, they are more productive and more effective along a variety of dimensions (Fulton, 2011, 2014; Pearson and McGhee, 2013; Anzia and Berry, 2011; Lazarus and Steigerwalt, 2018). Moreover, recent studies examining the impact of gender on candidate evaluation and vote choice have found that American voters actually often express a preference for female candidates, and tend to view female politicians as more honest, ethical, collaborative, and better at working to improve quality of life (Schwarz et al., 2018; Barnes and Beaulieu, 2019; Cassese and Holman, 2017; Pew Research Center, 2015).

Despite these ostensible advantages, men and women are re-elected at the same rates and tend to perform equally well in elections (Lawless, 2015), suggesting these performance and public opinion advantages do not translate to an electoral edge. An untested explanation for this paradox is that female incumbents receive relatively less credit for their performance in office. Recent studies show voters hold female candidates to higher qualification and competence standards (Bauer, 2020c; Mo, 2014; Ditonto, 2017), and myriad studies from management, education, public administration, and sociology literatures consistently suggest that men and women who hold the same roles and perform similar work are evaluated in different ways (Heilman and Haynes, 2005; Botelho and Abraham, 2017; Foschi and Valenzuela, 2012; Foschi, 2000). Particularly when they occupy leadership roles, women often suffer evaluative penalties, face stricter performance standards, and receive less credit for their accomplishments. If women must do more as candidates to convince voters of their suitability for political roles, it is intuitive that they may also have to work harder as officeholders to earn credit for their accomplishments.

There are various avenues through which voters might assign credit to officeholders for political outcomes, but a main avenue is credit claiming by the politician herself. Credit claiming is a fundamental activity of officeholders, especially so for incumbents seeking reelection (Mayhew, 1974; Grimmer et al., 2012). Credit claiming may be even more important for women if they must do more to convince voters of their suitability for political roles. If credit claiming attempts by female politicians are relatively less effective, this may explain why women's over performance as representatives does not garner them proportionate electoral rewards.

In this study, I design a survey experiment to directly test whether officeholder gender conditions the effectiveness of political credit claiming. I assign respondents to listen to short reelection campaign speeches

by fictional male or female gubernatorial incumbents in which they either claim direct credit for or simply describe accomplishments in office. I then elicit respondent evaluations of the incumbent's performance, traits, and electoral prospects to determine whether the effect of direct credit claiming differs significantly between the male and female governor. Counter to expectations, I find that the effect of direct credit claiming on performance and trait evaluations does not differ significantly between the male and female governor. I also find that the female governor is rated significantly higher than the male governor in terms of past performance evaluation, future performance expectations, and various positive traits; moreover, these advantages are actually *stronger* when she claims credit directly. Despite these advantages, respondents are no more likely to vote for the female governor, and see no significant difference in their chances of reelection.

4.1 Literature

4.1.1 Political Credit Claiming

Credit claiming is a centrally important activity for officeholders, and is especially crucial for incumbents seeking reelection (Mayhew, 1974; Grimmer et al., 2012). It is especially important because it helps officeholders strategically control how constituents attribute responsibility for policy outcomes. The attribution of political responsibility has been shown to be electorally consequential, particularly with regard to executive office (**?**Grimmer et al., 2012). Yet, attributing political responsibility is difficult for even the most politically attentive citizens, particularly in political environments in which contextual and institutional factors obfuscate clarity of responsibility (Powell and Whitten, 1993). For instance, a variety of studies indicate that voters engage in state-level economic voting in gubernatorial elections, but this relationship is conditioned by institutional factors that either clarify or obscure political responsibility (Rudolph, 2003; Lowry et al., 1998; Leyden and Borrelli, 1995; Orth, 2001). Credit claiming is one way elected officials can guide voters in assigning responsibility.

Elected officials frequently engage in a variety of credit claiming strategies: they issue press releases, make appearances at events, send e-mails, and make speeches touting their accomplishments (Grimmer et al., 2014). Moreover, credit claiming has been shown to be an effective way to build support among voters. For instance, credit claiming can increase overall support for legislators and improve ratings of their ability to deliver funding to the district and pass beneficial legislation (Grimmer et al., 2012). Slight differences in the way policy accomplishments are framed can impact levels of voter support for incumbents (Jensen and Maleksky, 2018, 206).

Very little research has examined gender differences in the volume and nature of political credit claiming strategies by male and female officeholders. There is evidence that compared to men, female officeholders claim credit more often in their newsletters and devote more space to discussing their accomplishments on

their websites (Dolan and Kropf, 2004; **?**), that female governors emphasize their clout more than male governors in state of the state addresses (Guthrie, 2018), and that female members of Congress are equally or slightly more likely than male members to use franking privileges and trips home to promote their policy accomplishments (Lazarus and Steigerwalt, 2018). Bauer (2020b) analyzed the websites of Senate candidates in 2016 and found that female candidates provided a significantly higher volume of qualification information, particularly with regard to policy accomplishments and professional qualifications.[1] On the other hand, Fridkin and Kenney (2015) show that male and female Senators are equally likely to discuss their policy accomplishments on their campaign websites and in press releases, and Bligh and Kohles (2008) find that female Senators tend to use less aggressive speech, and are less likely to use terms denoting accomplishment and praise.

However, no studies of which I am aware have examined gender differences in the effectiveness and consequences of direct credit claiming by elected officials in the political realm. Some past work has shown that women are less likely than men to receive credit for equal performance in areas outside of politics, particularly when there is ambiguity about who is responsible for outcomes, and especially in traditionally masculine domains (Heilman and Haynes, 2005). For instance, in the professional realm women's successes are more likely than men's to be attributed to luck rather than skill (Cecchi-Dimeglio, 2017), and men - but not women - experience positive status effects associated with voicing ideas (Martin, 2017). Along similar lines, women are likely to be seen as less influential and authoritative than men in deliberative settings, even when they speak up equally or more often (Mendelberg et al., 2014; Karpowitz et al., 2012).

4.1.2 Gender & Political Evaluation

A broad swath of social psychological research indicates that people are likely to perceive similar behaviors by men and women in political office differently due to stereotypical gender role expectations (Eagly and Karau, 2002; Prentice and Carranza, 2004). Gender stereotypes tend to portray women as possessing communal characteristics such as warmth, honesty, and modesty; and men as possessing agentic traits such as strength and decisiveness. Because masculine stereotypes align more closely with leadership roles, the 'lack of fit' between feminine and leadership stereotypes casts women as unlikely to succeed in leadership roles (Eagly and Karau, 2002). Behavior that violates prescriptive or descriptive social role expectations can generate negative affective reactions (Eagly and Karau, 2002). Thus, women who occupy leadership roles often experience an evaluative backlash against counter-stereotypic behavior. These penalties tend to be most severe where lack

[1] There was no difference in in male and female Senate candidates' likelihood of mentioning policy accomplishments or academic qualifications; key gender differences arose in the amount of qualification information provided. Female Senate candidates were also significantly more likely to mention their professional qualifications than male candidates, and to discuss their families. Bauer theorizes that the discussion of family life may be a strategy aimed at reducing the potentially negative consequences of violating social role expectations (Bauer, 2020b, 77).

of fit is strongest, particularly in stereotypically masculine realms such as politics.

For women in politics, many of the activities associated with political leadership risk violating stereotypical expectations and incurring a backlash. Indeed, a variety of studies show female politicians experience relatively stronger penalties for engaging in behaviors that run counter to feminine stereotypes, including negative campaigning and engaging in corrupt behavior (Krupnikov and Bauer, 2014; Cassese and Holman, 2017; Eggers et al., 2018; Courtemanche and Connor Green, 2020).[2] Women - but not men - are also punished for expressing disagreement with their party's platform, and for taking longer to respond to constituent mail (Vraga, 2017; Costa, 2020). Importantly, women who engage in self-promotion or display power-seeking intentions have also been found to suffer disproportionate penalties (Okimoto and Brescoll, 2010; Rudman, 1998).

On the other hand, traits and behavior that conform with feminine gender stereotypes can be detrimental for female politicians, while male politicians are often rewarded rather than punished for displaying feminine traits (Ditonto, 2017; Krupnikov and Bauer, 2014). For example, constituents reward men - but not women - for perceived friendliness in constituent communications (Costa, 2020). Descriptive trait expectations tend to cast women as unlikely to possess the agentic characteristics required to succeed in leadership roles, particularly in traditionally masculine realms such as politics (Eagly and Karau, 2002). This effect, often referred to as role incongruity or 'lack of fit,' tends to be strongest for women when the responsibilities associated with the role are more strongly associated with masculine traits, as in the traditionally masculine domains of politics and leadership (Heilman, 2001, 2012; Gaucher et al., 2011; Cejka and Eagly, 1994). Lack of fit can also be exacerbated or alleviated by context; for instance, factors such as economic anxieties and terrorist threat have been found to undermine support for female candidates (Holman et al., 2016; Lei and Bodenhousen, 2018). Thus, activating feminine trait expectations and stereotypes can exacerbate lack of fit and hurt perceptions of women's competence (Bauer, 2019, 2015a). This overall dynamic is often referred to as the 'double-bind': stereotype-affirming behavior can hurt perceptions of women's fitness for political leadership roles, but counter-stereotypic behavior risks generating an evaluative backlash (Costa, 2020; Teele et al., 2018).

Various studies have also uncovered a gender affinity effect in the way female politicians are evaluated. Women tend to express higher approval ratings of female officeholders' performance than men (Costa and Schaffner, 2018), and are more likely than men to support female candidates, even across party lines (Brians, 2005; **?**). On average, women tend to exhibit a preference for female candidates, and men for male can-

[2]Importantly, this has not been found to apply to *all* counter-stereotypical behaviors. Various studies have found that men and women politicians are judged similarly for behaviors such as displays of emotion or anger (Brooks, 2011; **?**) or for retaliating against campaign attacks (Krupnikov and Bauer, 2014). It should also be noted that gender stereotypes may help women in elections in which stereotypically feminine issues such as education or healthcare are especially salient (Anzia and Bernhard, 2019; Kahn, 1996).

didates (Sanbonmatsu, 2002a; Dolan, 1997), and women tend to express higher approval ratings of female representatives (Lawless, 2004). Both men and women also tend to express a preference for and warmer affect toward candidates of their own gender (Sanbonmatsu, 2002a), although this relationship is moderated by partisanship (Dolan, 2008).

A variety of studies also show that partisanship conditions the way voters evaluate female candidates. In general, Democrats are more likely than Republicans to support female candidates (Fulton and Dhima, 2020a). This may be because of the overlap between feminine and Democratic trait and issue stereotypes (Roberts and Utych, 2020). For example, Republicans tend to be seen as more masculine, while Democrats tend to be seen as more feminine (Winter, 2010; **?**). More liberal voters place greater importance on traits such as compassion and care when evaluating political leaders, while more conservative voters are more likely to prize toughness and strength (Clifford, 2020). Democratic voters are also more likely than Republican voters to care about gender equality in government (Sanbonmatsu and Dolan, 2009), while Republican voters tend to hold more traditional gender role attitudes (Knuckey, 2005). Moreover, women tend to be stereotyped as more liberal than men and as more liberal than they actually are, which can put Republican women at a disadvantage among conservative voters and Democratic women at an advantage among liberal voters (Sanbonmatsu and Dolan, 2009; King and Matland, 2003; Dolan, 2004; Koch, 2000). Indeed, studies show women actually experience an advantage among Democratic voters and a disadvantage among Republican voters (Burden et al., 2017; Burden and Ono, 2018), and that the interaction of gender and partisan stereotypes interacts to advantage Democratic female candidates but damage Republican female candidates (Bauer, 2018). There is also considerable overlap in issue area ownership: for instance, the public tends to perceive men and Republicans as better able to handle security or business and finance issues, and women and Democrats as better able to handle education and healthcare (Holman et al., 2016; Krook and O'Brien, 2012; Dolan, 2010; Fridkin and Kenney, 2009).

4.2 Theory

Credit claiming is in essence a form of persuasive speech (Dolan and Kropf, 2004). Therefore, the same factors that make persuasive speech effective should also apply to political credit claiming. Past work on persuasion suggests a variety of conditions under which persuasive speech should be most effective. Persuasive speech tends to be most effective when it is positively valenced, when it appeals to preferences for cognitive consistency, and when it aligns with prior attitudes and beliefs (Hovland et al., 1953; Chong and Druckman, 2007). Perhaps the most important aspect of effective persuasion is source credibility: sources that are seen as highly credible and competent are likely to be the most effective in prompting individuals to update their beliefs (Pornpitakpan, 2004). Perceptions of competence and expertise are key characteristics of credibility

(Dholakia and Sternthal, 1977; Hovland et al., 1953). Thus, credit claiming should be most effective when it is positively valenced, when claims align with expectations, and when its source is high in credibility.

Individuals tend to be most persuasive in contexts that correspond with salient aspects of their identity, as this increases inferences of competence and expertise, thereby heightening credibility (**?**). Thus, I expect gender to reduce the effectiveness of political credit for women because of the lack of fit between feminine and leadership stereotypes. I expect lack of fit to reduce women's persuasiveness for three main reasons. First, lack of fit can generate low performance expectations, which can in turn prevent recognition of women's successes and contributions. This is especially likely in traditionally masculine realms (Heilman, 2001; Heilman and Haynes, 2005; Mendelberg et al., 2014). Second, lack of fit may reduce women's credibility because women tend to be viewed as less competent, expert, and influential in traditionally masculine realms, even when they are objectively more influential or hold equal- or higher-status roles (McClean et al., 2018; Heilman and Haynes, 2005; Mendelberg et al., 2014; Eagly, 1983; Huddy and Terkildsen, 1993b). Women also have more difficulty than men establishing credibility in these areas, including politics (Neal et al., 2012; Armstrong and Mcadams, 2009; Bernstein, 2000; Ridgeway, 2001; Ditonto, 2017; Ditonto et al., 2014a). Reduced credibility in turn renders women relatively less persuasive, reducing the effectiveness of their political credit claiming attempts.

Third, women who hold leadership positions or engage in stereotypically masculine behaviors often experience evaluative penalties stemming from lack of fit (Eagly and Karau, 2002). Negative affective reactions may reduce the effectiveness of credit claiming for women in leadership roles because persuasive speech is more effective when it is positively valenced (Greenwald, 1968). Negative reactions may also be exacerbated when women perform leadership roles successfully, because success in these roles can strengthen inferences that women are engaging in behavior that violates stereotypical expectations (Heilman and Okimoto, 2007). Moreover, credit claiming is essentially an agentic behavior that fits well with masculine stereotypes because it entails touting accomplishments and asserting responsibility (Eagly and Karau, 2002). Thus, women may find themselves in a lose-lose in touting their performance records: for the sake of re-election, they must claim credit for their accomplishments, and yet they may experience a backlash for violating gender role expectations when they do so, rendering their claims less effective (Heilman, 2001; Gimenez et al., 2016, 2).

Credit claiming attempts by women (men) are also likely to be more effective among women (men). In general, men tend to feel an affinity for men and women for women (Pornpitakpan, 2004). This is likely to generate a positive bias toward in-group members, which can strongly, positively bias evaluation (Huddy et al., 2015). Indeed, literature on persuasion suggests that greater homophily (the degree to which individuals see themselves as similar to a source) tends to increase persuasion (McCroskey and Weiner, 1974; Wheeless, 1974). This is in part because individuals tend to view sources as more credible when they believe the speaker

shares their interests (Lupia, 2015, 1994), and gender may serve as a heuristic for shared interests.

Finally, credit claiming attempts by women are likely to be more (less) effective among more liberal (conservative) voters. The lack of fit between gender and leadership stereotypes is likely to be more severe among voters who hold more traditional gender role attitudes. Moreover, because women tend to be stereotyped as more liberal, more liberal (conservative) respondents are likely to view female officeholders more (less) favorably, increasing positive valence and boosting credibility. Thus, I expect direct credit claiming attempts by to be more effective among Democratic respondents than Republican respondents.

In sum, I expect to find that credit claiming attempts made by women are less effective compared to similar attempts made by men. I expect this because because the the mismatch between gender and leadership stereotypes should make voters less likely to attribute responsibility for political successes to women, either by reducing women's credibility or by generating a negative affective reaction. I further expect that even when direct credit claiming is effective, its electoral impact will be more positive for men than for women. Finally, I expect gender affinity and partisanship to moderate these relationships, whereby credit claiming should be more effective and its consequences more positive among women and Democrats, relative to men and Republicans.

4.3 Experimental Design

I design a survey experiment in which respondents are randomly assigned to listen to a short speech by a hypothetical male or female incumbent American governor[3] who either claims direct credit for or passively describes policy accomplishments. This design enables me to compare a female governor who claims credit directly to a female governor who does not, and a male governor who claims credit directly to a male governor who does not. In the direct credit claiming condition, the governor claims direct ownership over policy accomplishments, using direct individual language. In the passive credit claiming condition, the governor describes accomplishments using without claiming direct responsibility, using language such as 'we' and 'our' rather than 'I' and 'my.' These conditions are designed to map on to gender stereotypes. The passive condition fits with feminine stereotypes of communality and team-based leadership, while the direct condition fits with masculine stereotypes of assertiveness and self-promotion (**?**).[4]

[3] I study executive office because executives are unitary in their role and function, so claims of direct responsibility are likely to be more believable for executive compared to legislative offices (Mayhew, 1974; Jacobson and Carson, 2020).

[4] Speech scripts are detailed in the Appendix. I chose to use scripted speeches rather than real speeches for two reasons: (1) by using scripts and actors, I can ensure that treatments are as similar as possible, and (2) in watching many hours of real campaign speeches, I found that women were more likely to use the more indirect "we" rather than "I" in claiming credit for policy accomplishments, while it was far more common for men to claim direct credit. For examples, see videos from Hillary Schieve, Brian Blad, Nancy DeBoer, and Eric Hinds. This pattern renders comparisons difficult, and also lends greater legitimacy to the question of whether voters react negatively to more direct credit claims by women. On the other hand, it also raises the possibility that men may be penalized for displaying more communal characteristics or behaviors; this could also be a 'counter-stereotypical' behavior, after all. Although I do not develop or test this hypothesis directly here, I pay attention to the possibility when conducting my analyses.

Respondents listen to the speeches alongside photos of the candidates, which were pre-tested to match on perceptions of competence, attractiveness, likeability, and age. Speeches were recorded using voice actors.[5] In the speeches, the governors discuss policy accomplishments in three issue areas that are specifically selected to be as gender-neutral and nonpartisan as possible: transportation, workforce development, and bipartisan election reform.[6] The use of gender-neutral issue areas is important because highly gendered policy areas may evoke gender stereotypes and prompt respondents to view male and female officeholders as more competent in areas that align with stereotypical strengths. For instance, credit claiming by female officeholders may be more effective in policy areas such as education or health care, and less so in issue areas such as crime and security.[7]

Respondents were also randomly assigned to either a Democratic or Republican governor. The inclusion of explicit partisanship is important because of the primacy of party affiliation in American politics. Moreover, past work indicates partisanship can be more powerful than gender stereotypes (Hayes, 2011; Dolan and Sanbonmatsu, 2009; Burden and Ono, 2018; King and Matland, 2003). Thus, in the absence of explicit party cues, respondents may assume the female governor is a Democrat due to the close association between feminine, liberal, and Democratic party stereotypes (Roberts and Utych, 2020; Clifford, 2020; Sanbonmatsu and Dolan, 2009; Sanbonmatsu, 2002b). Random assignment of partisanship enables me to target the *average* effect of officeholder gender across the two parties, control for co-partisanship, and measure heterogeneous effects by partisanship.

To measure the effectiveness of credit claiming, I ask respondents a series of questions designed to measure (1) whether they attribute responsibility for the accomplishments described in the speech to the governor, and (2) whether respondents believe the performance claims the governor makes in the speech. To capture the extent to which respondents attribute responsibility to the governor, I first ask respondents if they believe the governor will do a good job if re-elected, with yes or no as the only answer choices.[8] I then ask them to indicate on a continuous scale how well they believe the Governor will perform if re-elected. I ask about fu-

[5]The candidates are both White. Matching was conducted based on a MTurk sample of 200 respondents who rated a total of approximately 150 images.

[6]Transportation and election reform are both nominally gendered-neutral issue areas. Workforce development itself has not been shown to be gendered, although it does pertain to economic issues in a more general sense, and past studies have shown that voters tend to view men as better at handling the economy. However, more recent work indicates that voters see no significant differences in the competence of male or female officeholders on economic issues (Dolan, 2010). Moreover, workforce development programs are a form of government assistance, which is associated with feminine stereotypes (Winfrey and Schnoebelen, 2019), so to the extent that workforce development evokes masculine stereotypes as an economic issue, this should be counterbalanced by government assistance stereotypes.

[7]Women tend to be viewed as better at handling issues that overlap with compassion and care, while men tend to be viewed as better able to handle issues that overlap with strength and assertiveness (Winfrey and Schnoebelen, 2019; Holman et al., 2016; Dolan, 2014; Sanbonmatsu and Dolan, 2009).

[8]Importantly, I also include an option for respondents to explain their answer choice to this question as well as to the vote intention question later in the survey. I do this in order to mitigate social desirability bias concerns, as previous studies indicate that surveys can underestimate actual levels of gender bias because respondents may be uncomfortable negatively assessing female candidates, but providing respondents a chance to justify their answers can assuage respondents' discomfort and enable researchers to more accurately capture gender bias (i.e., Krupnikov et al., 2016).

ture performance because positive performance that is attributed to an individual directly is seen as repeatable (Weiner and Rosenbaum, 1971).[9] I use these two questions about future performance to create the variable *Future Performance*. Next, I ask respondents to indicate on a continuous scale how well they believe the governor handled each of the three policy issues mentioned in the speech, and combine these three questions to create the variable *Past Performance*.[10] I solicit past performance evaluations to measure whether respondents believe the claims the governor makes in the speech. This approach has the added advantage of tapping into perceptions of credibility.

> **H1:** *Direct credit claiming will increase performance evaluations in the direct compared to the passive credit claiming condition more for the male governor than for the female governor.*

Next, I ask respondents to evaluate the governor on a variety of trait dimensions. I select traits that past literature has identified as (a) relevant to both source credibility and electoral performance, and (b) as fitting well with feminine and masculine gender stereotypes. These include confidence, competence, and expertise (stereotypically masculine), as well as authenticity and warmth (stereotypically feminine). The inclusion of authenticity and warmth also has the added advantage of capturing affective reactions. Specifically, I ask respondents to place the governors on a scale from 1-10 for the following: Ineffective-Competent, Arrogant-Confident, Amateur-Expert, Fake-Authentic, Cold-Warm. I use these trait evaluations to create the variables *Competence*, *Confidence*, *Expertise*, which I combine to create the variable *Masculine Traits*; and *Authenticity*, and *Warmth*, which I combine to create the variable *Feminine Traits*.[11]

> **H2:** *Relative to passive credit claiming, direct credit claiming will increase masculine trait evaluations significantly more for the male than for the female governor.*
>
> **H2a:** *Relative to passive credit claiming, direct credit claiming will reduce feminine trait evaluations significantly more for the female than for the male governor.*

Finally, I ask respondents to estimate on a continuous scale how likely they think it is that the governor will be reelected, and to indicate whether they would vote for the governor if they were voting in this election. I use these questions to create the variables *Electoral Viability* and *Vote Intention*. This enables me not only to test the effectiveness of credit claiming, but also to understand how credit claiming impacts perceptions of electoral viability and vote intention.

> **H3:** *Relative to passive credit claiming, direct credit claiming will have a stronger positive effect*

[9] This follows similar methods employed by Heilman and Haynes (2005, 907).
[10] Cronbach's α = 0.81 for *Future Performance* and 0.88 for *Past Performance*.
[11] Cronbach's α = 0.89 for *Masculine Traits* and 0.87 for *Feminine Traits*.

on perceptions of electoral viability and vote intention for the male governor compared to the female governor.

H4: *Effective credit claiming (positive performance evaluations) will have a stronger positive effect on perceptions of electoral viability and vote intention for the male governor compared to the female governor.*

H4a: *Effective credit claiming (positive performance evaluations) will have a stronger negative effect on perceptions of feminine traits for the female governor compared to the male governor. Conversely, effective credit claiming (positive performance evaluations) will have a stronger positive effect on perceptions of masculine traits for the male governor compared to the female governor.*

I ask respondents to indicate their partisanship and gender at the end of the survey, in order to avoid any potential bias incurred by question ordering effects. I use respondents' reported partisanship to control for the effects of shared partisanship and to test for gender differences in evaluation between the parties. I use respondents' reported gender to test for gender affinity effects.

H5: *The impacts of direct credit claiming and effective credit claiming will be more positive for respondents who share the governor's gender.*

H6: *The impacts of direct credit claiming and effective credit claiming will be more positive for the female (male) governor among Democrats (Republicans).*

4.4 Analysis & Results

I recruited 1,532 American adult respondents through Amazon MTurk to complete the survey experiment in May-June 2021. The sample is 44% male, with an average age of 41. Approximately 60% of the respondents either identify as Democrats or lean toward the Democratic party, 29% of respondents either identify or lean Republican, and just over 11% are politically independent. Respondents are well-balanced on gender, partisanship, and age between treatment conditions.[12]

Table 1 presents average treatment effects within the credit claiming condition for the female and male governor. Several patterns are of particular note. First, the results provide validation for the direct credit claiming manipulation: respondents' expectations of the incumbent's future performance, traits, and perceptions of electoral viability are all significantly higher in the direct compared to the passive credit claiming condition, for both the male and female governor. They also indicate some interesting gender differences. For

[12]Table 13 summarizes demographics by treatment condition.

the female governor, every dependent variable is significantly higher in the direct credit claiming condition *except* for vote intention. For the male governor, every dependent variable is significantly higher in the direct credit claiming condition except for past performance evaluation.

Table 1: Average Treatment Effects

	Passive Credit Claiming		Direct Credit Claiming		
	(1)	(2)	(3)	(4)	(5)
	Mean	Std. Dev.	Mean	Std. Dev.	Difference
Female Governor					
Future Performance	0.71	0.33	0.78	0.28	0.07***
Past Performance	0.57	0.25	0.61	0.23	0.04***
Feminine Traits	0.71	0.22	0.75	0.20	0.04***
Masculine Traits	0.66	0.25	0.69	0.24	0.03*
Electoral Viability	0.69	0.20	0.72	0.19	0.03**
Vote Intention	0.65	0.48	0.67	0.47	0.02
N	387		387		
Male Governor					
Future Performance	0.68	0.33	0.78	0.28	0.10***
Past Performance	0.56	0.25	0.59	0.24	0.03
Electoral Viability	0.69	0.19	0.72	0.18	0.03**
Feminine Traits	0.69	0.22	0.73	0.20	0.04**
Masculine Traits	0.62	0.24	0.65	0.24	0.03*
Vote Intention	0.62	0.49	0.67	0.47	0.05*
N	379		379		

*** p<0.01, ** p<0.05, * p<0.10

4.4.1 Performance and Trait Evaluation

To test the hypotheses that direct credit claiming will increase positive performance evaluations (H1) and positive trait evaluations (H2) significantly more for the male governor than for the female governor, I use an interactive regression model with past performance, future performance, and each trait as dependent variables, denoted in model (4.1).

$$Y_{\text{Performance Evaluations}} = \beta_0 + \beta_1 \text{Female} + \beta_2 \text{Direct Credit Claiming} + \beta_3 \text{Direct Credit Claiming*Female} + \beta_4 \text{Co-Partisan Governor} \quad (4.1)$$

Table 2 presents the interactive effect of gender and direct credit claiming on *Future Performance* and *Past Performance*. As the insignificant coefficients on the interaction variable *Direct Credit Claiming*Female Governor* indicate, there are no significant gender differences in the effect of direct credit claiming on performance evaluation.[13] Thus, this analysis produces no support for H1.

Table 2: OLS Results, Effect of Direct Credit Claiming on Performance Evaluation

	(1) Future Performance	(2) Past Performance
Direct Credit Claiming	0.058***	0.018
	(0.021)	(0.017)
Female Governor	0.014	-0.004
	(0.021)	(0.017)
Direct Credit Claiming*Female Governor	0.016	0.027
	(0.030)	(0.023)
Co-Partisan Governor	0.226***	0.153***
	(0.015)	(0.012)
Constant	0.592***	0.500***
	(0.016)	(0.013)
Observations	1,532	1,532
R^2	0.142	0.104

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10

Table 3 summarizes the interactive effect of incumbent gender and direct credit claiming on trait evaluations. The effect of direct credit claiming on masculine and feminine trait perceptions does not differ significantly between men and women, indicating that H2 and H2a are not supported. To further investigate this result, I also examined the interactive effects of of direct credit claiming and gender on each individual trait component of *Masculine Traits* and *Feminine Traits*.[14] I summarize the marginal effects of direct credit

[13] When past performance variable is broken down into its component parts, the gender interaction remains insignificant.

[14] To do this, I used the same interactive model with each individual trait as the dependent variable. Marginal effects were recovered by running regression models for the male and female governor separately to recover the effect of direct credit claiming on each trait.

claiming on each trait for the male and female governor in Table 4. Several gender differences in the marginal effects of direct credit claiming on trait evaluations are worth noting. First, direct credit claiming significantly increases perceptions of confidence, expertise, and authenticity for the female but not the male governor. Differences of means tests indicate that there are no significant gender differences in overall levels of perceived expertise in either condition, which suggests that although direct credit claiming increases estimations of the female governor's expertise, they remain virtually indistinguishable from the male governor's. Conversely, the female governor is rated significantly higher on confidence and authenticity than the male governor in both the passive and direct conditions; thus, not only does direct credit claiming increase perceptions of these traits for the female governor, but she also has a significant advantage on these traits to begin with.

Table 3: OLS Results, Effect of Direct Credit Claiming on Trait Perceptions

	(1)	(2)
Variables	Masculine Traits	Feminine Traits
Direct Credit Claiming	0.029*	0.031**
	(0.017)	(0.015)
Female Governor	0.038**	0.015
	(0.017)	(0.015)
Direct Credit Claiming*Female Governor	0.003	0.011
	(0.024)	(0.021)
Co-Partisan Governor	0.149***	0.127***
	(0.012)	(0.010)
Constant	0.556***	0.641***
	(0.013)	(0.011)
Observations	1,532	1,532
R^2	0.102	0.099

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 4: Marginal Effects of Direct Credit Claiming on Trait Perceptions

	(1) Competence	(2) Confidence	(3) Expertise	(4) Authenticity	(5) Warmth
Direct Credit Claiming	0.051***	0.045***	0.029**	0.039**	0.024
(Female Governor)	(0.017)	(0.016)	(0.014)	(0.018)	(0.017)
Observations	774	774	774	774	774
Direct Credit Claiming	0.052***	0.020	0.020	0.031	0.027
(Male Governor)	(0.016)	(0.017)	(0.014)	(0.019)	(0.017)
Observations	758	758	758	758	758

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10

4.4.2 Electoral Prospects

To test whether direct credit claiming impacts perceptions of electoral viability and vote intention more for men than for women (H3), I use the same interactive model (4.1), substituting *Electoral Viability* and *Vote Intention* as dependent variables. Table 5 summarizes the results. Counter to expectations, direct credit claiming has a virtually identical impact on perceptions of electoral viability and vote intention for the male and female governor; thus, this analysis produces no support for H3.

Table 5: OLS Results, Effect of Direct Credit Claiming on Electoral Viability and Vote Intention

	(1) Electoral Viability	(2) Vote Intention
Direct Credit Claiming	0.029**	0.035
	(0.014)	(0.030)
Female Governor	0.003	0.012
	(0.013)	(0.030)
Direct Credit Claiming*Female Governor	-0.004	-0.008
	(0.019)	(0.042)
Co-Partisan Governor	0.059***	0.484***
	(0.010)	(0.021)
Constant	0.664***	0.422***
	(0.010)	(0.023)
Observations	1,532	1,532
R^2	0.029	0.255

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

If women receive fewer electoral rewards for their performance in political office, then even where credit claiming is effective it should have a weaker effect on women's electoral prospects (H4). To test this expectation, I examine the impact of *Future Performance* and *Past Performance* on *Electoral Viability* and *Vote Intention* using an interactive regression model with performance interacted with direct credit claiming, run separately the male and female governor.[15] I then performed effect tests to determine whether the impact of performance evaluation is significantly different for the male compared to the female governor. Results are presented in Table 6. Although performance evaluations increase both perceptions of electoral viability and vote intention, effect tests reveal that the impact of effective credit claiming does not differ significantly between the male and female governor. Thus, H4 is not supported.

[15] Equation 4.2 details the full model.

Table 6: OLS Results, Effect of Performance Evaluations on Electoral Viability and Vote Intention

Variables	(1) Electoral Viability, Female Governor	(2) Electoral Viability, Male Governor	(3) Vote Intention, Female Governor	(4) Vote Intention, Male Governor
	Future Performance			
Future Performance	0.317***	0.253***	0.853***	0.861***
	(0.027)	(0.027)	(0.053)	(0.053)
Direct Credit Claiming	-0.006	0.013	-0.022	-0.056
	(0.032)	(0.031)	(0.063)	(0.060)
Future Performance*Direct Credit Claiming	0.010	0.003	-0.019	0.055
	(0.040)	(0.039)	(0.078)	(0.077)
Co-Partisan Governor	-0.001	-0.012	0.291***	0.288***
	(0.013)	(0.013)	(0.026)	(0.025)
Constant	0.470***	0.520***	-0.083**	-0.087**
	(0.020)	(0.020)	(0.039)	(0.039)
Observations	774	758	774	758
R^2	0.263	0.186	0.524	0.539
	Past Performance			
Past Performance	0.424***	0.377***	0.982***	1.028***
	(0.035)	(0.034)	(0.073)	(0.074)
Direct Credit Claiming	0.014	0.025	0.032	0.051
	(0.032)	(0.030)	(0.066)	(0.065)
Past Performance*Direct Credit Claiming	-0.012	-0.004	-0.082	-0.060
	(0.050)	(0.048)	(0.104)	(0.104)
Co-Partisan Governor	0.006	-0.010	0.336***	0.335***
	(0.013)	(0.012)	(0.027)	(0.027)
Constant	0.452***	0.481***	-0.055	-0.096**
	(0.021)	(0.020)	(0.044)	(0.044)
Observations	774	758	774	758
R^2	0.276	0.245	0.468	0.479

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

If perceptions of success in stereotypically masculine roles increase inferences that women are engaging in counter-stereotypic behavior, then effective credit claiming should reduce feminine trait perceptions more for the female governor than for the male governor. Conversely, demonstrated success in a masculine realm should improve evaluations of masculine traits significantly more for the male governor. To test this expectation, I examine the impact of *Future Performance* and *Past Performance* on *Feminine Traits* and *Masculine Traits* using the same interactive regression model as above, with performance interacted with direct credit claiming for the male and female governor, and then perform effect tests. The results are presented in Table

7. Effect tests indicate that the impact of performance evaluations on *Feminine Traits* is not significantly different for the male and female governor, suggesting no support for H4a. In fact, positive performance evaluations significantly increase feminine trait perceptions for both the male and the female governor, and direct credit claiming does not significantly change this effect. Effect tests further indicate that the effects of performance evaluations on *Masculine Traits* are similarly positive and significant for the male and female governor. This also does not support H4a.

It is also worth noting that the coefficients on *Direct Credit Claiming* are different here compared to the results in Table 3. In tests that do not control for performance evaluations, the effect of *Direct Credit Claiming* on *Electoral Viability*, *Masculine Traits*, and *Feminine Traits* perceptions is significant and positive. But when performance evaluations are included in the model along with *Direct Credit Claiming*, the significance of *Direct Credit Claiming* changes completely. In all but one specification, the effect of *Direct Credit Claiming* is insignificant; and in that one exception, the coefficient is *negative* and significant for the female governor. This suggests that the positive effects of directly claiming credit do not occur independently; instead, direct credit claiming increases electoral viability and trait perceptions *through* its positive impact on performance evaluations. In other words, performance evaluations mediate the effect of direct credit claiming.[16]

[16]Another way to examine this is to examine the impact of *Direct Credit Claiming* on *Masculine Traits* for the with and without controlling for *Future Performance*. When I perform these regressions, I find that the effect of *Direct Credit Claiming* on *Masculine Traits* is 0.031 ($p = 0.062$, s.e.$= 0.016$) without the control, and -0.013 ($p = 0.255$, s.e.$= 0.011$) when controlling for *Future Performance* for the female governor. For the male governor, the effect of *Direct Credit Claiming* on *Masculine Traits* is 0.029 ($p = 0.083, s.e.$=0.016) without the control, and -0.003 ($p = 0.767, s.e.$=0.011) with the control. Results are similar for *Feminine Traits* and *Electoral Viability*.

Table 7: OLS Results, Effect of Performance Evaluations on Feminine Trait Perceptions

Variables	(1) Feminine Traits, Female Governor	(2) Feminine Traits, Male Governor	(3) Masculine Traits, Female Governor	(4) Masculine Traits, Male Governor
	Future Performance			
Future Performance	0.550***	0.529***	0.579***	0.530***
	(0.020)	(0.021)	(0.026)	(0.026)
Direct Credit Claiming	-0.018	0.001	-0.052*	-0.049
	(0.023)	(0.023)	(0.031)	(0.030)
Future Performance*Direct Credit Claiming	0.024	-0.002	0.052	0.063
	(0.029)	(0.030)	(0.038)	(0.038)
Co-Partisan Governor	-0.005	0.013	0.015	0.020
	(0.010)	(0.010)	(0.013)	(0.013)
Constant	0.326***	0.326***	0.244***	0.246***
	(0.015)	(0.015)	(0.019)	(0.019)
Observations	774	758	774	758
R^2	0.669	0.636	0.583	0.549
	Past Performance			
Past Performance	0.668***	0.691***	0.728***	0.684***
	(0.028)	(0.026)	(0.035)	(0.035)
Direct Credit Claiming	0.003	0.033	-0.021	-0.001
	(0.025)	(0.023)	(0.031)	(0.031)
Past Performance*Direct Credit Claiming	0.013	-0.026	0.032	0.031
	(0.039)	(0.037)	(0.049)	(0.049)
Co-Partisan Governor	0.017*	0.030***	0.036***	0.041***
	(0.010)	(0.010)	(0.013)	(0.013)
Constant	0.328***	0.293***	0.233***	0.215***
	(0.017)	(0.016)	(0.021)	(0.021)
Observations	774	758	774	758
R^2	0.617	0.653	0.562	0.539
	Standard errors in parentheses			
	*** p<0.01, ** p<0.05, * p<0.1			
	Standard errors in parentheses			
	*** p<0.01, ** p<0.05, * p<0.1			

4.4.3 Gender Affinity

To understand the impact of gender affinity, I begin by exploring differences of means for all dependent variables for male and female respondents in each treatment condition.[17] On average across both conditions,

[17]These are summarized in Appendix Tables 16 and 17.

female respondents rate the female governor significantly higher than the male governor on both feminine and masculine traits, and these advantages are stronger in the direct credit claiming condition. In the direct credit claiming condition, women's evaluations of the female governor's performance are also significantly higher, both relative to the male governor and relative to male respondents' evaluations of the female governor. These patterns suggest support for the gender affinity hypothesis. Despite this, women are no more likely to say they would vote for the female governor than the male governor in any condition, nor do they see any difference in electoral viability between the male and female governor. Among male respondents, the only significant differences that emerge between evaluations of the male and female governor are in the passive credit claiming condition: men rate the female governor significantly higher than the male governor on masculine traits, and are significantly more likely to say they would vote for her. This finding runs counter to the gender affinity hypothesis.

Next, I run the interactive models used to test H1, H2, and H3 for men and women separately. Tables 8 and 9 summarize the effects of direct credit claiming on perceptions of incumbent performance, electoral viability, vote intention, and traits for male and female respondents. The main result that emerges is that among female respondents, the effect of direct credit claiming on performance evaluations is significantly more positive for the female governor.[18] Consistent with H5, this suggests direct credit claiming by the female governor is most effective among women.

[18]I also examined the impact of performance evaluation on electoral viability, vote intention, and feminine traits for the subsamples of male and female respondents separately, using the same approach I used to test H4 and H4a. Among both men and women, the effect of performance evaluations on these outcomes is virtually identical for the male and female governor.

Table 8: Interactive Effects of Gender & Direct Credit Claiming on Performance & Electoral Outcomes by Respondent Gender

	Female Respondents				Male Respondents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Future Performance	Past Performance	Electoral Viability	Vote Intention	Future Performance	Past Performance	Electoral Viability	Vote Intention
Direct Credit Claiming	0.036	-0.001	0.042**	0.013	0.070**	0.029	0.017	0.051
	(0.029)	(0.024)	(0.024)	(0.020)	(0.030)	(0.023)	(0.019)	(0.041)
Female Governor	0.006	-0.008	0.011	-0.025	0.018	-0.001	-0.003	0.039
	(0.029)	(0.024)	(0.020)	(0.045)	(0.029)	(0.022)	(0.018)	(0.040)
Direct Credit Claiming*Female Governor	0.081**	0.076**	0.003	0.062	-0.028	-0.006	-0.007	-0.061
	(0.041)	(0.034)	(0.028)	(0.063)	(0.042)	(0.032)	(0.026)	(0.057)
Co-Partisan Governor	0.191***	0.139***	0.070***	0.475***	0.250***	0.161***	0.050***	0.490***
	(0.021)	(0.017)	(0.014)	(0.032)	(0.021)	(0.016)	(0.013)	(0.029)
Constant	0.638***	0.537***	0.662***	0.446***	0.560***	0.475***	0.666***	0.405***
	(0.023)	(0.019)	(0.016)	(0.035)	(0.022)	(0.017)	(0.014)	(0.030)
Observations	678	678	678	678	854	854	854	854
R^2	0.139	0.106	0.048	0.254	0.151	0.108	0.019	0.258

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 9: Interactive Effects of Gender & Direct Credit Claiming on Trait Evaluations by Respondent Gender

	Female Respondents		Male Respondents	
	(1)	(2)	(3)	(4)
Variables	Feminine Traits	Masculine Traits	Feminine Traits	Masculine Traits
Direct Credit Claiming	0.015	0.025	0.038*	0.027
	(0.020)	(0.023)	(0.020)	(0.024)
Female Governor	0.013	0.053**	0.015	0.026
	(0.021)	(0.024)	(0.020)	(0.023)
Direct Credit Claiming*Female Governor	0.047	0.011	-0.011	0.000
	(0.029)	(0.034)	(0.028)	(0.033)
Co-Partisan Governor	0.103***	0.137***	0.143***	0.156***
	(0.015)	(0.017)	(0.014)	(0.017)
Constant	0.681***	0.583***	0.614***	0.537***
	(0.016)	(0.019)	(0.015)	(0.018)
Observations	678	678	854	854
R^2	0.089	0.108	0.112	0.100

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

4.4.4 Party Effects

To explore how partisanship conditions the effects of direct credit claiming for male and female officeholders, I start by examining mean differences for Democrats and Republicans in each treatment condition.[19] Democrats evaluate the female governor significantly higher than the male governor on *Feminine Traits*, *Masculine Traits*, and *Future Performance* across all conditions. However, Democrats do not evaluate the female governor's performance more positively, are not more likely to say they would vote for her, and do not see her as more electorally viable. Compared to the male governor, Republicans evaluate the female governor's past performance significantly more positively, and rate her significantly higher than the male governor on both *Masculine Traits* and *Feminine Traits*, but only in the direct credit claiming condition. However, between Democrats and Republicans, there are no significant differences in evaluations of the male or female governor on any variable.

Next, I run the same series of interactive models used to test H1-H3 for Democrats and Republicans separately.[20] The results for performance and electoral variables are summarized in Table 10.[21] Counter to expectations, only among Republicans does direct credit claiming significantly enhance past performance

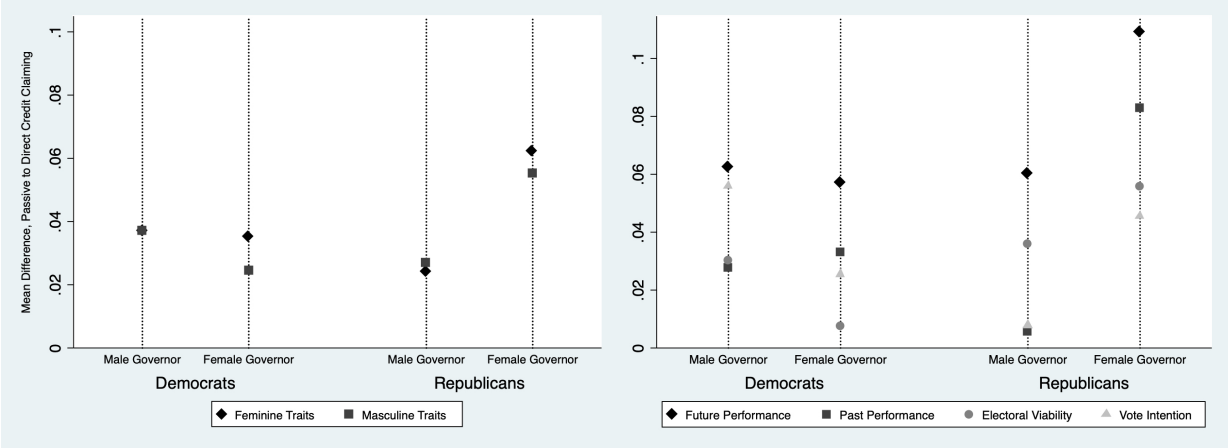
[19]These are summarized in Appendix Tables 18 and 19.

[20]I also perform the same tests used to test H4 and H4a. I find no significant gender difference in the effects of performance evaluation on electoral outcomes or trait evaluation within either party.

[21]The results for electoral outcomes and trait evaluations are presented in Appendix Table 20.

evaluations for the female governor; thus, H6 is not supported.[22] In other words, for Republicans, direct credit claiming benefits the female governor significantly more than the male governor.

Figure 4.1: Mean Differences in Dependent Variables, Passive vs. Direct Credit Claiming



These results are also visualized in Figure 4.1, which plots the mean differences between the passive and direct conditions for the male and female governor within the two parties. The effect of direct credit claiming is consistently larger among Republican respondents evaluating the female governor. Moreover, this is not attributable to lower initial evaluations of the female governor relative to the male governor, as Republicans do not evaluate the male and female governor differently in the passive condition.

There is some past work that suggest gendered evaluation patterns may be conditioned by shared co-partisanship (Ditonto et al., 2014b). Thus, in order to explore whether these results are driven by shared co-partisanship, I perform the interactive models separately for co-partisan and out-partisan governors for the entire sample, for Democrats, and for Republicans. These tests essentially ask whether the effects of direct credit claiming depend on in-party status. Table 11 presents the results for Democrats.[23] These results suggest that for Democrats, the interactive effects of direct credit claiming and governor gender are not dependent on shared partisanship with the governor. Interestingly, they also indicate that the significant, positive effect of direct credit claiming on future performance expectations only holds for co-partisan governors. Table 12 presents the results for Republicans. These results similarly suggest that the effects of direct credit claiming and gender are not contingent on governor partisanship. In other words, the finding that direct credit claiming is significantly more effective for the female governor among Republicans is not conditioned by the governor's party affiliation.[24]

[22]The effect of direct credit claiming on past performance evaluations among Republicans for the female governor is 0.080 ($p = 0.009$), while for the male governor it is 0.001 ($p = 0.973$). For Democrats, the effect of direct credit claiming is 0.037 ($p = 0.057$), compared to 0.022 ($p = 0.278$) for the male.

[23]The results for the entire sample are presented in Appendix Table 21.

[24]Effect tests also indicate that the effect of direct credit claiming on past performance evaluation does not differ significantly by governor partisanship for Republicans.

Table 10: OLS Results, Effects of Direct Credit Claiming Effects on Performance & Electoral Outcomes by Respondent Party

Variables	(1) Future Performance	(2) Past Performance	(3) Electoral Viability	(4) Vote Intention
		Democrats		
Direct Credit Claiming	0.054**	0.022	0.028*	0.038
	(0.026)	(0.020)	(0.016)	(0.037)
Female Governor	0.031	0.011	0.008	0.023
	(0.026)	(0.020)	(0.016)	(0.037)
Direct Credit Claiming*Female Governor	0.010	0.015	-0.019	0.001
	(0.037)	(0.029)	(0.023)	(0.052)
Co-Partisan Governor	0.253***	0.162***	0.065***	0.532***
	(0.018)	(0.014)	(0.012)	(0.026)
Constant	0.566***	0.489***	0.657***	0.368***
	(0.020)	(0.016)	(0.013)	(0.029)
Observations	915	915	915	915
R^2	0.184	0.127	0.037	0.318
		Republicans		
Direct Credit Claiming	0.054	0.001	0.034	-0.008
	(0.041)	(0.033)	(0.027)	(0.056)
Female Governor	-0.009	-0.029	-0.001	0.000
	(0.039)	(0.031)	(0.026)	(0.053)
Direct Credit Claiming*Female Governor	0.051	0.079*	0.021	0.043
	(0.056)	(0.045)	(0.038)	(0.077)
Co-Partisan Governor	0.191***	0.133***	0.052***	0.500***
	(0.028)	(0.022)	(0.019)	(0.038)
Constant	0.617***	0.518***	0.670***	0.413***
	(0.031)	(0.025)	(0.021)	(0.042)
Observations	446	446	446	446
R^2	0.116	0.089	0.031	0.281

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 11: OLS Results, Effects of Direct Credit Claiming and Co-Partisanship, Democrats

	(1)	(2)	(3)	(4)
Variables	Future Performance	Past Performance	Chances of Reelection	Vote Intention
	Co-Partisan Governor			
Direct Credit Claiming	0.043*	0.008	0.029	-0.000
	(0.023)	(0.025)	(0.023)	(0.034)
Direct Credit Claiming*Female Governor	-0.005	0.017	-0.042	0.030
	(0.032)	(0.035)	(0.032)	(0.048)
Female Governor	0.053**	0.027	0.040*	-0.000
	(0.023)	(0.025)	(0.023)	(0.034)
Constant	0.817***	0.650***	0.711***	0.924***
	(0.016)	(0.018)	(0.017)	(0.025)
Observations	448	448	448	448
R^2	0.034	0.011	0.007	0.002
	Out-Partisan Governor			
Direct Credit Claiming	0.066	0.036	0.028	0.073
	(0.045)	(0.032)	(0.023)	(0.063)
Direct Credit Claiming*Female Governor	0.024	0.014	0.004	-0.024
	(0.065)	(0.045)	(0.033)	(0.091)
Female Governor	0.009	-0.004	-0.024	0.042
	(0.046)	(0.032)	(0.024)	(0.064)
Constant	0.567***	0.490***	0.667***	0.347***
	(0.032)	(0.022)	(0.016)	(0.045)
Observations	467	467	467	467
R^2	0.013	0.008	0.010	0.005

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 12: OLS Results, Effects of Direct Credit Claiming and Co-Partisanship, Republicans

Variables	(1) Future Performance	(2) Past Performance	(3) Chances of Reelection	(4) Vote Intention
	Co-Partisan Governor			
Direct Credit Claiming	0.037	-0.036	0.042	-0.062
	(0.044)	(0.044)	(0.036)	(0.054)
Direct Credit Claiming*Female Governor	0.012	0.087	-0.030	0.097
	(0.060)	(0.060)	(0.049)	(0.074)
Female Governor	0.005	-0.040	0.034	-0.026
	(0.042)	(0.042)	(0.035)	(0.052)
Constant	0.820***	0.673***	0.713***	0.940***
	(0.031)	(0.031)	(0.026)	(0.038)
Observations	215	215	215	215
R^2	0.011	0.011	0.010	0.010
	Out-Partisan Governor			
Direct Credit Claiming	0.069	0.034	0.028	0.042
	(0.067)	(0.048)	(0.041)	(0.095)
Direct Credit Claiming*Female Governor	0.090	0.073	0.068	-0.006
	(0.092)	(0.066)	(0.056)	(0.131)
Female Governor	-0.023	-0.021	-0.032	0.023
	(0.063)	(0.046)	(0.039)	(0.090)
Constant	0.607***	0.500***	0.678***	0.390***
	(0.046)	(0.033)	(0.028)	(0.065)
Observations	231	231	231	231
R^2	0.032	0.027	0.028	0.002

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4.5 Discussion

Women in American politics tend to be higher-quality as candidates and more effective as officeholders. Moreover, voters often express warmer affect for female politicians, rating them more favorably than men along a variety of positive, electorally relevant trait dimensions. Yet, these factors do not translate to an advantage for women's electoral prospects: men and women are elected and reelected at similar rates, and garner similar vote shares at the ballot box. In this study, I explored a possible explanation for this paradox: that women receive relatively less credit for their policy accomplishments than men. I designed an experiment to directly test for gender differences in the effectiveness of political credit claiming. I recruited a sample of American adults to listen to a hypothetical American governor speak about his or her performance record in office, assigning them randomly to either a governor who claims credit directly or a governor who passively describes accomplishments. I then solicited respondents' evaluations of the governor's performance, traits, and electoral prospects in order to capture the effectiveness of the credit claiming attempts. This design enabled me to test whether there are gender differences in the effectiveness of direct credit claiming compared to passive credit claiming.

I hypothesized that direct credit claiming would be less effective and less electorally beneficial for the female governor compared to the male governor. This stemmed from past theoretical and empirical work suggesting that the lack of fit between feminine and leadership stereotypes would render women less persuasive in the political realm, particularly because both success and claiming direct credit should increase inferences that they are engaging in counter-stereotypic behavior. I also anticipated that direct credit claiming would improve performance evaluations, trait evaluations, electoral viability perceptions, and vote intention more for the male than for the female governor. I further anticipated that effective credit claiming - measured through strong performance evaluations - would have larger electoral benefits for the male governor. Overall, these expectations are not supported by the results of the experiment. While I do find that direct credit claiming increases performance expectations, competence and warmth assessments, and electoral viability perceptions, these effects are statistically indistinguishable for the male and female governor. Moreover, more positive performance evaluations raise respondents' perceptions of the governor's electoral prospects and their willingness to vote for the governor, and these effects are also virtually identical for the male and female governor. On average, these results do not generally suggest that women experience a penalty associated with direct credit claiming, nor that women receive significantly smaller rewards for similar policy accomplishments.

Where I do find gender differences, they tend to cut in women's favor. For example, counter to expectations, direct credit claiming significantly *increases* respondents' perceptions of positive masculine and

feminine traits for the female governor, and these increases are more common for the female governor. This is particularly surprising when it comes to the feminine traits of warmth and authenticity, as I expected direct credit claiming to reduce respondent perceptions of these traits, given that it is an assertive, agentic behavior. However, I observe the opposite. It could be that respondents' perceptions of these traits in women are sticky enough that they act as a sort of counterweight to counter-stereotypic behavior. This may be a question for future research to explore.

I also find that direct credit claiming is significantly more beneficial for the female governor among women and Republicans. Among female and among Republican respondents, direct credit claiming raises evaluations of the female governor's performance significantly more than it does for the male governor. Women also rate the female governor significantly higher on a variety of traits, both compared to their ratings of the male governor and to male respondents' ratings of the female governor, while men only rate the female governor significantly higher on feminine traits. It is quite puzzling that direct credit claiming would be more effective for the female governor among *both* women and Republicans. I speculate that there may be two different mechanisms behind these findings. Among women, I believe this finding points to support for the gender affinity hypothesis. As discussed earlier in this chapter, women are likely to feel a sense of solidarity, pride, or protectiveness for female officeholders. This possibility seems especially likely given the salience of gender in the current political climate. Conversely, one possible explanation for this finding among Republicans is the phenomenon of "reverse double standards" Foschi (2000, 37). This perspective would essentially hold that because individuals with more traditional gender attitudes are likely to have low expectations regarding women's performance in leadership roles, a stronger than expected performance can result in a disproportionate boost rather than a penalty, because "standards act as filters mediating the relationship between evaluations and expectations" (Foschi, 2000, 24). In other words, Republicans may reward women officeholders for success more than men because they did not expect it to begin with. Future research would do well to examine the ways performance expectations filter performance evaluations among different sections of the American electorate.

Along similar lines, is also worth mentioning the possibility that rather than creating a backlash effect, direct credit claiming may instead provide a way for women to prove themselves. If women are stereotyped as less likely to possess the traits necessary to lead effectively, then direct credit claiming could help to combat these stereotypes and persuade voters of women's suitability for political roles. Indeed, recent work has shown that emphasizing feminine issue strengths increases support for female candidates, while emphasizing feminine traits harms them (Bauer, 2020a). It is possible that an emphasis on performance has a similar effect as an emphasis on substantive policy issues. Because I do not have a pre-treatment measure of performance expectations, and because the passive credit claiming condition does explicitly mention policy accomplish-

ments, I cannot say whether performance expectations would be significantly lower for women in the absence of any performance record information. That is also a question for future research.

Although my initial expectations were not supported, I do find several patterns that may be consistent with some degree of gender bias in political evaluation. First, although direct credit claiming boosts evaluations of the female governor's performance, feminine and masculine traits, and electoral viability, it does not increase respondents' willingness to vote for her. Even female respondents expressed the same willingness to vote for the male and female governor. In fact, the only group that expressed a significantly greater willingness to vote for the female governor than the male governor was male respondents in the passive condition. This finding is consistent with the broader patterns motivating this study: namely, that despite women's over-performance and more positive view in the public eye, their electoral prospects remain on par with men's. This also provides some indication that women may need to be better than men in order to garner the same electoral support, although because I used equivalent performance information for the male and female governor, I can only speculate about this.

However, on a more encouraging note, this study also indicates that when voters *do* receive positive performance information about female officeholders, they respond favorably. This suggests that the uneven relationship between performance and electoral support is not entirely attributable to gender bias in the electorate, but rather to the lack of positive performance information voters receive about female officeholders. For example, Bauer (2020b) found that in the 2016 election cycle, media coverage of female Senate candidates focused significantly less on their political and professional qualifications, and more on coverage of their personal lives, relative to male candidates - despite the fact that female candidates provided significantly more qualification-related information on their campaign websites. If voters receive relatively less information about women's successes in office, this could help explain why these successes do not translate to proportionate increases in electoral support.

4.6 Appendix

4.6.1 Speech Scripts

Direct Credit Claiming: I'm very proud of my record as Governor over the past four years. To highlight just a few of my key accomplishments, I have significantly enhanced the safety and efficiency of our transportation system by making targeted investments in our state's critical infrastructure. I have also directed renewed resources to our state's workforce development programs, which have trained and employed more than 30,000 citizens this year alone. Last year, I spearheaded the creation of six new local partnerships across the state to support efforts to connect people with in-demand job skills. Under my leadership, my Administration also worked with the legislature to secure the passage of several important election reform bills with strong, bipartisan support. These measures have improved the integrity and security of our state's elections.

Passive Credit Claiming: This Administration has made major improvements to our state over the past four years. To highlight just a few of our key accomplishments, our transportation system is substantially safer and more efficient, thanks to a series of targeted investments in critical infrastructure. We have directed renewed resources to our state's workforce development programs, which have trained and employed more than 30,000 citizens this year alone. Last year, six new local partnerships were created across the state to support efforts to connect people with in-demand job skills. Our Administration has also worked with the legislature to secure the passage of several important election reform bills with strong, bipartisan support. These measures have improved the integrity and security of our state's elections.

4.6.2 Treatment Conditions

Table 13: Demographic Summary Statistics by Treatment Condition

	Female Passive	Female Direct	Male Passive	Male Direct
Men	221	221	217	195
Women	166	166	162	184
Democrats	231	221	226	237
Republicans	121	116	109	100
Independents	35	50	44	42
Mean Age	41	40	41	42
N	387	387	379	379

4.6.3 Models

$$\mathrm{Y}_{\text{Electoral Viability}} = \beta_0 + \beta_1 \text{Future Performance} + \beta_2 \text{Direct Credit Claiming} + \beta_3 \text{Future Performance*Direct Credit Claiming} + \beta_4 \text{Co-partisan Gov} \tag{4.2}$$

$$\mathrm{Y}_{\text{Performance Evaluation}} = \beta_0 + \beta_1 \text{Competence} + \beta_2 \text{Confidence} + \beta_3 \text{Expertise} + \beta_4 \text{Authenticity} + \beta_5 \text{Warmth} + \beta_6 \text{Competence*Female Gov} + \beta_7 \text{Confidence*Female Gov} + \beta_8 \text{Expertise*Female Gov} + \beta_9 \text{Authenticity*Female Gov} + \beta_{10} \text{Warmth*Female Gov} + \beta_{11} \text{Female Gov} + \beta_{11} \text{Co-partisan Gov} \tag{4.3}$$

$$\mathrm{Y}_{\text{Performance Evaluation}} = \beta_0 + \beta_1 \text{Competence} + \beta_2 \text{Confidence} + \beta_3 \text{Expertise} + \beta_4 \text{Authenticity} + \beta_5 \text{Warmth} + \beta_6 \text{Competence*Direct Credit Claiming} + \beta_7 \text{Confidence*Direct Credit Claiming} + \beta_8 \text{Expertise*Direct Credit Claiming} + \beta_9 \text{Authenticity*Direct Credit Claiming} + \beta_{10} \text{Warmth*Direct Credit Claiming} + \beta_{11} \text{Direct Credit Claiming} + \beta_{11} \text{Co-partisan Gov} \tag{4.4}$$

4.6.4 Mediation Analysis: Entire Sample

Table 14: OLS Results, Effects of Direct Credit Claiming and Performance Evaluations

Variables	(1) Chances of Reelection	(2) Vote Intention	(3) Feminine Traits	(4) Masculine Traits
	Past Performance			
Past Performance	0.401***	1.005***	0.679***	0.706***
	(0.024)	(0.052)	(0.019)	(0.025)
Direct Credit Claiming	0.020	0.042	0.018	-0.012
	(0.022)	(0.046)	(0.017)	(0.022)
Direct Credit Claiming*Past Performance	-0.010	-0.072	-0.006	0.032
	(0.035)	(0.074)	(0.027)	(0.035)
Co-Partisan Governor	-0.001	0.336***	0.024***	0.039***
	(0.009)	(0.019)	(0.007)	(0.009)
Constant	0.466***	-0.075**	0.310***	0.224***
	(0.015)	(0.031)	(0.011)	(0.015)
Observations	1,532	1,532	1,532	1,532
R^2	0.259	0.473	0.634	0.549
	Future Performance			
Future Performance	0.286***	0.857***	0.540***	0.556***
	(0.019)	(0.037)	(0.014)	(0.018)
Direct Credit Claiming	0.005	-0.039	-0.008	-0.050**
	(0.022)	(0.043)	(0.017)	(0.021)
Direct Credit Claiming*Future Performance	0.004	0.018	0.011	0.057**
	(0.028)	(0.054)	(0.021)	(0.027)
Co-Partisan Governor	-0.006	0.289***	0.004	0.018**
	(0.009)	(0.018)	(0.007)	(0.009)
Constant	0.495***	-0.085***	0.326***	0.244***
	(0.014)	(0.028)	(0.011)	(0.014)
Observations	1,532	1,532	1,532	1,532
R^2	0.223	0.531	0.653	0.566

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4.6.5 Summary Statistics with Difference of Means Tests by Governor Gender

Table 15: Dependent Variables & Difference Tests by Governor Gender

	Passive Credit Claiming			Direct Credit Claiming		
	(1)	(2)	(3)	(4)	(5)	(6)
	Female Governor	Male Governor	Gender Difference	Female Governor	Male Governor	Gender Difference
Female Respondents						
Future Performance	0.731	0.724	0.008	0.848	0.756	**0.092*****
Past Performance	0.592	0.599	0.007	0.666	0.595	**0.071*****
Competence	0.725	0.718	0.006	0.803	0.743	**0.060*****
Confidence	0.777	0.749	0.027	0.836	0.756	**0.079*****
Expertise	0.724	0.715	0.009	0.768	0.721	**0.047*****
Authenticity	0.722	0.668	**0.054****	0.763	0.690	**0.072*****
Warmth	0.674	0.620	**0.053****	0.706	0.643	**0.062*****
Electoral Viability	0.704	0.693	0.011	0.749	0.734	0.015
Vote Intention	0.638	0.660	0.021	0.710	0.663	0.048
N	166	162		166	184	
Male Respondents						
Future Performance	0.690	0.654	0.036	0.721	0.742	0.021
Past Performance	0.546	0.535	0.011	0.561	0.576	0.014
Competence	0.673	0.637	**0.034***	0.693	0.717	0.023
Confidence	0.715	0.690	0.025	0.745	0.721	0.023
Expertise	0.691	0.674	0.016	0.705	0.710	0.006
Authenticity	0.642	0.617	0.025	0.671	0.659	0.013
Warmth	0.624	0.575	**0.049****	0.636	0.610	0.026
Electoral Viability	0.686	0.685	0.001	0.693	0.705	0.012
Vote Intention	0.665	0.589	**0.075***	0.633	0.676	0.043
N	221	217		221	195	

*** p<0.01, ** p<0.05, * p<0.10

Table 16: Dependent Variables & Difference Tests by Respondent Gender

	Future Performance			Past Performance			Electoral Viability			Vote Intention		
	Female Respondents	Male Respondents	Difference	Female Respondents	Male Respondents	Difference	Female Respondents	Male Respondents	Difference	Female Respondents	Male Respondents	Difference
Passive Credit Claiming												
Female Governor	0.731	0.690	0.041	0.592	0.546	**0.045****	0.704	0.686	0.019	0.638	0.665	0.026
Male Governor	0.724	0.654	**0.069****	0.599	0.535	**0.064*****	0.693	0.685	0.008	0.660	0.589	**0.070***
Difference	0.008	0.036		0.007	0.011		0.011	0.001		0.021	**0.075***	
Direct Credit Claiming												
Female Governor	0.848	0.721	**0.126*****	0.666	0.561	**0.104*****	0.749	0.693	**0.055*****	0.710	0.633	**0.077***
Male Governor	0.756	0.742	0.013	0.595	0.576	0.019	0.734	0.705	**0.027***	0.663	0.676	0.014
Difference	**0.092*****	0.021		**0.071*****	0.014		0.015	0.012		0.048	0.043	

*** p<0.01, ** p<0.05, * p<0.10

Difference columns indicate whether there is a significant difference between male and female respondents perceptions of the governor, while difference rows indicate whether there is a significant difference between perceptions of the male and female governor for female and then for male respondents. N Women, Passive Female Gov = 166; N Men, Passive Female Gov = 216. N Women, Passive Male Gov = 162; N Men, Passive Male Gov = 217. N Women, Direct Female Gov = 166; N Men, Direct Female Gov = 221. N Women, Direct Male Gov = 184; N Men, Direct Male Gov = 195.

Table 17: Dependent Variables & Difference Tests by Respondent Gender

	Feminine Traits			Masculine Traits		
	Female Respondents	Male Respondents	Difference	Female Respondents	Male Respondents	Difference
Passive Credit Claiming						
Female Governor	0.742	0.693	0.048**	0.698	0.633	0.064***
Male Governor	0.728	0.668	0.060***	0.645	0.596	0.049**
Difference	0.014	0.025		0.053**	0.037*	
Direct Credit Claiming						
Female Governor	0.803	0.715	0.088***	0.734	0.654	0.080***
Male Governor	0.740	0.717	0.023	0.667	0.634	0.032*
Difference	0.062***	0.002		0.067***	0.019	

*** p<0.01, ** p<0.05, * p<0.10

Difference columns indicate whether there is a significant difference between male and female respondents perceptions of the governor, while difference rows indicate whether there is a significant difference between perceptions of the male and female governor for female and then for male respondents.

N Women, Passive Female Gov = 166; N Men, Passive Female Gov = 216. N Women, Passive Male Gov = 162; N Men, Passive Male Gov = 217. N Women, Direct Female Gov = 166; N Men, Direct Female Gov = 221. N Women, Direct Male Gov = 184; N Men, Direct Male Gov = 195.

Table 18: Dependent Variables & Difference Tests by Respondent Party

	Future Performance			Past Performance			Electoral Viability			Vote Intention		
	Democrats	Republicans	Difference	Democrats	Republicans	Difference	Democrats	Republicans	Difference	Democrats	Republicans	Difference
Passive Credit Claiming												
Female Governor	0.727	0.700	0.027	0.583	0.553	0.030	0.698	0.694	0.004	0.662	0.653	0.009
Male Governor	0.684	0.705	0.021	0.565	0.580	0.015	0.687	0.694	0.006	0.615	0.642	0.027
Difference	0.042*	0.005		0.018	0.026		0.011	0.000		0.047	0.010	
Direct Credit Claiming												
Female Governor	0.784	0.809	0.025	0.616	0.636	0.019	0.706	0.750	0.044**	0.688	0.698	0.010
Male Governor	0.746	0.765	0.019	0.592	0.585	0.007	0.717	0.730	0.012	0.671	0.650	0.020
Difference	0.037*	0.043		0.024	0.051*		0.011	0.020		0.017	0.048	

*** p<0.01, ** p<0.05, * p<0.10

Difference columns indicate whether there is a significant difference between Democratic and Republican respondents' perceptions of the governor, while difference rows indicate whether there is a significant difference between perceptions of the male and female governor for Democratic and for Republican respondents.
Female Passive: N=231 Democrats, 121 Republicans; Male Passive: N=226 Democrats, 109 Republicans. Female Direct: N=221 Democrats, 116 Republicans; Male Direct: N=237 Democrats, 100 Republicans.

Table 19: Dependent Variables & Difference Tests by Respondent Party

	Feminine Traits			Masculine Traits		
	Democrats	Republicans	Difference	Democrats	Republicans	Difference
Passive Credit Claiming						
Female Governor	0.721	0.715	0.006	0.666	0.669	0.003
Male Governor	0.695	0.708	0.013	0.607	0.639	0.031
Difference	0.026*	0.006		0.058***	0.029	
Direct Credit Claiming						
Female Governor	0.756	0.777	0.020	0.690	0.724	0.033
Male Governor	0.732	0.732	0.000	0.645	0.666	0.021
Difference	0.024*	0.044*		0.046**	0.058**	

*** p<0.01, ** p<0.05, * p<0.10

Difference columns indicate whether there is a significant difference between Democratic and Republican respondents' perceptions of the governor, while difference rows indicate whether there is a significant difference between perceptions of the male and female governor within groups.

Female Passive: N=231 Democrats, 121 Republicans; Male Passive: N=226 Democrats, 109 Republicans. Female Direct: N=221 Democrats, 116 Republicans; Male Direct: N=237 Democrats, 100 Republicans.

Table 20: OLS Results, Effect of Performance Evaluations on Electoral Viability, Vote Intention, & Trait Evaluations

Variables	(1) Electoral Viability	(2) Vote Intention	(3) Feminine Traits	(4) Masculine Traits
		Democrats		
Direct Credit Claiming	0.028*	0.038	0.032*	0.032
	(0.016)	(0.037)	(0.018)	(0.021)
Female Governor	0.008	0.023	0.020	0.051**
	(0.016)	(0.037)	(0.018)	(0.021)
Direct Credit Claiming*Female Governor	-0.019	0.001	0.006	-0.003
	(0.023)	(0.052)	(0.026)	(0.030)
Co-Partisan Governor	0.065***	0.532***	0.133***	0.167***
	(0.012)	(0.026)	(0.013)	(0.015)
Constant	0.657***	0.368***	0.633***	0.530***
	(0.013)	(0.029)	(0.014)	(0.016)
Observations	915	915	915	915
R^2	0.037	0.318	0.115	0.135
		Republicans		
Direct Credit Claiming	0.034	-0.008	0.020	0.023
	(0.027)	(0.056)	(0.029)	(0.033)
Female Governor	-0.001	0.000	0.004	0.027
	(0.026)	(0.053)	(0.028)	(0.032)
Direct Credit Claiming*Female Governor	0.021	0.043	0.039	0.030
	(0.038)	(0.077)	(0.040)	(0.045)
Co-Partisan Governor	0.052***	0.500***	0.117***	0.128***
	(0.019)	(0.038)	(0.020)	(0.023)
Constant	0.670***	0.413***	0.654***	0.581***
	(0.021)	(0.042)	(0.022)	(0.025)
Observations	446	446	446	446
R^2	0.031	0.281	0.088	0.082

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 21: OLS Results, Effects of Direct Credit Claiming by Co-Partisanship

	(1)	(2)	(3)	(4)
Variables	Future Performance	Past Performance	Chances of Reelection	Vote Intention
	Co-Partisan Governor			
Direct Credit Claiming	0.041**	-0.005	0.033*	-0.019
	(0.021)	(0.022)	(0.019)	(0.029)
Direct Credit Claiming*Female Governor	-0.000	0.039	-0.037	0.050
	(0.029)	(0.030)	(0.027)	(0.040)
Female Governor	0.037*	0.005	0.038**	-0.009
	(0.020)	(0.022)	(0.019)	(0.029)
Constant	0.818***	0.658***	0.712***	0.929***
	(0.015)	(0.016)	(0.014)	(0.021)
Observations	663	663	663	663
R^2	0.021	0.008	0.007	0.003
	Out-Partisan Governor			
Direct Credit Claiming	0.071**	0.035	0.027	0.075
	(0.033)	(0.024)	(0.019)	(0.048)
Direct Credit Claiming*Female Governor	0.030	0.020	0.022	-0.051
	(0.047)	(0.034)	(0.026)	(0.067)
Female Governor	-0.005	-0.012	-0.025	0.025
	(0.033)	(0.024)	(0.019)	(0.048)
Constant	0.592***	0.497***	0.673***	0.406***
	(0.023)	(0.017)	(0.013)	(0.033)
Observations	869	869	869	869
R^2	0.016	0.008	0.011	0.003

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

CHAPTER 5

Conclusion

In this dissertation, I asked: do voters evaluate performance in political office, and attribute political responsibility for outcomes, differently for male and female officeholders? These questions originated from the puzzle that although women systematically outperform men in political office, they garner virtually equivalent electoral support. An untested explanation for this enigma is that women are held to systematically higher performance standards. Previous research indicates that voters hold women to higher qualification standards as candidates, but do they continue to hold women to higher performance standards once they reach office? Female officeholders often express the belief that they must do more to prove that they can effectively perform political leadership roles, but whether this is empirically true has not been fully explored.

I approached this question in three different ways. In Part I, I conducted a survey experiment designed to isolate the impact of officeholder gender on performance evaluations and the attribution of responsibility. I randomly assigned a nationally representative sample of American respondents to receive positive, negative, or neutral economic performance information about a hypothetical male or female American governor. I then solicited respondents' evaluations of the governor's performance and the extent to which they attributed performance to the governor. I also examined heterogeneous effects among men and women, and Democrats and Republicans. I expected respondents would reward the female governor relatively less than the male governor for positive performance and punish her relatively more for negative performance. I further anticipated gender and party would condition the impact of gender on political performance evaluation and attribution, such that gender disparities would be more pronounced among men and Republicans. On average, the results of this experiment do not comport with my expectations: instead, I find that respondent evaluations of positive performance by the male and female governor are virtually identical. Moreover, the female governor is attributed with significantly less influence than the male governor for negative performance, and significantly more influence for neutral performance. Moreover, instead of finding more pronounced gender bias against the female governor among men and Republicans, I find that women and Democrats were significantly more likely to blame the male governor for negative performance.

In Part II, I asked whether legislative activity garners differential electoral consequences for men and women. I examined the impact of legislative productivity and policy congruence on electoral returns for members of the U.S. House of Representatives from 1973-2017. In part because women have significantly higher levels of both legislative productivity and policy congruence but perform as well as men in elections, I expected the electoral returns on these activities would be larger for men than for women, and that these

disparities would be more pronounced within the Republican party. I further expected that women would be punished less than men when their voting records deviated away from district preferences in a liberal direction, and more than men for conservative deviation. Here, my findings are somewhat more consistent with expectations. For the entire chamber, I find that the electoral consequences of bill sponsorship are virtually identical for men and women, but that policy congruence is significantly more positive for men. Additional analyses indicate that these patterns operate differently within the two parties. Although Republican men sponsor significantly fewer bills than Republican women, bill sponsorship is associated with a significant electoral reward for Republican men but not for Republican women. Similarly, although Democratic women hew more closely to their districts' preferences than Democratic men, policy congruence incurs a significant electoral reward for Democratic men but not for Democratic women. However, Democratic women are punished significantly *less* than Democratic men for voting more liberally than their districts.

Finally, in Part III I conducted an experiment to test whether directly claiming credit for policy accomplishments has different consequences for men and women. A sample of American adult respondents were randomly assigned to listen to a speech by a hypothetical male or female American governor either passively describing or directly claiming responsibility for policy accomplishments. I then solicited respondents' evaluations of the governor's performance and traits, and measured the extent to which they attributed responsibility to the governor. This design enabled me to test whether directly claiming credit - as opposed to simply providing information about policy accomplishments - changes evaluations of men and women in systematically different ways. Because direct credit claiming is an assertive behavior that fits well with agentic masculine stereotypes, I expected it would incur significantly larger evaluative benefits for the male governor and result in penalties for the female governor. As in Parts I and II, I further expected that respondent gender and party would condition these effects. My broad expectations are largely not borne out by the results. I find that direct credit claiming incurs significant and similar benefits for the male and female governor, raising performance evaluations, future performance expectations, and electoral viability perceptions. However, I do find that direct credit claiming increases vote intention for the male governor, but this was not the case for the female governor.[1] I also find a gender affinity effect among women: specifically, direct credit claiming by the female governor is significantly more effective than direct credit claiming by the male governor among female respondents. Counter to expectations, I also find that direct credit claiming increases performance evaluations significantly more for the female governor than for the male governor among Republican respondents.

[1] However, the difference-in-differences for vote intention (from passive to direct for the male compared to passive to direct for the female) was not significant.

5.1 Major Themes

Several main themes emerge from these findings. First, the effect of gender on performance evaluation and the attribution of political responsibility is strongly moderated by partisanship. In the first experiment, I found that Democrats blamed the female governor significantly less for poor economic performance, while in the second I found that direct credit claiming by the female governor was significantly more effective among Republicans. These findings are somewhat in line with past work that suggests women and Democrats tend to favor female candidates, and view female legislators as more competent and ethical than male legislators (i.e., Costa and Schaffner, 2017; Thomsen and Swers, 2017; Fulton and Dhima, 2020a). It may be that Democrats are engaging in motivated reasoning in order to avoid blaming an officeholder they have an underlying preference for. While the results do not necessarily suggest that women and Democrats disproportionately reward female officeholders for strong performance in office, they do indicate that they may blame them less for negative outcomes. However, these findings also depart from past work that shows gendered double standards against female candidates tend to be strongest among Republican voters (i.e., Fulton and Dhima, 2020a; King and Matland, 2003; Burden et al., 2017; Burden and Ono, 2018).

Although I can only speculate as to why this might be, I believe the most plausible explanation is that Democrats and Republicans start out with different performance expectations for women. Because Democrats view women as more competent than men, strong performance by women might be expected - and thus not disproportionately rewarded - among Democratic respondents. Conversely, because Republicans tend to hold more traditional attitudes with regard to gender roles, strong performance information may in some sense correct low expectations. Additionally, better-than-expected performance can result in disproportionate rewards, as information that is surprising tends to factor more heavily into assessments. Future work that investigates performance expectations for men and women officeholders among Democrats and Republicans might help determine if this explanation holds.

Another possible explanation is that voters within the two parties respond differently to different types of performance. In Part II, I found that legislative activities impact men and women within the Democratic and Republican parties differently. This suggests not only that Republican and Democratic voters react to different legislative activities in different ways, but also that these reactions vary across officeholder gender. Hence, the disproportionate benefits associated with credit claiming by the female governor among Republicans in Part III could be because Republicans and Democrats respond differently to certain types of performance information. Although I cannot adjudicate that possibility here, future work might vary the nature of the performance information provided. For instance, might Democrats reward women more for information that emphasizes their collaboration across party lines, or highlights their role in committee hearings? Might

Republicans reward women more for information that emphasizes their constituency service or foreign affairs accomplishments?

Another major theme that emerges is a gender affinity effect in evaluation. In both experiments, I found that female officeholders fared better along some dimension among women. These findings comport well with past literature suggesting that attitudes toward - and evaluations of - women representatives are generally more favorable among women (i.e., Burns and Verba, 2001; Carroll, 2002; Lawless, 2004; Costa and Schaffner, 2018). Also consistent with past literature, my findings suggest that the effects of gender affinity are not uniform across all outcomes. For instance, in Part I I find that women blame the male governor significantly more than the female for negative performance (and that they blame her significantly less than men), but they do not reward her disproportionately for positive performance. In Part III, I find that among female respondents, direct credit claiming benefits the female governor significantly more than the male governor, but does not increase women's willingness to vote for her. Conversely, I found no significant gender affinity effects among men, although past literature indicates men tend to exhibit a preference for male candidates (Dolan, 2008; Sanbonmatsu, 2002a). It is possible that although men prefer male candidates, this baseline preference does not result in performance evaluation or attribution bias. In any case, these evaluative differences between men and women are also consistent with past literature that shows men and women react differently to other types of political behaviors, such as negative campaign attacks and campaign debates (Fridkin and Kenney, 2011; Fridkin et al., 2021). In sum, these findings both comport with and expand our understanding of the role of gender affinity in American politics.

Another consistent feature of this dissertation project was the finding that the expectations I drew from role congruity theory were generally not borne out by the empirical tests. In general, I expected that the mismatch between feminine and leadership stereotypes would generate evaluative penalties for women that would in turn lead to evaluation and attribution bias. Moreover, I expected that these negative effects would be amplified by factors that exacerbate lack of fit, such as more traditional beliefs among particular groups of voters, or more agentic behaviors by female officeholders (i.e., success; direct credit claiming). Instead, in Part I, I found that men and women were generally rewarded and punished for strong performance at similar rates. In Part II, I found that men and women experienced similar penalties for voting more conservatively than their districts. In Part III, I found that women benefited equally if not more than men from directly claiming credit for their accomplishments. Perhaps most surprising, I also found that direct credit claiming significantly *increased* perceptions of the stereotypically feminine traits of authenticity and warmth for both the male and female governor, and this effect was actually somewhat stronger for the female governor. Moreover, while I might have expected perceptions of feminine and masculine traits to act as counterweights, instead I found that direct credit claiming increased perceptions of *both* feminine and masculine traits at

statistically indistinguishable rates.

This could mean several different things for role congruity theory. First, it may be the case that role congruity theory - specifically with regard to women in leadership roles - simply needs updating. In recent years, the political ethos around gender has shifted, and the salience of gender in American politics has increased in the wake of the Hillary Clinton's 2016 loss, the MeToo movement, and the 'pink waves' of recent election cycles. Particularly among Democrats, women now seem to have a significant gender-based advantage (i.e., Burden and Ono, 2018). It is possible that prescriptive and descriptive stereotypes surrounding femininity and the suitability of women for leadership roles are also changing. It also may be that beliefs surrounding the nature of leadership roles and the traits necessary for success in those roles are changing. Particularly in a highly polarized, acrimonious climate, beliefs surrounding the importance of congeniality, kindness, and compassion to leadership may be changing.

Second, it is possible that rather than generating a backlash, the lack of fit between feminine and leadership stereotypes instead puts women at an advantage. This could be for several reasons. First, role incongruity may create a reverse double standard; if feminine stereotypes generate low performance expectations for women in politics, then when these expectations are exceeded, women may experience a disproportionate reward. Even if feminine stereotypes do not create low performance expectations, voters may believe that women face disproportionate, gender-based challenges in the political realm. For instance, a majority of American adults believe gender discrimination is a major reason why there aren't more women in top positions in business and politics (Pew Research Center, 2018). Thus, performance expectations may be low to begin with not because gender stereotypes generate negative beliefs about women's abilities, but because beliefs surrounding the sexist nature of American political institutions and power structures. Future research that examines performance expectations for men and women in the American electorate could shed further light on this possibility. Along similar lines, it may be the case that feminine trait stereotypes are sticky enough that they actually help women maintain a positive impression in masculine realms, or when engaging in agentic leadership behaviors. In other words, voters may be more give women the benefit of the doubt that they are *still* warm, kind, or authentic even when they engage in self-promotion or succeed in politics. It is also possible that women in politics are viewed differently than women in general; in other words, women in politics may be stereotyped *both* as women and as politicians. Thus, political women may be stereotyped as *both* compassionate and competent, warm and assertive, kind and strong. Indeed, the results from Part III would suggest that these feminine and masculine traits are not dichotomous.

Another possibility is that the effects of role incongruity are negative for some behaviors but not for others. The experimental manipulations I used in this project were neutral in tone; the governors did not display anger, emotion, brashness, combativeness, or defensiveness. The manipulations were also focused on

policy accomplishments rather than on officeholder traits. There is work that indicates that gender stereotypes must be activated to be harmful to women, and that an emphasis on feminine *traits* activates harmful gender stereotypes, while an emphasis on feminine policy issues does helps female candidates (Bauer, 2013, 2015b, 2020a). Much of the past work that has shown women are punished disproportionately for various behaviors has examined behaviors that are likely to activate beliefs about the officeholder's personal traits: corruption, negative campaigning, power-seeking, and displays of friendliness (Eggers et al., 2018; Krupnikov and Bauer, 2014; Cassese and Holman, 2017; Costa, 2020; Okimoto and Brescoll, 2010). Thus, the negative effects of role incongruity may be contingent on the activation of trait stereotypes. It is therefore possible that performance information may not *independently* activate harmful gender stereotypes.

5.2 Limitations and Opportunities

This dissertation has several major limitations. First, in the first experiment I do not include any information on the governor's partisanship. Respondents may assume the female governor is a Democrat, and the male governor a Republican, which may contribute to the female governor's advantage among Democrats and women, as women tend to be more politically liberal than men. In Part III, I make partisanship explicit. The results with regard to gender affinity from Part III assuage the concern that the results for women from Part I stem from the assumption that the female governor is a Democrat rather than from gender affinity. However, consistent with canonical American politics literature, the results from Part III show that co-partisanship is extremely important in shaping political evaluations and electoral support. The results from Part II further suggest that partisanship conditions which political behaviors matter for electoral support. Thus, the exclusion of partisanship from Part I is a limitation, as I cannot control for its effects.

Another major limitation centers on policy area. Voters tend to assume women (men) are better at handling policy areas that correspond with stereotypical feminine (masculine) traits, and these beliefs in turn impact the way voters evaluate legislator behavior. For instance, voters - particularly male Democrats - are more likely to take cues from women in the areas of birth control and sexual harassment (Anderson-Nilsson and Clayton, 2020; Clayton et al., 2019). Moreover, gender and party stereotypes overlap in ways that are likely to strengthen assumptions about which legislators are most competent in which areas. For instance, voters are likely to believe Republicans - particularly men - are best able to handle terrorism, while Democrats - especially women - are best able to handle healthcare and education (Holman et al., 2016, 2017; Winfrey and Schnoebelen, 2019; Sanbonmatsu and Dolan, 2009; Dolan, 2014). In Parts I and III, I use gender-neutral policy areas in order to isolate the effects of gender on evaluation sans independent of the potential effects of policy area. I am also unable to distinguish bill sponsorship patterns by policy area in Part II. This is an important limitation, as the legislation sponsored by men and women differs in systematic ways (Atkinson and

Windett, 2019). For instance, women sponsor significantly more bills related to women's issues, civil rights, and education (Volden et al., 2018), and female governors devote significantly more attention to social welfare issues (Heidbreder and Scheurer, 2012). Moreover, women's policy preferences tend to be better represented in districts represented by Democrats, while men's tend to be better reflected by Republican representatives (Griffin et al., 2012). This further suggests that policy area, gender, and party may interact to condition the ways voters evaluate the political behaviors of men and women. In sum, future research would do well to examine the ways gender conditions performance evaluation and the attribution of political responsibility in highly gendered policy areas.

Next, in this project I do not analyze the interactive effects of race and gender. In the experiment in Part I, I do not include any information about the governor's race, and in the credit claiming experiment in Part III, I use photos of White politicians. The data I use for Part II does also not distinguish by legislator race. This presents another limitation for this project, as well as an opportunity for future research. Like gender, both candidate and voter race has been shown to impact the way voters evaluate politicians (Junn and Masuoka, 2020). For instance, voters stereotype women and Black candidates as more liberal (**?**), and black women are the strongest supporters of black female candidates (Philpot and Walton Jr., 2007). Moreover, racial identity may overcome or condition the salience of gender identity in shaping vote choice (Gay and Tate, 2003). Women of color also exhibit different policy priorities in the legislation they sponsor, both compared to White women and to men of color, and also serve in political office at higher rates than White women (Barrett, 1995; Brown, 2014; Scola, 2013). Women of color are also likely to face systematically different campaign terrain related to both gender and racial stereotypes, which can intersect to pose added challenges (e.g., Scola, 2013; Krupnikov et al., 2016), as well as generating unique advantages (i.e., Bejarano, 2013; Fraga et al., 2008). For instance, in 2020, 41% of female Republican candidates and 55% of Democratic female candidates won their primary elections; for women of color, these figures drop to 29% and 46%. This suggests a promising avenue for future work that accounts for the interactive effects of race, gender, and partisanship.

Finally, it is possible that my largely null experimental results are a result of the fundamental limitations of survey experiments. In both experiments, I provide respondents with only a small sliver of information about a hypothetical governor. In the experiment in Part I, the governor's gender is communicated only using name and pronouns. In the experiment in Part III, I use photos and voice recordings. Yet, this still cannot possibly reflect the complex, multi-faceted nature of political evaluation and the ways voters form opinions about politicians.[2] Survey experiments are helpful in the sense that they can hold all else equal,

[2]For instance, as discussed above, traits more likely than policy issues to activate gender stereotypes, and these manipulations do not reflect differences in personality or tone.

isolating the effect of gender in the absence of everything else. But the absence of everything else does not reflect reality. Here, it bears emphasizing that in Part II, I show that women do not incur significant electoral benefits associated with the legislative activities on which they outperform men. Yet, this real-world result does not comport well with the equivalent rewards I find in the experiments in Parts I and III. For these reasons, I believe future work would benefit from the use of real officeholders in experimental manipulations. Although this approach introduces a host of potentially confounding variables, I believe it would be a more effective method of understanding the true nature of gendered political evaluation.

5.3 Implications

In this dissertation project, I find limited evidence of gendered double standards in the evaluation of officeholder performance in American politics. In Parts I and III, I show that men and women reward and punish political behavior in different ways, and that officeholder gender further conditions these patterns. In Part II, I find that different types of legislative activities impact men and women differently within the two political parties. This analysis indicates that although women vote more closely in line with their districts' ideological preferences and sponsor more legislation, they do not incur any significant electoral rewards associated with these patterns. This indicates some degree of gender bias in the political credit women receive for their accomplishments. However, this finding stands in contrast with the experimental results from Parts I and III, which either show no significant gender differences in performance evaluation or indicate that women are actually at an *advantage* relative to men in the way voters assess their performance and hold them accountable for outcomes. However, this finding from Part II is more consistent with the finding from Part III that although direct credit claiming improves performance and trait evaluations for both men and women, it only increases vote intention for men. Yet even this result is tenuous, given that the gender difference in the effect of direct credit claiming on vote intention is not significant.

Overall, it is not clear that gender bias in performance evaluation and the attribution of political responsibility explains the puzzle that motivated this dissertation project: namely, the mismatch between women's over-performance and electoral returns. However, it is also far from certain that it does not. It is my hope that in the future, research that incorporates variation in policy area and race, and deals explicitly with the interaction of gender and partisanship will fill in more pieces of this important puzzle.

References

Achen, C. H. and Bartels, L. M. (2004). Musical Chairs: Pocketbook Voting and the Limits of Democratic Accountability.

Achen, C. H. and Bartels, L. M. (2015). *Democracy for Realists*. Princeton University Press, Princeton and Oxford.

Anderson-Nilsson, G. and Clayton, A. (2020). Gender and Policy Persuasion. *Political Science Research & Methods*.

Anderson-Nilsson, G. and Clayton, A. (2021). Gender and Policy Persuasion. *Political Science Research & Methods*.

Anestaki, A., Sabharwal, M., Connelly, K., and Cayer, N. J. (2016). Race and Gender Representation in Presidential Appointments, SES, and GS Levels, During Clinton, Bush and Obama Administrations. *Administration and Society*, pages 1–32.

Ansolabehere, S. and Jones, P. E. (2010). Constituents' responses to congressional roll-call voting. *American Journal of Political Science*, 54(3):583–597.

Ansolabehere, S. D., Snyder, J., and III, C. S. (2001). Candidate positioning in u.s. house elections. *American Journal of Political Science*, 45(1):136–159.

Anzia, S. F. and Bernhard, R. (2019). How Does Gender Stereotyping Affect Women at the Ballot Box? Evidence from Local Elections in California, 1995-2016.

Anzia, S. F. and Berry, C. R. (2011). The Jackie (and Jill) Robinson effect: Why do congresswomen outperform congressmen? *American Journal of Political Science*, 55(3):478–493.

Armstrong, C. L. and Mcadams, M. J. (2009). Blogs of Information: How Gender Cues and Individual Motivations Influence Perceptions of Credibility. *Journal of Computer-Mediated Communication*, 14:435–456.

Astor, M. (2019). 'women are held to a higher standard,' klobuchar says at debate. *New York Times*, November 20.

Atkinson, M. L. and Windett, J. H. (2019). Gender Stereotypes and the Policy Priorities of Women in Congress. *Political Behavior*, 41(3):769–789.

Ballard, J. (2018). Standards are different for male and female politicians, americans say. *YouGov*, September 10.

Barnes, T. D. and Beaulieu, E. (2019). Women Politicians, Institutions, and Perceptions of Corruption. *Comparative Political Studies*, 52(1):134–167.

Barnes, T. D., Branton, R. P., and Cassese, E. C. (2017). A Reexamination of Women's Electoral Success in Open Seat Elections: The Conditioning Effect of Electoral Competition. *Journal of Women, Politics and Policy*, 38(3):298–317.

Barrett, E. J. (1995). The policy priorities of african american women in state legislatures. *Legislative Studies Quarterly*, 20(2):223–247.

Bartels, L. M. (2002). Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior*, 24(2):117–150.

Bartels, L. M. (2008). The Irrational Electorate. *Wilson Quarterly*, 32(4):44–50.

Bauer, N. M. (2013). Rethinking stereotype reliance: Understanding the connection between female candidates and gender stereotypes. *Politics and the Life Sciences*, 32(1):22–42.

Bauer, N. M. (2015a). Emotional, sensitive, and unfit for office? gender stereotype activation and support female candidates. *Political Psychology*, 36(6):691–708.

Bauer, N. M. (2015b). Who stereotypes female candidates? identifying individual differences in feminine stereotype reliance. *Politics, Groups, and Identities*, 3(6):691–708.

Bauer, N. M. (2017). The Effects of Counterstereotypic Gender Strategies on Candidate Evaluations. *Political Psychology*, 38(2):279–295.

Bauer, N. M. (2018). Untangling the Relationship between Partisanship, Gender Stereotypes, and Support for Female Candidates. *Journal of Women, Politics and Policy*, 39(1):1–25.

Bauer, N. M. (2019). A Feminine Advantage? Delineating the Effects of Feminine Trait and Feminine Issue Messages on Evaluations of Female Candidates. *Politics and Gender*, 16(2020):660–680.

Bauer, N. M. (2020a). A Feminine Advantage? Delineating the Effects of Feminine Trait and Feminine Issue Messages on Evaluations of Female Candidates. *Politics and Gender*, 16(3):660–680.

Bauer, N. M. (2020b). *The Qualifications Gap: Why Women Must Be Better than Men to Win Political Office*.

Bauer, N. M. (2020c). Shifting standards: How voters evaluate the qualifications of female and male candidates. *Journal of Politics*, 82(1):1–12.

Bejarano, C. E. (2013). *The Latina Advantage: Gender, Race, and Political Success*.

Belmore, S. M. (1987). Determinants of Attention During Impression Formation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3):480–489.

Bendor, J. (2002). Bounded Rationality in Political Science. *International Encyclopedia of the Social & Behavioral Sciences*, pages 1303–1307.

Bernhard, W. and Sulkin, T. (2018). *Legislative Style*.

Bernstein, A. G. (2000). The effects of message theme, policy explicitness, and candidate gender. *Communication Quarterly*, 48(2):159–173.

Berry, C. R. and Howell, W. G. (2007). Accountability and local elections: rethinking retrospective voting. *Journal of Politics*, 69(3):844–858.

Bianchini, S., Lissoni, F., , and Pezzoni, M. (2012). Instructor characteristics and students' evaluations of teaching effectiveness. *Journal of Engineering Education*, pages 1–20.

Biernat, M., Carnes, M., Filut, A., and Kaatz, A. (2019). Gender, Race, and Grant Reviews: Translating and Responding to Research Feedback. *Personality and Social Psychology Bulletin*, 46(1):140–154.

Biernat, M., Fuegen, K., and Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36(7):855–868.

Biernat, M., Tocci, M. J., and Williams, J. C. (2012). The language of performance evaluations: Gender-based shifts in content and consistency of judgment. *Social Psychological and Personality Science*, 3(2):186–192.

Bisgaard, M. (2015). Bias will find a way: Economic perceptions, attributions of blame, and partisan-motivated reasoning during crisis. *Journal of Politics*, 77(3):849–860.

Bligh, M. C. and Kohles, J. C. (2008). Negotiating gender role expectations: Rhetorical leadership and women in the US senate. *Leadership*, 4(4):381–402.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27–41.

Bos, A. L. (2011). Out of control: Delegates' information sources and perceptions of female candidates. *Political Communication*, 28(1):87–109.

Botelho, T. L. and Abraham, M. (2017). Pursuing Quality: How Search Costs and Uncertainty Magnify Gender-based Double Standards in a Multistage Evaluation Process. *Administrative Science Quarterly*, 62(4):698–730.

Box-Steffensmeier, J. M. and Grant, J. T. (1999). All in a day's work: The financial rewards of legislative effectiveness. *Legislative Studies Quarterly*, 24:511–523.

Box-Steffensmeier, J. M., Kimball, D. C., Meinke, S. R., and Tate, K. (2003). The Effects of Political Representation on the Electoral Advantages of House Incumbents. *Political Research Quarterly*, 56(3):259–270.

Branton, R., English, A., Pettey, S., and Barnes, T. D. (2018). The impact of gender and quality opposition on the relative assessment of candidate competency. *Electoral Studies*, 54(February):35–43.

Brewster, J. (2021). July 2.

Brians, C. L. (2005). Women for women?: Gender and party bias in voting for female candidates. *American Politics Research*, 33(3):357–375.

Brooks, D. J. (2011). Testing the Double Standard for Candidate Emotionality: Voter Reactions to the Tears and Anger of Male and Female Politicians. *The Journal of Politics*, 73(2):597–615.

Brown, N. E. (2014). *Sisters in the Statehouse: Black Women and Legislative Decision Making*.

Bruckmüller, S., Ryan, M. K., Rink, F., and Haslam, S. A. (2014). Beyond the glass ceiling: The glass cliff and its lessons for organizational policy. *Social Issues and Policy Review*, 8(1):202–232.

Burden, B. C. and Ono, Y. (2018). Testing Alternative Explanations for Gender-Office Congruency. In *Midwest Political Science Association*, pages 1–39.

Burden, B. C., Ono, Y., and Yamada, M. (2016). Reassessing Public Support for a Female President. *Journal of Politics*, 79(3):1073–1078.

Burden, B. C., Ono, Y., and Yamada, M. (2017). Reassessing public support for a female president. *Journal of Politics*, 79(3):1073–1078.

Burns, Nancy, K. L. S. and Verba, S. (2001). *The Private Roots of Public Action: Gender, Equality, and Political Participation*. Harvard University Press.

Butler, D. and Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: Implications for leadership evaluations. *Journal of Personality and Social Psychology*, 58(1):48–59.

Bystrom, D. and Dimitrova, D. V. (2014). Migraines, Marriage, and Mascara: Media Coverage of Michele Bachmann in the 2012 Republican Presidential Campaign. *American Behavioral Scientist*, 58(9):1169–1182.

Caleo, S., . H. M. E. (2010). Who gets the credit and who gets the blame?: Differential reactions to men and women's joint work. *Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA*.

Canes-Wrone, B., Brady, D., and Cogan, J. (2002). Out of step, out of office: Electoral accountability and house members' voting. *American Political Science Review*, 96(1):127–140.

Carey, T. E. and Lizotte, M.-K. (2019). Political experience and the intersection between race and gender. *Politics, Groups and Identities*, 7(2):243–266.

Carnevale, A. P., Smith, N., and Campbell, K. P. (2019). May the best woman win? education and bias against women in american politics. *Georgetown University Center on Education and the Workforce*.

Carroll, S. J. (2002). Representing women: Congresswomen's perceptions of their representational roles. In Rosenthal, C. S., editor, *Women Transforming Congress*.

Carroll, S. J. and Sanbonmatsu, K. (2013). Entering the mayor's office: Women's decisions to run for municipal positions. In Rose, M., editor, *Women and Executive Office: Pathways and Performance*. Lynne Rienner Publications, Inc.

Carson, J. L., Koger, G., Lebo, M. J., and Young, E. (2010). The electoral costs of party loyalty in congress. *American Journal of Political Science*, 54(3):598–616.

Cash, T. F., Gillen, B., and Burns, D. S. (1976). Sexism and "beautyism" in personnel consultant decision making. *Journal of Applied Psychology*, 62(3):301–310.

Cassese, E. C. and Holman, M. R. (2017). Party and Gender Stereotypes in Campaign Attacks. *Political Behavior*, pages 1–23.

Cecchi-Dimeglio, P. (2017). How Gender Bias Corrupts Performance Reviews, and What to Do About It. *Harvard Business Review*, April 12.

Cejka, M. A. and Eagly, A. H. (1994). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin*, 25(4):413–423.

Center for American Women in Politics (2021a). Current numbers.

Center for American Women in Politics (2021b). Women mayors in u.s. cities 2021.

Choi, S. (2011). Diversity and Representation in the Government: Analysis of the Trends of Federal Employment. *Public Personnel Management*, 40(1):25–47.

Chong, D. and Druckman, J. N. (2007). Framing Public Opinion in Competitive Democracies. *The American Political Science Review*, 101(4):637–655.

Clark, R. C., Ochs, H. L. A., and Frazier, M. (2013). Representative bureaucracy: The politics of access to policy-making positions in the federal executive service. *Public Personnel Management*, 42(1):75–89.

Clayton, A., O'Brien, D. Z., and Piscopo, J. M. (2019). All male panels? representation and democratic legitimacy. *American Journal of Political Science*, 63(1):113–129.

Clayton, A., Robinson, A. L., Johnson, M. C., and Muriaas, R. (2020). (how) do voters discriminate against women candidates? experimental and qualitative evidence from malawi. *Comparative Political Studies*, 53(3-4):601–630.

Clifford, S. (2020). Compassionate democrats and tough republicans: How ideology shapes partisan stereotypes. *Political Behavior*, 42(4):1269–1293.

Conn, A. B., Hanges, P. J., Sipe, W. P., and Salvaggio, A. N. (1999). The search for ambivalent sexism: A comparison of two measures. *Educational and Psychological Measurement*, 59(6):898–909.

Coppock, A. and McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research and Politics*, pages 1–14.

Costa, M. (2020). He Said, She Said: The Gender Double Bind in Legislator-Constituent Communication. *Politics and Gender*, (March).

Costa, M. and Schaffner, B. F. (2017). How gender conditions the way citizens evaluate and engage with their representatives. *Political Research Quarterly*, 71(1):46–58.

Costa, M. and Schaffner, B. F. (2018). How Gender Conditions the Way Citizens Evaluate and Engage with Their Representatives. *Political Research Quarterly*, 71(1):46–58.

Courtemanche, M. and Connor Green, J. (2020). A Fall from Grace: Women, Scandals, and Perceptions of Politicians. *Journal of Women, Politics and Policy*, 41(2):219–240.

CQ Press (2018). Election results.

Crumpacker, M. and Crumpacker, J. M. (2008). The U.S. Federal Senior Executive Service (SES) and Gender Diversity: Would Proposed Legislation Enhancing the Representational Diversity of Selection Panels Necessarily Increase the Number of Female Appointments to the SES? *Public Personnel Management*, 37:417–433.

Dancey, L. and Sheagley, G. (2016). Inferences Made Easy: Partisan Voting in Congress, Voter Awareness, and Senator Approval. *American Politics Research*, 44(5):844–874.

Deaux, K. and Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29(1):80–85.

Dholakia, R. R. and Sternthal, B. (1977). Highly credible sources: Persuasive facilitators or persuasive liabilities? *3*, 4(Journal of Consumer Research):223–232.

Dickhäuser, O. (2006). Perceived teachers' ability evaluations and boys' and girls' concepts of their mathematical ability in elementary school. *Psychologie in Erziehung und Unterricht*, 50:182–190.

Diekman, A. B., Eagly, A. H., and Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal ofExperimental Social Psychology*, 38:268–282.

Ditonto, T. (2017). A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns. *Political Behavior*, 39(2):301–325.

Ditonto, T. M., Hamilton, A. J., and Redlawsk, D. P. (2014a). Gender stereotypes, information search and voting behavior in political campaigns. *Political Behavior*, 36:335–238.

Ditonto, T. M., Hamilton, A. J., and Redlawsk, D. P. (2014b). Gender Stereotypes, Information Search, and Voting Behavior in Political Campaigns. *Political Behavior*, 36(2):335–358.

Dittmar, K. (2015). *Navigating gendered terrain: stereotypes and strategy in political campaigns*. Temple University Press, Philadelphia.

Dittmar, K. (2020). More republican women are running for office than ever, but how are they running? *Center for American Women in Politics*, June 23.

Dittmar, K. and Hill, C. (2020). A record number of women will serve in state legislatures in 2021. *Center for American Women in Politics*, December 4.

Dokoupil, T. (2011). Why female politicians are more effective. *Newsweek*, January 22.

Dolan, J. and Kropf, J. S. (2004). Credit claiming from the u.s. house: Gendered communication styles? *The International Journal of Press/Politics*, 9(1):41–59.

Dolan, K. (1997). Gender differences in support for women candidates: Is there a glass ceiling in american politics? *Women and Politics*, 17(2):27–41.

Dolan, K. (2004). *Voting for women: How the public evaluates women candidates*. Westview Press.

Dolan, K. (2008). Is there a "gender affinity effect" in american politics? *Political Research Quarterly*, 61:79–89.

Dolan, K. (2010). The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior*, 32(1):69–88.

Dolan, K. (2014). Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates. *Political Research Quarterly*, 67(1):96–107.

Dolan, K. and Sanbonmatsu, K. (2009). Gender Stereotypes and Attitudes Toward Gender Balance in Government. *American Politics Research*, 37(3):409–428.

Druckman, J. N., Peterson, E., and Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, 107(1):57–79.

Duerst-Lahti, G. (2002). Knowing congress as a gendered institution: Manliness and the implications of women in congress. In Rosenthal, C. S., editor, *Women transforming Congress*, pages 20–49. University of Oklahoma Press.

Dunaway, J., Lawrence, R. G., Rose, M., and Weber, C. R. (2013). Traits versus Issues. *Political Research Quarterly*, 66(3):715–726.

Eagly, A. H. (1983). Gender and Social Influence: A Social Psychological Analysis. *American Psychologist*, 38(9):971–981.

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

Eagly, A. H., Makhijani, M. G., and Klonsky, B. G. (1992). Gender and the Evaluation of Leaders: A Meta-Analysis. *Psychological Bulletin*, 111(1):3–22.

Eggers, A. C., Vivyan, N., and Wagner, M. (2018). Corruption, Accountability, and Gender: Do Female Politicians Face Higher Standards in Public Life? *Journal of Politics*, 80(1):321–326.

Escobar-Lemmon, M. and Taylor-Robinson, M. M. (2009). Getting to the Top: Career Paths of Women in Latin American Cabinets. *Political Research Quarterly*, 62(4):685–699.

Federalist 52 (1788). The federalist papers: No. 52. *From the New York Packet. Friday*, February 8, 1788.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66(2):127–148.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1):5–25.

Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. Yale, New Haven.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Social Cognition: Selected Works of Susan Fiske*, 38(6):889–906.

Fiske, S. T. (2012). Managing Ambivalent Prejudices: Smart-but-Cold and Warm-but-Dumb Stereotypes. *The ANNALS of the American Academy of Political and Social Science*, 639(1):33–48.

Fiske, S. T., Lin, M., and Neuberg, S. L. (1999). The continuum model: Ten years later. In Chaiken, S. and Trope, Y., editors, *Dual-process theories in social psychology*. Guilford Press, New York.

FiveThirtyEight (2020). When women run. *FiveThirtyEight*.

Foschi, M. (2000). Double Standards for Competence: Theory and Research. *Annual Review of Sociology*, 26:21–42.

Foschi, M. and Valenzuela, J. (2012). Who is the better applicant? Effects from gender, academic record, and type of decision. *Social Science Research*, 41(4):949–964.

Fowler, L. L. and Lawless, J. L. (2009). Looking for sex in all the wrong places: Press coverage and the electoral fortunes of gubernatorial candidates. *Perspectives on Politics*, 7(3):519–536.

Fox, R. L. and Oxley, Z. M. (2003a). Gender Stereotyping in State Executive Elections: Candidate Selection and Success. *The Journal of Politics*, 65(3):833–850.

Fox, R. L. and Oxley, Z. M. (2003b). Gender Stereotyping in State Executive Elections: Candidate Selection and Success. *The Journal of Politics*, 65(3):833–850.

Fraga, L. R., Lopez, L., Martinez-Ebers, V., and Ramirez, R. (2008). Gender and ethnicity: Patterns of electoral success and legislative advocacy among latina and latino state officials in four states. *Journal of Women, Politics Policy*, 28(3-4):121–145.

Francia, P. L. (2001). Early fundraising by nonincumbent female congressional candidates. *Women & Politics*, 23(1-2):7–20.

Fridkin, K. L., Gershon, S. A., Courey, J., and LaPlant, K. (2021). *Gender Differences in Emotional Reactions to the First 2016 Presidential Debate*, volume 43. Springer US.

Fridkin, K. L. and Kenney, P. J. (2009). The Role of Gender Stereotypes in U.S. Senate Campaigns. *Politics & Gender*, 5(2009):301–324.

Fridkin, K. L. and Kenney, P. J. (2011). Variability in Citizens' Reactions to Different Types of Negative Campaigns. *American Journal of Political Science*, 55(2):307–325.

Fridkin, K. L. and Kenney, P. J. (2015). *The Changing Face of Representation: the Gender of U.S. Senators and Constituent Communications*.

Fridkin, K. L., Kenney, P. J., and Woodall, G. S. (2009). Bad for men, better for women: The impact of stereotypes during negative campaigns. *Political Behavior*, 31(1):53–77.

Fulton, S. (2011). Running Backwards and in High Heels: The Gendered Quality Gap and Incumbent Electoral Success. *Political Research Quarterly*, 65(2):303–314.

Fulton, S. A. (2014). When Gender Matters: Macro-dynamics and Micro-mechanisms. *Political Behavior*, 36(3):605–630.

Fulton, S. A. and Dhima, K. (2020a). The Gendered Politics of Congressional Elections. *Political Behavior*.

Fulton, S. A. and Dhima, K. (2020b). The Gendered Politics of Congressional Elections. *Political Behavior*, (0123456789).

Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., and Verkuilen, J. (2007). Same facts, different interpretations: Partisan motivation and opinion on Iraq. *Journal of Politics*, 69(4):957–974.

Gasper, J. T. and Reeves, A. (2011). Make it rain? retrospection and the attentive electorate in the context of natural disasters. *American Journal of Political Science*, 55(2):340–355.

Gaucher, D., Friesen, J., and Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, 101.

Gay, C. and Tate, K. (2003). Doubly bound: The impact of gender and race on the politics of black women. *Political Psychology*, 19(1):169–184.

Gimenez, A. T., Karpowitz, C. F., Monson, J. Q., and Preece, J. R. (2016). The Double Bind Still Constricts: Gendered Self-Presentation and Electoral Success in Republican Neighborhood Caucuses. In *Presented at the Annual Meeting of the Southern Political Science Association, San Juan, Puerto Rico*, pages 1–37.

Glick, P. and Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism Peter. *Journal of Personality and Social Psychology*, 70(450):491–512.

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In Greenwald, A., Brock, T. C., and Ostrom, T. M., editors, *Psychological Foundations of Attitudes*, pages 147–170. Academic Press, New York.

Grethen, D. (2018). 2018: The next year of the woman. *The Seattle Times*, October 1.

Grier, K. B. and McGarrity, J. P. (2002). Presidential party, incumbency, and the effects of economic fluctuations on House Elections, 1916-1996. *Public Choice*, 110(1-2):143–162.

Griffin, J. D. and Flavin, P. (2007). Racial differences in information, expectations, and accountability. *Journal of Politics*, 69(1):220–236.

Griffin, J. D., Newman, B., and Nickerson, D. W. (2019). A God of Vengeance and of Reward? Voters and Accountability. *Legislative Studies Quarterly*, 44(1):133–162.

Griffin, J. D., Newman, B., and Wolbrecht, C. (2012). A gender gap in policy representation in the u.s. congress? *Legislative Studies Quarterly*, 37(1):35–66.

Grimmer, J., Messing, S., and Westwood, S. J. (2012). How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation. *American Political Science Review*, 106(4):703–719.

Grimmer, J., Westwood, S. J., and Messing, S. (2014). *The Impression of Influence : Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press, Princeton, NJ, 1 edition.

Guisinger, A. (2009). Determining trade policy: Do voters hold politicians accountable? *International Organization*, 63(3):533–557.

Guthrie, S. (2018). *He Said, They Did: Using Executive Rhetoric to Understand Interbranch Relationships in the American States*. PhD thesis, American University.

Hayes, D. (2011). When gender and party collide: Stereotyping in candidate trait attribution. *Politics and Gender*, 7(2):133–165.

Hayes, D. and Lawless, J. (2016). There's much less gender bias in politics than you think. here's why.

Hayes, D. and Lawless, J. L. (2015). A non-gendered lens? Media, voters, and female candidates in contemporary congressional elections. *Perspectives on Politics*, 13(1):95–118.

Hayes, D., Lawless, J. L., and Baitinger, G. (2014). Who cares what they wear? Media, gender, and the influence of candidate appearance. *Social Science Quarterly*, 95(5):1194–1212.

Healy, A. and Malhotra, N. (2009). Myopic voters and natural disaster policy. *American Political Science Review*, 103(3):387–406.

Healy, A. J., Malhotra, N., Mo, C. H., and Laitin, D. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences of the United States of America*, 107(29):12804–12809.

Heidbreder, B. and Scheurer, K. F. (2012). Gender and the gubernatorial agenda. *State and Local Government Review*, 45(1):3–13.

Heilman, Madeline E., C. B. and Martell, R. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality*, 10:237–252.

Heilman, M. E. (1995). Sex Stereotypes and Their Effects in the Workplace: What We Know and What We Don't Know. *Journal of Social Behavior and Personality*, 10(6):3–26.

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*, 32:113–135.

Heilman, M. E. and Chen, J. J. (2005). Same behavior, different consequences: Reactions to men's and women's altruistic citizenship behavior. *Journal of Applied Psychology*, 90(3):431–441.

Heilman, M. E. and Haynes, M. C. (2005). No Credit Where Credit Is Due: Attributional Rationalization of Women's Success in Male-Female Teams. *Journal of Applied Psychology*, 90(5):905–916.

Heilman, M. E. and Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, 92(1):81–92.

Heilman, M. E., Wallen, A. S., Fuchs, D., and Tamkins, M. M. (2004). Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks. *Journal of Applied Psychology*, 89(3):416–427.

Hengel, E. (2017). Publishing while Female: Are women held to higher standards? Evidence from peer review.

Herrnson, P. S., Lay, J. C., and Stokes, A. K. (2003). Women Running "as Women": Candidate Gender, Campaign Issues, and Voter-Targeting Strategies. *Journal of Politics*, 65(1):244–255.

Hilton, J. L., Klein, J. G., and von Hippel, W. (1991). Attention allocation and impression formation. *Personality and Social Psychology Bulletin*, 17(5):548–559.

Hogarth, R. M. (1987). Judgement and Choice. In *Judgement and Choice*, chapter 1, 10, pages 1–11. John Wiley & Sons, New York, 2 edition.

Hollibaugh, G. E., Rothenberg, L. S., and Rulison, K. K. (2013). Does It Really Hurt to Be Out of Step? *Political Research Quarterly*, 66(4):856–867.

Holman, M. R. and Mahoney, A. (2018). Stop, Collaborate, and Listen: Women's Collaboration in US State Legislatures. *Legislative Studies Quarterly*.

Holman, M. R., Merolla, J. L., and Zechmeister, E. J. (2011). Sex, stereotypes, and security: An experimental study of the effect of crises on assessments of gender and leadership. *Journal of Women, Politics, & Policy*, 32(3):173–192.

Holman, M. R., Merolla, J. L., and Zechmeister, E. J. (2016). Terrorist Threat, Male Stereotypes, and Candidate Evaluations. *Political Research Quarterly*, 69(1):134–147.

Holman, M. R., Merolla, J. L., and Zechmeister, E. J. (2017). Can experience overcome stereotypes in times of terror threat? *Research & Politics*, 4(1):205316801668812.

Hovland, C. I., Janis, I. L., and Kelley, H. H. (1953). *Communication and Persuasion: Psychological Studies of Opinion Change*. Yale University Press, New Haven, CT.

Huang, L. M. and Sherman, J. W. (2018). Attentional Processes in Social Perception. In *Advances in Experimental Social Psychology*, volume 58, pages 199–241. Elsevier Inc.

Huddy, L. and Capelos, T. (2002). Gender stereotyping and candidate evaluation: Good news and bad news for women politicians. In Ottati, V. C., Tindale, S., Edwards, J., Bryant, F. B., Heath, L., Suarez-Balcazar, Y., and J., P. E., editors, *The Social Psychology of Politics*.

Huddy, L., Mason, L., and Aarøe, L. (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, 109(1):1–17.

Huddy, L. and Terkildsen, N. (1993a). The consequences of gender stereotypes for women candidates at different levels and types of office. *Political Research Quarterly*, 46(3):503–525.

Huddy, L. and Terkildsen, N. (1993b). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, 37(1):119–147.

Inesi, M. E. and Cable, D. M. (2015). When Accomplishments Come Back to Haunt You: The Negative Effect of Competence Signals on Women's Performance Evaluations. *Personnel Psychology*, 68(3):615–657.

Iyengar, S. (1987). Television news and citizens' explanations for national affairs. *American Political Science Review*, 81:815–832.

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, 12:19–40.

Jacobson, G. C. and Carson, J. L. (2020). *The Politics of Congressional Elections*. Rowman Littlefield Publishers, Lanham, Maryland.

Jensen, N. M. and Maleksky, E. J. (2018). *Incentives to Pander: How politicians Use Corporate Welfare for Political Gain*. Cambridge University Press, New York.

Jeydel, A. and Taylor, A. (2003). Are women legislators less effective? evidence from the u.s. house in the 103rd-105th congress. *Political Research Quarterly*, 56(1):19–27.

Jones, P. E. (2011). Which buck stops here? Accountability for policy positions and policy outcomes in congress. *Journal of Politics*, 73(3):764–782.

Jones, P. E. (2014). Does the descriptive representation of gender influence accountability for substantive representation? *Politics and Gender*, 10(2):175–199.

Junn, J. and Masuoka, N. (2020). The Gender Gap Is a Race Gap: Women Voters in US Presidential Elections. *Perspectives on Politics*, 18(4):1135–1145.

Kahn, K. F. (1993). Gender Differences in Campaign Messages: The Political Advertisements of Men and Women Candidates for U.S. Senate. *Political Research Quarterly*, 46(3):481–502.

Kahn, K. F. (1996). *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. Columbia University Press, New York.

Kamarck, E. (2018). 2018: Another 'year of the woman'. *Brookings*, November 7.

Karol, D. and Miguel, E. (2007). Iraq war casualties and the 2004 u.s. presidential election. *Journal of Politics*, 69(3):633–648.

Karpowitz, C. F., Mendelberg, T., and Shaker, L. (2012). Gender Inequality in Deliberative Participation. *American Political Science Review*, 106(3):533–547.

Key, V. J. (1964). *Public Opinion and American Democracy*. Knopf, New York.

King, D. C. and Matland, R. E. (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*, 31:595–612.

Kitchens, K. E. and Swers, M. L. (2016). Why Aren't There More Republican Women in Congress? Gender, Partisanship, and Fundraising Support in the 2010 and 2012 Elections. *Politics & Gender*, 12:1–29.

Klar, S. (2018). When Common Identities Decrease Trust: An Experimental Study of Partisan Women. *American Journal of Political Science*, 62(3):610–622.

Knuckey, J. (2005). A new front in the culture war?: Moral traditionalism and voting behavior in u.s. house elections. *American Politics Research*, 33:645–671.

Koch, J. (2000). Do citizens apply gender stereotypes to infer candidates' ideological orientations? *The Journal of Politics*, 62:414–429.

Koch, J. W. (2002). Gender Stereotypes and Citizens' Impressions of House Candidates' Ideological Orientations. *American Journal of Political Science*, 46(2):453.

Koch, S. C. (2005). Evaluative affect display toward male and female leaders of task-oriented groups. *Small Group Research*, 36(6):678–703.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., and Ristikari, T. (2011). Are leader stereotypes masculine? a meta-analysis of three research paradigms. *Psychological Bulletin*, 137:616–642.

Kramer, G. (1971). Short-term fluctuations in u.s. voting behavior, 1896-1964. *The American Political Science Review,*, 65(1):131–143.

Krook, M. L. and O'Brien, D. Z. (2012). All the President's Men? The Appointment of Female Cabinet Ministers Worldwide. *Journal of Politics*, 74(3):840–855.

Krupnikov, Y. and Bauer, N. M. (2014). The Relationship Between Campaign Negativity, Gender and Campaign Context. *Political Behavior*, 36(1):167–188.

Krupnikov, Y., Piston, S., and Bauer, N. (2016). Saving face: Identifying voter responses to black candidates and female candidates. *Political Psychology*, 37(2):253–273.

Kulich, C., Ryan, M. K., and Haslam, S. A. (2007). Where is the romance for women leaders? The effects of gender on leadership attributions and performance-based pay. *Applied Psychology*, 56(4):582–601.

Kulich, C., Trojanowski, G., Ryan, M. K., Haslam, S. A., and Renneboog, L. D. R. (2011). Who Gets the Carrot and Who Gets the Stick? Evidence of Gender Disparities in Executive Renumeration. *Strategic Management Journal*, 32:301–321.

Lawless, J. and Hayes, D. (2015). A non-gendered lens? media, voters, and female candidates in contemporary congressional elections. *Perspectives on Politics*, 13(1):94–118.

Lawless, J. L. (2004). Politics of Presence? Congresswomen and Symbolic Representation. *Political Research Quarterly*, 57(1):81–99.

Lawless, J. L. (2015). Female Candidates and Legislators. *Annual Review of Political Science*, 18(1):349–366.

Lawless, J. L. and Pearson, K. (2008). The primary reason for women's underrepresentation? Reevaluating the conventional wisdom. *Journal of Politics*, 70(1):67–82.

Lazarus, J. and Steigerwalt, A. (2018). *Gendered Vulnerability: How Women Work Harder to Stay in Office*. University of Michigan Press, Ann Arbor.

Leeper, T. J. and Slothuus, R. (2014). Political Parties, Motivated Reasoning, and Public Opinion Formation. *Political Psychology*, 35(1):129–156.

Lei, R. F. and Bodenhousen, G. V. (2018). Economic anxieties undermine support for female (but not male) political candidates. *Analyses of Social Issues and Public Policy*, 18(1):244–264.

Lenz, G. S. (2012). *Follow the Leader?: How Voters Respond to Politicians' Policies and Performance*. The University of Chicago Press, Chicago and London.

Lerman, A. E. and Sadin, M. L. (2016). Stereotyping or Projection? How White and Black Voters Estimate Black Candidates' Ideology. *Political Psychology*, 37(2):147–163.

Leyden, K. M. and Borrelli, S. A. (1995). The Effect of State Economic Conditions on Gubernatorial Elections: Does Unified Government Make a Difference? *Political Research Quarterly*, 48(2):275–290.

Lips, K. A. (2020). The year of the republican woman. *The Hill*, November 8.

Lowry, R. C., Alt, J. E., and Ferree, K. E. (1998). Fiscal policy outcomes and electoral accountability in american states. *American Political Science Review 92(4)*, 92(4):759–74.

Lupia, A. (1994). Shortcuts versus encyclopedias: Information and voting behavior in california insurance reform elections. *American Political Science Review*, 88(1):63–76.

Lupia, A. (2015). *Uninformed: Why People Seem to Know So Little about Politics and What We Can Do About It*. Oxford University Press, New York.

Luque, D., Vadillo, M. A., Pelley, M. E. L., and Beesley (2017). Prediction and uncertainty in associative learning: Examining controlled and automatic components of learned attentional biases.

MacNell, L., Driscoll, A., and Hunt, A. N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291–303.

Martin, L. L. (2016). Gender, Teaching Evaluations, and Professional Success in Political Science. *PS: Political Science & Politics*, (April):313–319.

Martin, S. R. (2017). Research: Men Get Credit for Voicing Ideas, but Not Problems. Women Don't Get Credit for Either. *Harvard Business Review*, November 2.

Mayhew, D. (1974). *Congress: The Electoral Connection*. Yale University Press, New Haven, 1 edition.

Mayhew, D. (2004). *Congress: The Electoral Connection*. Yale University Press, New Haven, 2 edition.

McClean, E., Martin, S. R., Emich, K. J., and Woodruff, T. (2018). The social consequences of voice: An examination of voice type and gender on status and subsequent leader emergence. *Academy of Management Journal*.

McCroskey, James C., P. R. H. and Weiner, A. N. (1974). The effect of interaction behavior on source credibility, homophily, and interpersonal attraction. *Human Communication Research*, 1(1):42–52.

Mcdermott, M. L. (1998). Race and Gender Cues in Low-Information Elections. *Political Research Quarterly*, 51(4):895–918.

McGarrity, J. P. (2008). A Test of Prospective Voting in House Elections Using Leading Economic Indicators. *Journal of Political Science*, 36(1).

McGraw, K. (2001). Political accounts and attribution processes. In Kuklinski, J. H., editor, *Citizens and Politics: Perspectives from Political Psychology*. Cambridge University Press, Cambridge, UK.

Mendelberg, T., Karpowitz, C. F., and Oliphant, J. B. (2014). Gender Inequality in Deliberation: Unpacking the Black Box of Interaction. *Perspectives on Politics*, 12(1):18–44.

Miller, C. (2016). The problem for women is not winning. it's deciding to run. *The New York Times: The Upshot*, October 25.

Miller, J. A. and Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology*, 28(4):283–298.

Minkin, R. (2020). Most americans support gender equality, even if they don't identify as feminists. *Pew Research Center*, July 14.

Miquel, G. P. I. and Snyder, J. M. (2006). Legislative Effectiveness and Legislative Careers. *Legislative Studies Quarterly*, 31(3):347–381.

Mitchell, K. M. and Martin, J. (2018). Gender Bias in Student Evaluations. *PS - Political Science and Politics*, 51(3):648–652.

Mo, C. H. (2014). The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters. *Political Behavior*, 37(2):357–395.

Neal, T. M., Guadagno, R. E., Eno, C. A., and Brodsky, S. L. (2012). Warmth and Competence on the Witness Stand: Implications for the Credibility of Male and Female Expert Witnesses. *The Journal of the American Academy of Psychiatry and the Law*, 40(4):488–497.

Nieva, V. and Gutek, B. A. (1980). Sex Effects on Evaluation. *The Academy of Management Review*, 5(2):267–276.

Nyhan, B., McGhee, E., Sides, J., Masket, S., and Greene, S. (2012). One Vote Out of Step? The Effects of Salient Roll Call Votes in the 2010 Election. *American Politics Research*, 40(5):844–879.

Okimoto, T. G. and Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, 36(7):923–936.

Olohan, M. M. (2020). 'the year of the republican woman': Here's a list of gop women who won 2020 elections. *The Daily Caller*, November 4.

Ondercin, H. L. and Welch, S. (2009). Comparing Predictors of Women's Congressional Election Success: Candidates, Primaries, and the General Election. *American Politics Research*, 37(February 2016):593–613.

Ono, Y. and Burden, B. C. (2017). The Contingent Effects of Candidate Sex on Voter Choice. In *Paper prepared for the 2017 Annual Meeting of the American Political Science Association*, pages 1–20.

Orth, D. A. (2001). Accountability in a Federal System: The Governor, the President, and Economic Expectations. *State Politics & Policy Quarterly*.

Parker, D. C. and Goodman, C. (2009). Making a good impression: Resource allocation, home styles, and Washington work. *Legislative Studies Quarterly*, 34(4):493–524.

Pearson, K. and Dancey, L. (2011). Elevating women's voices in congress: Speech participation in the house of representatives. *Political Research Quarterly*, 64(4):910–923.

Pearson, K. and McGhee, E. (2013). What It Takes to Win: Questioning "Gender Neutral" Outcomes in U.S. House Elections. *Politics & Gender*, 9(04):439–462.

Pew Research Center (2015). Women and leadership. *Social Trends*, January 14.

Pew Research Center (2018). Women and leadership. *Social Trends*, September 20.

Philpot, T. S. and Walton Jr., H. (2007). One of Our Own: Black Female Candidates and the Voters Who Support Them. *American Journal of Political Science*, 51(1):49–62.

Player, A., de Moura, G. R., Leite, A. C., Abrams, D., and Tresh, F. (2019). Overlooked leadership potential: The preference for leadership potential in job candidates who are men vs. women. *Frontiers In Psychology*.

Popkin, S. (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. University of Chicago Press, Chicago.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decade's evidence. *Journal of Applied Social Psychology*, 34(2):243–281.

Powell, G. B. and Whitten, G. (1993). A cross-national analysis of economic voting: Taking account of the political context. *Rudolph, Thomas J. 2003. "Institutional Context and the Assignment of Political Responsibility." Journal of Politics 65(1): 190–215.*, 37(2):391–414.

Pratto, F. and John, O. P. (1991). Automatic vigilance: The Attention-grabbing Power of Negative Social Information. *Journal of Personality and Social Psychology*, 61(3):380–391.

Prentice, D. A. and Carranza, E. (2004). Sustaining Cultural Beliefs in the Face of Their Violation: The Case of Gender Stereotypes. *Psychological Foundations of Culture*.

Ragsdale, L. and Cook, T. (1987). Representatives' actions and challengers' reactions: Limits to candidate connections in the house. *American Journal of Political Science*, 31(1):45–81.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12):4403–4408.

Ridgeway, C. L. (2001). Gender, Status, and Leadership. *Journal of Social Issues*, 57(4):637–655.

Roberts, D. C. and Utych, S. M. (2020). Linking Gender, Language, and Partisanship: Developing a Database of Masculine and Feminine Words. *Political Research Quarterly*, 73(1):40–50.

Romero, G. J. and Garza, R. T. (1986). Attributions for the Occupational Success/Failure of Ethnic Minority and Nonminority Women. *Sex Roles*, 14(7/8):445–452.

Rose, M. (2013). *Executive Women: Pathways and Performance*. Lynne Rienner Publishers, Boulder, CO.

Rosenthal, C. S. (1998). *When Women Lead: Integrative Leadership in State Legislatures*. Oxford University Press.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypic impression management. *Journal of Personality and Social Psychology*, 74:629–645.

Rudman, L. A. and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57:743–762.

Rudolph, T. J. (2003). Institutional context and the assignment of political responsibility. *Journal of Politics*, 65(1):190–215.

Sabharwal, M. (2013). From glass ceiling to glass cliff: Women in senior executive service. *Journal of Public Administration Research and Theory*, 24:399–426.

Sanbonmatsu, K. (2002a). Gender Stereotypes and Vote Choice. *American Journal of Political Science*, 46(1):20–34.

Sanbonmatsu, K. (2002b). Political parties and the recruitment of women to state legislatures. *The Journal of Politics*, 64(3):791–809.

Sanbonmatsu, K. and Dolan, K. (2009). Do Gender Stereotypes Transcend Party? *Political Research Quarterly*, 62(3):485–494.

Sanbonmatsu, K. and Rogers, K. (2020). Advancing research on gender and gubernatorial campaign finance. *Journal of Women, Politics, and Policy*.

Schneider, M. C. and Bos, A. L. (2016). The Interplay of Candidate Party and Gender in Evaluations of Political Candidates. *Journal of Women, Politics and Policy*, 37(3):274–294.

Schwarz, S., Hunt, W., and Coppock, A. (2018). What have we learned about gender from candidate choice experiments? a meta-analysis of 30 factorial survey experiments. Working Paper.

Scola, B. (2013). *Gender, Race, and Office Holding in the United States*.

Sekaquaptewa, D. (2011). Discounting Their Own Success: A Case for the Role of Implicit Stereotypic Attribution Bias in Women's STEM Outcomes. *Psychological Inquiry*, 22:291–295.

Shaffner, B. F. (2005). Priming gender: Campaigning on women's issues in u.s. senate elections. *American Journal of Political Science*.

Shah, P., Scott, J., and Juenke, E. G. (2018). Women of color candidates: examining emergence and success in state legislative elections. *Politics, Groups and Identities*, 7(2):429–443.

Shames, S. L. (2018). Higher Hurdles for Republican Women: Ideology, Inattention, and Infrastructure. In Och, M. and Shames, S. L., editors, *The Right Women: Republican Party Activists, Candidates, and Legislators*. Praeger, Santa Barbara, CA.

Simon, S. and Hoyt, C. L. (2008). Exploring the gender gap in support for a woman president. *Analyses of Social Issues and Public Policy*, 8:157–181.

Skowronski, J. J. and Carlston, D. E. (1987). Social Judgment and Social Memory The Role of Cue. *Journal of Personality and Social Psychology*, 52(4):689–699.

Slothuus, R. and de Vreese, C. H. (2010). Political Parties, Motivated Reasoning, and Issue Framing Effects. *The Journal of Politics*, 72(3):630–645.

Snyder, K. (2014). The abrasiveness trap: High-achieving men and women are described differently in reviews. *Fortune*, August 26.

Stokes, D. E. (1963). Spatial models of party competition. *he American Political Science Review*, 57(2):368–377.

Streb, M. J., Burrell, B., Frederick, B., and Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly*, 72(1):76–89.

Sulkin, T. (2005). *Issue Politics in Congress*. Cambridge University Press, New York.

Sulkin, T., Testa, P., and Usry, K. (2015a). What Gets Rewarded? Legislative Activity and Constituency Approval. *Political Research Quarterly*, 68(4):690–702.

Sulkin, T., Testa, P., and Usry, K. (2015b). What Gets Rewarded? Legislative Activity and Constituency Approval. *Political Research Quarterly*, 68(4):690–702.

Swers, M. (2013). *Women in the Club: Gender and Policy Making in the Senate*. University of Chicago Press, Chicago.

Swim, J. K. and Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, 22:507–519.

Taber, C. S., Lodge, M., and Glathar, J. (2001). The motivated construction of political judgments. In Kuklinski, J. H., editor, *Citizens and Politics: Perspectives from Political Psychology*. Cambridge University Press, Cambridge, UK.

Tate, K. (2003). *Black Faces in the Mirror: African Americans and Their Representatives in the U.S. Congress*.

Taylor, A. J. (2017). Which U.S. House Members Present Their Legislative Records? Models of Electoral Accountability and the Content of Press Releases. *Congress and the Presidency*, 44(1):102–119.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1):67–85.

Teele, D. L., Kalla, J., and Rosenbluth, F. (2018). The ties that double bind: Social roles and women's underrepresentation in politics. *American Political Science Review*, 112(3):525–541.

Teele, D. L. and Thelen, K. (2017). Gender in the Journals: Publication Patterns in Political Science. *PS: Political Science & Politics*, (April):433–447.

The Cook Political Report (2020). Election results.

Thomsen, D. M. (2019). Which women win? partisan changes in victory patterns in us house elections. *Politics, Groups and Identities*, 7(2):412–428.

Thomsen, D. M. (2020). Ideology and Gender in U.S. House Elections. *Political Behavior*, 42:415–442.

Thomsen, D. M. and Sanders, B. K. (2020). Gender Differences in Legislator Responsiveness. *Perspectives on Politics*, 18(4):1017–1030.

Thomsen, D. M. and Swers, M. L. (2017). Which Women Can Run? Gender, Partisanship, and Candidate Donor Networks. *Political Research Quarterly*, 70(2):449–463.

Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics*, 41(2):191–207.

Tosi, H. L. and Einbender, S. W. (1985). The Effects of the Type and Amount of Information in Sex Discrimination Research: A Meta-Analysis. *Academy of Management Journal*, 28(2):712–723.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

Ulbig, S. G. (2007). Gendering municipal government: Female descriptive representation and feelings of political trust. *Social Science Quarterly*, 88(5):1106–1123.

U.S. Census Bureau (2018). Economic indicators.

Volden, C., Wiseman, A. E., and Wittmer, D. E. (2013). When Are Women More Effective Lawmakers Than Men? *The American Economic Review*, 57(2):326–341.

Volden, C., Wiseman, A. E., and Wittmer, D. E. (2018). Women's Issues and Their Fates in the US Congress. *Political Science Research and Methods*, 6(4):679–696.

Voteview (2020). Realtime nominate ideology and related data.

Vraga, E. K. (2017). Which Candidates Can Be Mavericks? The Effects of Issue Disagreement and Gender on Candidate Evaluations. *Politics and Policy*, 45(1):4–30.

Watson, D. and Moreland, A. (2014). Perceptions of Corruption and the Dynamics of Women's Representation. *Politics & Gender*, 10(03):392–412.

Weiner, B., I. F. A. K. S. R. and Rosenbaum, R. (1971). *Perceiving the Causes of Success and Failure*. General Learning Press, Morristown, N.J.

Wheeless, L. R. (1974). The relationship of attitude and credibility to comprehension and selective exposure. *Western Journal of Communication*, 38(2):88–97.

Wilson, J. M. and Gronke, P. (2000). Concordance and Projection in Citizen Perceptions of Congressional Roll-Call Voting. *Legislative Studies Quarterly*, 25(3):445.

Winfrey, K. L. and Schnoebelen, J. M. (2019). Running as a Woman (or Man): A review of research on political communicators and gender stereotypes. *Review of Communication Research*, 7(2019):109–138.

Winter, N. J. G. (2010). Masculine Republicans and Feminine Democrat Gender and Americans' Explicit and Implicit of the Political Parties. *Political Behavior*, 32(32):587–618.

Witteman, H. O., Hendricks, M., and Straus, S. (2019). Are gender gaps due to evaluations of the applicant or the science? a natural experiment at a national funding agency. *The Lancet*, 393.

Yarvin, J. (2017). Will 2018 be the next 'year of the woman?'. *PBS News Hour*, November 7.