What Have We Learned About Gender From Candidate Choice Experiments?
A Meta-analysis of 42 Factorial Survey Experiments

Susanne Schwarz and Alexander Coppock[*]

August 13, 2019

Abstract

Candidate choice survey experiments in the form of conjoint or vignette experiments have become a standard part of the political science toolkit for understanding voters' multidimensional preferences over candidates. By our count, the most common attribute studied in these experiments is candidate gender. We collect 42 studies and reanalyze them using a standardized statistical approach. We find that respondents prefer women by an average of 2 percentage points. We find some evidence of heterogeneity as the female preference appears to be somewhat larger for white (versus non-white) candidates, among female (versus male) respondents, and among Democrats and Independents (versus Republicans). We demonstrate the external validity of these survey experimental estimates with an observational analysis of putatively male and female candidates in low-salience elections. Our results add to the growing body of evidence that voter preferences are not a major factor explaining the persistently low rates of women in elected office.

---

[*]Susanne Schwarz is a Doctoral Candidate in the Department of Politics, Princeton University. Alexander Coppock is Assistant Professor of Political Science, Yale University.

Do voters discriminate against women running for office? Owing to the real, pervasive, and pernicious biases against women in many areas of society, a reasonable guess would be that voters tend to prefer male over female political candidates. In the United States, the 2016 Presidential election in particular has confirmed for many observers that women seeking higher office face unique challenges, including gender-based discrimination.[1]

Setting aside that particular election, the empirical evidence of voter bias against female candidates (conditional on running) is surprisingly thin. Some of the earlier observational studies on women in politics indeed reported gender gaps in electoral outcomes. In the 1960s and early 1970s, men tended to outpoll women in many Western democracies, albeit by relatively small margins (e.g., Darcy and Schramm 1977; Hills 1981; Kelley and McAllister 1984). For example, Kelley and McAllister (1984) reported that, conditional on party affiliation, vote margins for women in Britain and Australia were on average 2.5 percent points and 4 percentage points lower, respectively. However, observational data from the late 1970s onwards suggest these gender-based discrepancies in elections have by and large dissipated: when women run for political office, they are not less likely to win than their male counterparts. In the United States, for example, women typically garner the same level of support as male candidates with similar characteristics, both in primary and general elections (e.g., Zipp and Plutzer 1985; Welch et al. 1987; Fox 2005). Likewise, an Australian study that tracked vote share for female parliamentary candidates between 1903 and 2004, found that sex-based discrepancies have diminished drastically after the 1980s, and win rates for women today are virtually identical to those of men (King and Leigh 2010). An analysis of elections to the U.S. House of Representatives from 1980 to 2012 comes to similar conclusions, even when comparing ideologically-similar men and women (Thomsen 2018). And even in 2016, when concerns about gender discrimination and misogyny in politics were heightened in the U.S. context, women in non-Presidential races encountered little resistance from voters. Democratic women running for the U.S. House of Representatives outpolled their male counterparts in both primary and general elections while Republican women performed only slightly worse than male candidates (Dittmar 2017).

These analyses of electoral returns yield the important descriptive finding that conditional on running, men and women tend to win elected office at similar rates. It is tempting to give this finding a causal interpretation, i.e., that the average causal effect of being a woman (versus a man) on the probability of winning an election is close to zero. However, the men and women who successfully arrive on the ballot are different from each other in more

---

[1]See for example: Burleigh (2016); Crockett (2016).

ways than gender alone. We have some evidence that women who run are of higher quality than their male counterparts (Anzia and Berry 2011; Fulton 2011). Further, we have some evidence that the women who *win* are (at least by some measures) superior elected officials (Jeydel and Taylor 2003; Brollo and Troiano 2016). The complex factors that eventuate in a person's name appearing on the ballot are surely different by gender so that the types of men and women who end up as candidates differ from each other in both observed and unobserved ways. If so, then comparisons of male and female candidates (even controlling for observables) may yield biased estimates of the true effect of gender on vote choice.

In response to this inferential difficulty, scholars have turned to a particular flavor of randomized experiment, the candidate choice survey experiment. In these studies, survey respondents evaluate hypothetical candidates – presented to them either in vignettes or statements of their personal characteristics in a conjoint table – and report whether they would or would not vote for a given candidate. Because the gender[2] of the candidate (and typically, many of the other candidate characteristics) is randomized, a comparison of the support for female candidates versus male candidates yields an estimate of the effect of gender that is not biased by unobserved confounding factors that may plague many observational analyses.

While survey experiments provide design-based assurances that the causal effect estimates are unbiased, one may still be concerned that the question they answer is not quite the question we care about. We want to know the average effect of switching the genders of male and female candidates in the real world, but the question answered by the hypothetical candidate choice experiment is subtly different. First, we have some evidence that subjects evaluate hypothetical and real candidates differently (McDonald 2019); the true effect of gender might be different for hypothetical versus real candidates. Second, we might be concerned that subjects spend less cognitive effort when filling out a survey questionnaire than they do when filling out a ballot. Third, we might worry that the observed response patterns are due to *experimenter demand effects* (Zizzo 2010) as respondents may anticipate the research objectives of a study and adjust their answers accordingly to "confirm" the researcher's hypotheses. Respondents may also desire to appear like "good" people who be-

---

[2]In this article, we will speak exclusively of candidate gender being either "male" or "female." We recognize that gender can take on many values beyond these two. However, none of the experiments in our sample assign candidates a gender beyond male or female. We also recognize that gender is socially-constructed and a distinct concept from biological sex (Bittner and Goodyear-Grant 2017). Our goal is to understand the effects of the social construction, but we grant that in these experiments, gender and sex are perfectly collinear in the sense that when respondents are informed that a candidate is female, they likely infer both her gender and her sex.

have in ways that are socially desirable, thus masking their true preferences (e.g., Krupnikov et al. 2016). Such explanations are difficult to square, however, with recent investigations that have found very little evidence for demand effects, especially in the context of studies administered online (De Quidt et al. 2017). For example, response patterns did not differ systematically when researchers varied the amount of information that participants received about the study objectives at the beginning of an online survey (Mummolo and Peterson 2017). Another study randomly altered demographic information about the researcher in an online experiment and reported no difference in how respondents answered subsequent questions on racial resentment, gender roles, or support for female and minority presidential candidates (White et al. 2016).

We aim in this short article to document and summarize in one place what has been learned about how gender influences voters' preference for political candidates from hypothetical candidate choice experiments. At this point, many such experiments have been conducted (indeed, by our count, gender is the most common candidate attribute to be randomized in these experiments). We collected 42 conjoint and factorial candidate-choice experiments where the researchers randomized candidate gender and studied respondents' vote choice. What emerges from our analysis aligns with existing observational studies: the balance of evidence shows that on average, voters *prefer* female to male candidates, all else equal, by approximately 2 percentage points. We attempt to evaluate the possibility that the survey experiments are generating systematically upwardly biased estimates by conducting an observational study of low salience, nonpartisan judicial elections in Colorado. Under the assumption that voters know nothing about these candidates beyond the information provided on the ballot, we leverage the fact that voters can infer gender from names to estimate the causal effect of gender on support. The results of this observational study are substantively similar to the experimental results, leading us to conclude that the survey experimental estimates are unlikely to be an artifact of the method itself and that indeed, voters prefer men and women at similar rates, possibly even displaying a slight preference for women.

Supply and demand explanations for candidate selection

Women make up half of the world's population but hold only 23 percent of the elected positions in national legislatures (Inter-Parliamentary Union 2018). The existing theory and evidence tends to group possible causes of this persistent gender gap in electoral politics into

factors that shape the supply and demand for politicians who are female (e.g., Karpowitz et al. 2017; Holman and Schneider 2018).

Supply-side explanations consider each of the critical junctures that may shape whether and what sorts of women end up standing for election. Women might not aspire to run for office at the same rates as men do (Lawless and Fox 2010; Kanthak and Woon 2015). They also might face higher entry costs into politics, especially in primary-based electoral systems (Lawless and Pearson 2008), and they might be more averse to highly competitive settings than men (Preece and Stoddard 2015). As many women continue to be primary caregivers for their families, they might shy away from politics when their political career adversely affects their work-life balance, for example in the form of extended commutes to legislative offices (Silbermann 2015) or even increased risk of divorce (Folke and Rickne 2019). In addition, women are less likely to be recruited by party gatekeepers to run for office (Crowder-Meyer 2013). By contrast, when parties face high levels of electoral competition, they might be more inclined to consider women and minority candidates if it increases their chances of winning (Folke and Rickne 2016).

Others have flagged demand-side barriers. Voters simply might not have a "taste" for female candidates and politicians, or gender stereotypes about leadership abilities may disadvantage women at the ballot box (e.g., Alexander and Andersen 1993; Brescoll and Okimoto 2010; Sanbonmatsu 2002; Smith et al. 2007). One possibility sometimes invoked in the popular press is that women might lack political skill or be otherwise legislatively deficient, so voters correctly elect them at lower rates. The empirical record on this count, however, indicates that women are effective legislators (Jeydel and Taylor 2003) and are often more likely to get things done than their male counterparts (Brollo and Troiano 2016), though this pattern may result from a selection process in which only the highest quality women are elected (Anzia and Berry 2011; Fulton 2011).

Hypothetical candidate choice experiments shed the most light on the demand side of the candidate selection process, mainly because they measure voter preferences for candidates of various types (one might argue that they also inform the supply side to the extent that they measure the tastes of party gatekeepers). Our paper aims to summarize the evidence from candidate choice survey experiments on three distinct demand-side explanations. First, we explore whether individuals discriminate against female candidates on average. Next, we ask whether or not individuals discriminate against *specific types* of female candidates, with a focus on the intersection between race and ethnicity and gender. Lastly, we test whether certain *groups* of respondents display stronger (or weaker) preferences for female candidates.

We briefly review the theory underlying each of these three before turning to our review of candidate choice experiments.

Voter Preferences and Gender Discrimination

Hostility against women or a general distaste for female politicians would predict a preference for male candidates, all else equal. Perceptions about gender roles may shape evaluations of political candidates, especially in the context of male-dominated domains such as politics. Female candidates may face electoral penalties because they are perceived as defying traditional sex roles and prescriptive gender norms, even when they are as qualified for the job as their male counterparts (the *gender-incongruency hypothesis*, e.g., Brescoll and Okimoto 2010). Indeed, in one study, respondents rated fictitious male presidential candidates as more skilled and as having more political potential than female candidates – even when candidates had otherwise identical profiles (Smith et al. 2007). In a study of state legislative races in the US between 1970-1980, male candidates were also found to be preferred to women by an average of two percentage points (Welch et al. 1985). However, more recent studies have suggested that outright discrimination against female political candidates may not be as prevalent as it once was. An observational study of candidates for US Congress did not find evidence of differential candidate evaluations by gender after conditioning on partisanship (Dolan and Lynch 2014). Similarly, a study of local media coverage of political candidates in nearly 350 Congressional races found no significant differences in the portrayal of female and male office-seekers (Hayes and Lawless 2015).

Interactions with Other Candidate Characteristics

Even if individuals do not discriminate against women candidates in general, they might evaluate certain *types* of female or male candidates differently. In other words, female candidates might face double standards in terms of the qualifications or attributes they need to bring to the table if they want to succeed in politics (the *double-standard hypothesis*, see Teele et al. 2018). Experimental research has demonstrated penalties for women who overtly "seek power" or who are as assertive as men (Brescoll and Okimoto 2010). In addition, respondents react negatively to women who show emotions like anger (Brescoll and Uhlmann 2008; Brooks 2011). However, in a recent conjoint experiment, Teele et al. (2018) found that women faced bigger electoral *advantages* than men when they had a larger family, and for all other characteristics – age, marital status, experience in politics, and previous occupation –

men and women were not rewarded or penalized differentially. As we will show below, the balance of evidence from a large set of candidate choice experiments also does not support the double-standard hypothesis.

Intersectional theories of gender and politics predict that whatever effects candidate gender may have on candidate support, the effects are likely to be different for candidates of different racial or ethnic groups (e.g., Hooks 1982; Crenshaw 1991; Hardy-Fanta 2013; Mügge and De Jong 2013; Holman and Schneider 2018). Even if voters prefer white women to white men, they might not prefer black women to black men. Candidate choice experiments often randomize both characteristics of candidates independently, so are well-placed to evaluate this possibility.

Interactions with Respondent Characteristics

Identity-based theories of vote choice suggest that individuals favor political candidates who "look" and "think" like themselves (Converse et al. 1961; Besley and Coate 1997). In the context of gender, we might therefore expect female voters to prefer female over male political candidates (the *gender affinity hypothesis*, see Dolan 2008). Findings from a number of survey and experimental studies lend some support to this hypothesis (Dolan 1998; Plutzer and Zipp 1996; Sanbonmatsu 2002). However, more recent studies report no such gender affinity effect (Dolan 2008; Teele et al. 2018). Some have suggested that gender-based group consciousness provides the necessary link between gender identity and political preferences (e.g., Hildreth and Dran 1994). Such a gender-based group consciousness is sometimes equated with feminist consciousness (Gurin 1985; Klatch 2001) but can extend beyond feminism and instead interact with conservatism in important ways (Mendoza and DiMaria 2019; Schreiber 2002).

Candidates' personal characteristics may also provide informational shortcuts for voters, allowing them to infer candidates' policy positions and ideological orientations in low-information environments (Downs 1957; Popkin 1991; Kirkland and Coppock 2018). Similar to candidate partisan affiliation, candidate gender may provide such a shortcut (the *gender heuristic hypothesis*). For example, female candidates are often believed to be more liberal than male candidates, all else equal, which can advantage female Democratic candidates over their male counterparts amongst liberal voters (McDermott 1997, 1998). Conversely, Republican women running for office may face additional barriers (Bucchianeri 2017).

In addition, gender stereotypes may mold perceptions of issue positions that candidates hold and of skills and leadership abilities they bring to the table. Women are seen as

"more dedicated to honest government" (McDermott 1998) and viewed as better suited to handling issues related to women, children, the aged, and the poor (Huddy and Terkildsen 1993). By contrast, male politicians are often more trusted with issues related to security or the economy (Holman et al. 2016). Both Democratic and Republican voters appear to hold these gendered stereotypes, but because these groups differ in their policy preferences and ideological positions, they may endorse female candidates to varying degrees (Sanbonmatsu and Dolan 2009). Indeed, as we will discuss in more detail below, we do find a larger preference for female candidates amongst Democrats than amongst Republicans across the studies we include in our analysis.

Design

Most existing candidate choice experiments were not designed specifically to study gender, but nevertheless vary candidate gender as one of the many candidate characteristics included in the description of hypothetical candidates. Our goal is to leverage the randomization of candidate gender in many countries, time periods, and contexts in order to gain a holistic understanding of the effects of gender on vote choice. We attempted to collect all candidate choice experiments ever conducted that met the following inclusion criteria:

- Candidate gender is randomized.

- The dependent variable is, or can be transformed into, a binary vote choice for or against the candidate.

We followed standard practices to locate our studies: Citation chains, internet searches using the terms ("factorial", "candidate choice", "voter preference experiment", "conjoint experiment", "gender, vote, experiment", "vignette"), and word of mouth using social media as well as personal conversations with scholars in the field. In total, we located 42 experiments from 30 papers. In 26 of the 42 cases, we were able to obtain replication data either through private communication or publicly-available repositories. In the remaining 17 cases, we attempted to recover the necessary statistics from the text or graphics of the articles. We did not exclude studies based on the manner in which candidate gender is signaled to the survey respondent. Some studies manipulate gender by indicating "Male" or "Female" in a matrix of candidate characteristics (e.g., Kirkland and Coppock 2018) while others use pictures (e.g., Crowder-Meyer et al. 2015). We did not limit our data collection to any specific geographic context. While most studies were conducted in the US context, we also include samples

from Afghanistan, Argentina, Australia, Brazil, Chile, Japan, Malawi, Switzerland, and the UK. Moreover, studies were included regardless of their sample type. Some use convenience samples like Mechanical Turk (MTurk), Survey Sampling International (SSI), or student samples. Others use samples that are nominally representative of the voting-age population in a given country at the time the study was conducted. We included experiments that used a variety of designs, including standard factorial experiments in which only a few characteristics are varied, highly factorial conjoint experiments in which many characteristics are varied, and vignette experiments which embed manipulations in a larger dose of information about the candidate. Some experiments asked respondents to rate one candidate at a time, others asked respondents to choose between two at a time. We excluded several excellent studies that randomized gender but measured favorability or perceptions of competence instead of vote choice. Overall, our database of studies includes 42 experiments from four continents across three and a half decades. Table 1 provides an overview the studies used in our analysis.

We will present three types of estimates. The first is the estimate of the average treatment effect (ATE) of being a female versus male candidate in each study. These ATEs are typically *sample* average treatment effects (SATEs), though some studies target *population* average treatment effects (PATEs), either by using a probability sampling scheme or poststratification weights. The estimand in conjoint experiments is often referred to as the average marginal component effect (Hainmueller et al. 2014, AMCE,). Referring to a SATE or a PATE as an AMCE correctly emphasizes that the estimand itself depends on the distribution of the other candidate attributes included in the study. This jumble of acronyms aside, we will obtain a single ATE estimate from each study. These ATEs are study-specific: we do not assume that the true ATEs in each study are equal to one another.

Because we require candidate gender to be randomized, the difference-in-means will be an unbiased estimator of the ATE in each case. Where the raw data are available, we estimate robust standard errors and 95% confidence intervals using the `estimatr` package for R (Blair et al. 2018). We include sampling weights when provided by the original researchers in their replication datasets. When subjects rate multiple candidate profiles, we cluster our standard errors by respondent. Where raw data were not available, we searched the original publications for estimates of the ATE as well as uncertainty estimates. Occasionally, this process involved digitally measuring coefficient plots for both point estimates and 95% confidence intervals.

Our second analysis will present estimates of the Conditional Average Treatment Effect (CATE) of candidate gender depending on other (randomly assigned) candidate character-

Table 1: Study Manifest

	N subjects	N ratings	Raw Data	Sample Type
Aguilar et al. (2015), Sao Paulo	608	608	Yes	convenience
Bansak et al. (2018), USA - MTurk	2,411	144,494	Yes	convenience
Bansak et al. (2018), USA - SSI	643	38,482	Yes	convenience
Campbell et al. (2016), UK Frequency of MP Dissent	1,899	18,990	Yes	representative
Campbell et al. (2016), UK Type of MP Dissent	1,919	19,190	Yes	representative
Carnes and Lupu (2016), Argentina	1,149	2,298	Yes	representative
Carnes and Lupu (2016), UK	5,548	11,096	Yes	representative
Carnes and Lupu (2016), USA	1,000	2,000	Yes	representative
Clayton et al. (2019), Malawi	604	3,624	Yes	representative
Eggers et al. (2018), UK	1,367	2,806	Yes	representative
Hainmueller et al. (2014), USA	311	3,466	Yes	convenience
Henderson et al. (2019), USA, CCES	2,791	22,328	Yes	representative
Holman et al. (2016), USA	1,001	1,001	Yes	representative
Hopkins (2014), USA	551	7,714	Yes	representative
Kirkland and Coppock (2017), USA - MTurk	1,204	12,032	Yes	convenience
Kirkland and Coppock (2017), USA - YouGov	1,200	11,432	Yes	representative
Mo (2015), Florida	407	5,700	Yes	convenience
Saha and Weeks (2019), DLABSS 1	551	3,280	Yes	convenience
Saha and Weeks (2019), UK, Prolific	869	8,682	Yes	convenience
Saha and Weeks (2019), USA, DLABSS 2	497	4,886	Yes	convenience
Saha and Weeks (2019), USA, SSI	1,248	7,480	Yes	representative
Sen (2017), USA, SSI 1	798	4,797	Yes	convenience
Sen (2017), USA, SSI 2	763	4,592	Yes	convenience
Teele et al. (2018), USA	1,052	6,312	Yes	representative
Visconti (2018), Chile	210	3,360	Yes	convenience
Armendariz et al. (2018), USA	1,495	2,990	No	convenience
Atkeson and Hamel (2018), USA	1,500	3,000	No	convenience
Bermeo and Bhatia (2017), Afghanistan	2,485	7,455	No	representative
Crowder-Meyer et al. (2015), MTurk	430	1,290	No	convenience
Crowder-Meyer et al. (2015), UC Merced	350	1,050	No	student
Fox and Smith (1998), UCSB	650	2,600	No	student
Fox and Smith (1998), University of Wyoming	990	3,960	No	student
Horiuchi et al. (2017), Japan	2,200	22,000	No	convenience
Kage et al. (2018), Japan	1,611	9,666	No	representative
Kang et al. (2018), Australia	2,290	4,580	No	representative
Ono and Burden (2018), USA, SSI	1,583	15,830	No	representative
Piliavin (1987), USA	245	245	No	student
Sigelman and Sigelman, (1982), USA	227	227	No	student
Tomz and Van Houweling (2016), USA	4,200	25,200	No	representative
Vivyan and Wagner (2015), UK, YouGov - 2012	1,899	1,899	No	representative
Vivyan and Wagner (2015), UK, YouGov - 2013	1,919	1,919	No	representative
Wuest and Pontusson (2017), Switzerland	4,500	9,000	No	representative
Total	64,474	503,835	26	

istics. For example, to estimate the CATE given that candidates are black, we condition the dataset to only include black candidates, then estimate the difference-in-means using the same procedure described above. We estimate CATEs for all candidate dimensions for which we have data. These dimensions are overlapping, but we do not estimate CATEs in the *intersection* of candidate characteristics (e.g., among 55-year-old Democratic former police officers) because we run out of data too quickly.

Our third and final analysis estimates the CATE of candidate gender conditional on respondent characteristics. Because the space of possible respondent characteristics is very large, we limit ourselves to those respondent characteristics that have emerged as most salient in previous literature: respondent gender and (in the US) respondent partisanship.
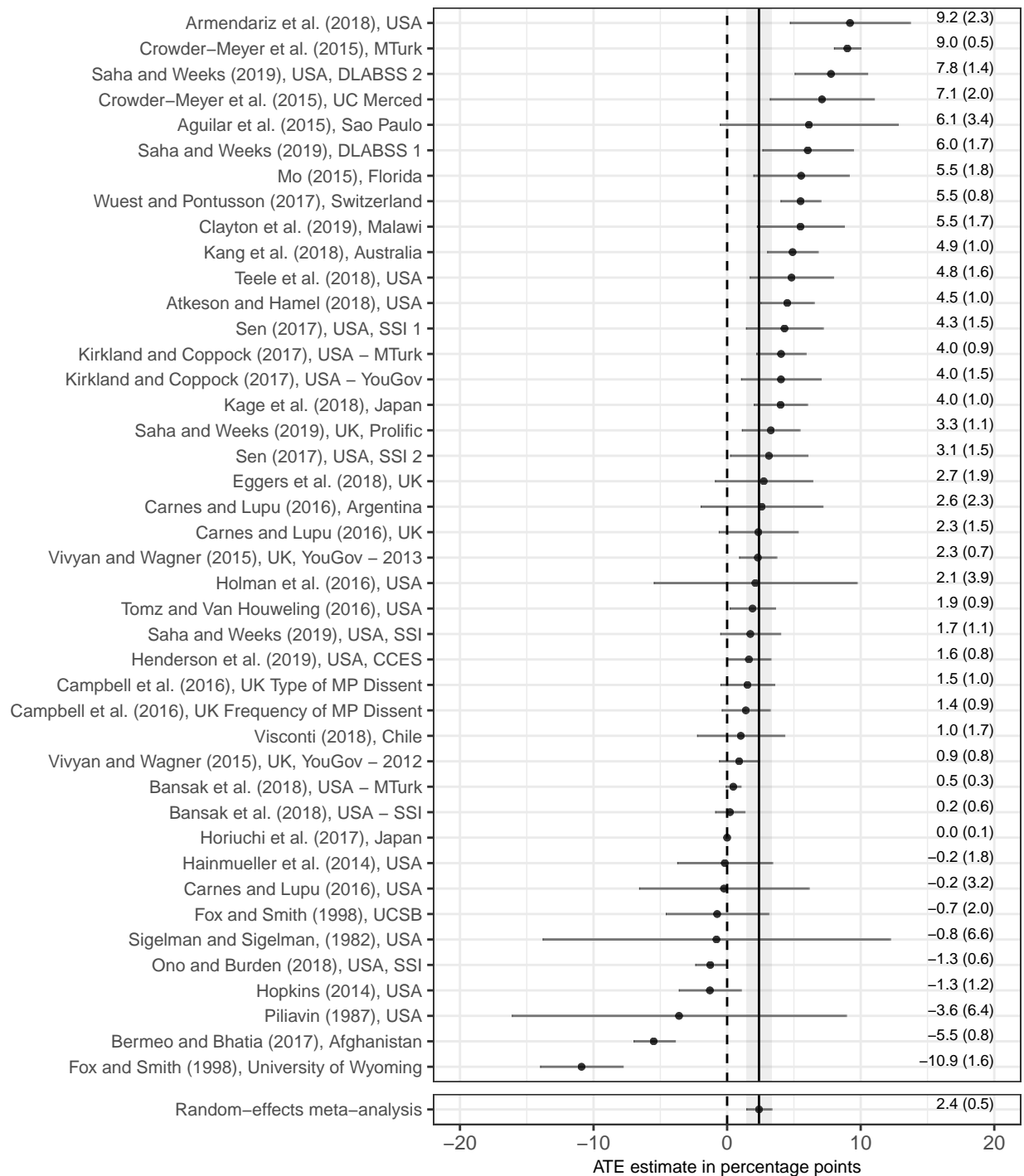
When averaging across studies, we will employ random-effects meta-analysis. Random-effects (rather than fixed-effects) is appropriate in this setting because we do not assume that the true average effect of gender is exactly the same across contexts. Instead, we assume that these effects vary from context-to-context, but are nevertheless drawn from a common distribution. The estimand in the random-effects meta analysis is the expectation (or average) of this distribution.

Results

Main Finding: Preference for Female Candidates

Our main result is presented in Figure 1. Using random-effects meta-analysis, the average SATE is 2.4 percentage points, with a confidence interval stretching from 1.5 to 3.4 points. We see a very consistent pattern in favor of women: 34 estimates are positive and 9 are negative. Only of these nine, only three are statistically significant: one conducted at the University of Wyoming in 1998 (Fox and Smith 1998), a second conducted in Afghanistan in 2017 (Bermeo and Bhatia 2017), and a third conducted on an SSI sample of Americans (Ono and Burden 2018)). Of the 42 experiments, 23 show statistically significant positive results and 16 produce estimates that cannot be distinguished from zero. While there are some exceptions and some very imprecise estimates, the results of these studies indicate an overall preference for female candidates over male candidates.

Figure 1: Results of 42 Candidate Choice Experiments on the Effect of Candidate Gender

Study	ATE (SE)
Armendariz et al. (2018), USA	9.2 (2.3)
Crowder–Meyer et al. (2015), MTurk	9.0 (0.5)
Saha and Weeks (2019), USA, DLABSS 2	7.8 (1.4)
Crowder–Meyer et al. (2015), UC Merced	7.1 (2.0)
Aguilar et al. (2015), Sao Paulo	6.1 (3.4)
Saha and Weeks (2019), DLABSS 1	6.0 (1.7)
Mo (2015), Florida	5.5 (1.8)
Wuest and Pontusson (2017), Switzerland	5.5 (0.8)
Clayton et al. (2019), Malawi	5.5 (1.7)
Kang et al. (2018), Australia	4.9 (1.0)
Teele et al. (2018), USA	4.8 (1.6)
Atkeson and Hamel (2018), USA	4.5 (1.0)
Sen (2017), USA, SSI 1	4.3 (1.5)
Kirkland and Coppock (2017), USA – MTurk	4.0 (0.9)
Kirkland and Coppock (2017), USA – YouGov	4.0 (1.5)
Kage et al. (2018), Japan	4.0 (1.0)
Saha and Weeks (2019), UK, Prolific	3.3 (1.1)
Sen (2017), USA, SSI 2	3.1 (1.5)
Eggers et al. (2018), UK	2.7 (1.9)
Carnes and Lupu (2016), Argentina	2.6 (2.3)
Carnes and Lupu (2016), UK	2.3 (1.5)
Vivyan and Wagner (2015), UK, YouGov – 2013	2.3 (0.7)
Holman et al. (2016), USA	2.1 (3.9)
Tomz and Van Houweling (2016), USA	1.9 (0.9)
Saha and Weeks (2019), USA, SSI	1.7 (1.1)
Henderson et al. (2019), USA, CCES	1.6 (0.8)
Campbell et al. (2016), UK Type of MP Dissent	1.5 (1.0)
Campbell et al. (2016), UK Frequency of MP Dissent	1.4 (0.9)
Visconti (2018), Chile	1.0 (1.7)
Vivyan and Wagner (2015), UK, YouGov – 2012	0.9 (0.8)
Bansak et al. (2018), USA – MTurk	0.5 (0.3)
Bansak et al. (2018), USA – SSI	0.2 (0.6)
Horiuchi et al. (2017), Japan	0.0 (0.1)
Hainmueller et al. (2014), USA	−0.2 (1.8)
Carnes and Lupu (2016), USA	−0.2 (3.2)
Fox and Smith (1998), UCSB	−0.7 (2.0)
Sigelman and Sigelman, (1982), USA	−0.8 (6.6)
Ono and Burden (2018), USA, SSI	−1.3 (0.6)
Hopkins (2014), USA	−1.3 (1.2)
Piliavin (1987), USA	−3.6 (6.4)
Bermeo and Bhatia (2017), Afghanistan	−5.5 (0.8)
Fox and Smith (1998), University of Wyoming	−10.9 (1.6)
Random–effects meta-analysis	2.4 (0.5)

ATE estimate in percentage points

12

Effects Conditional on Candidate Characteristics

We turn next to the effects of gender, conditional on other candidate characteristics. Since most studies included in our analysis were not primarily concerned with the effects of candidate gender on vote choice, they included a wide range of other candidate characteristics. Table 2 provides an overview of the attributes included in those studies for which we have raw data access.

Table 2: Overview of Attributes (Other Than Gender) Included in Study Designs

Study	Design Attributes
Aguilar et al. (2015)	Race, ballot length
Bansak et al. (2018)	Party ID, marital status, age, education, annual income, state of residence, main campaign contributor, policy positions on health care and abortion, religion, military service, favorite sport, car, favorite music
Campbell et al. (2016)	Ideology, years in politics, constituency effort, party dissent
Carnes and Lupu (2016)	Party ID, education, prior experience, race, profession
Clayton et al. (2018)	Education, leadership experience, family structure, occupation, policy priority, matrilineal vs. patrilineal kinship
Eggers et al. (2017)	Party ID, age, profession, corruption history
Hainmueller et al. (2013)	Age, race or ethnicity, education, income, profession, religion, military service
Henderson et al. (2019)	Race/ethnicity
Holman et al. (2016), USA	Party ID, terrorism threat
Hopkins (2016)	Race/ethnicity, party ID, religion, issue positions on health care, gun control, abortion, government spending and gay marriage, annual income
Kirkland and Coppock (2017)	Party ID, race, age, political experience, profession
Mo (2015)	Family details, ABA rating, education, previous experience, miscellaneous fact
Saha and Weeks (2019)	Age, political experience, political agenda, progressiveness, personality traits, profession
Sen (2017)	Age, profession, race/ethnicity, religion, education, court experience, party ID
Teele et al. (2018)	Age, marital status, number of children, profession, years in politics
Visconti (2017)	Ideology, age, profession, political experience, position on flood relief

Figure 2 shows the distribution of CATEs, conditional on each of the attributes listed in Table 2. For space reasons, we do not label each of the CATEs, though we provide figures of each separate estimate by study in the supplementary materials. Even without labels,

however, the figure tells a very consistent story. Not only do the *average* treatment effects of gender on vote choice tend to be positive, so too do the *conditional average* treatment effects of gender. Our main takeaway from Figure 2 is that the effect of gender does vary somewhat depending on the level of the other candidate characteristics but is typically positive.

In those studies that also randomize candidate race (Aguilar et al. 2015; Carnes and Lupu 2016; Hainmueller et al. 2014; Henderson et al. 2019; Hopkins 2014; Kirkland and Coppock 2018; Sen 2017), the preference subjects exhibit for women is slightly larger among white candidates than nonwhite candidates, as shown in Figure 3. Among white candidates, the preference for women is 2.9 percentage points. Among nonwhite candidates, the average preference is smaller, at 1 percentage point. The difference between these two estimates is not statistically significant ($p = 0.076$).

Effects Conditional on Respondent Characteristics

Lastly, we consider the effects of gender, conditional on the respondent characteristics, in particular by gender and partisans affiliations. Turning first to the *gender affinity hypothesis* discussed earlier, Figure 4(a) shows the CATEs of candidate gender, conditional on respondent gender. Both male and female respondents prefer female candidates to male candidates, by 4.6 points and 2.1 points, respectively. However, the effect seems to be larger among women, on average, as the difference between the two estimates is statistically significant ($p = 0.032$).

The pattern with respect to partisanship is somewhat clearer. In Figure 4(b), we summarize the results of 15 studies conducted in the US context, for which we were able access raw data and for which partisan identification of respondents was available. Republicans show a mild dispreference for women (-1.7 points, SE: 0.6 points) while Democrats and Independents display preference for female candidates on the order of 5 percentage points. The Republican estimate is statistically significantly different from the Democratic and Independent estimates. These estimates by party underline a general difficulty in interpreting the overall preference for women. The preference could represent a gender heuristic whereby people infer the sorts of policies women are likely to pursue when elected, or they reflect a taste-based preference for women. Yet, even when disaggregating by party, we cannot disentangle the two mechanisms because Democrats may both prefer the sorts of policies typically championed by women and also prefer women on taste-based grounds.

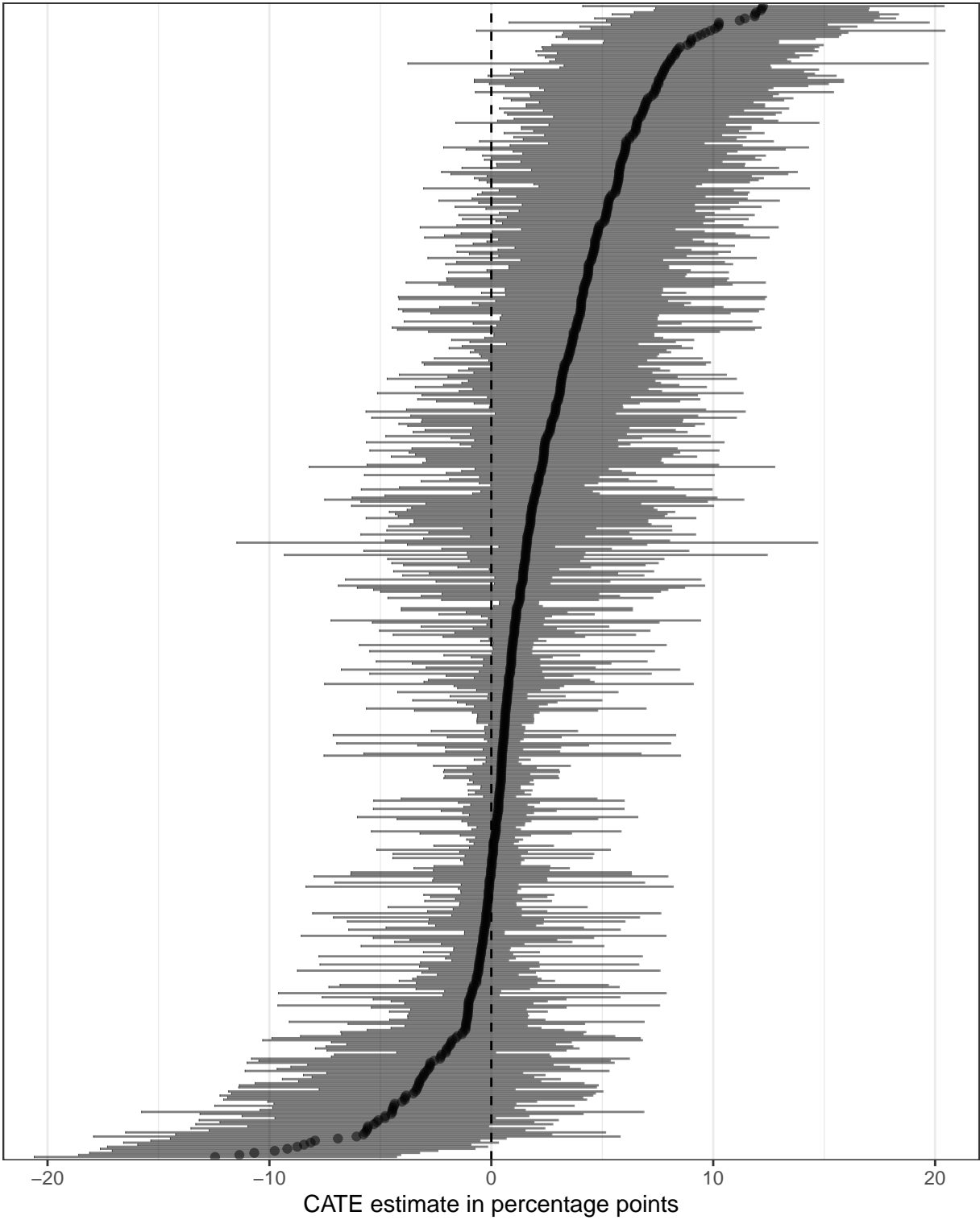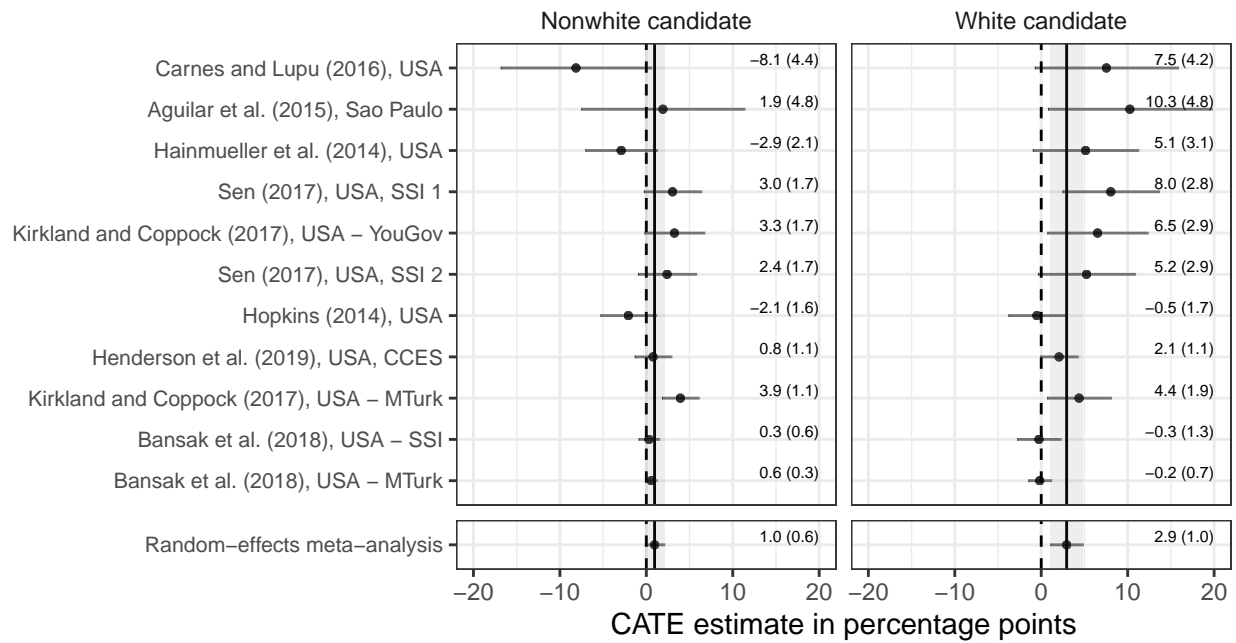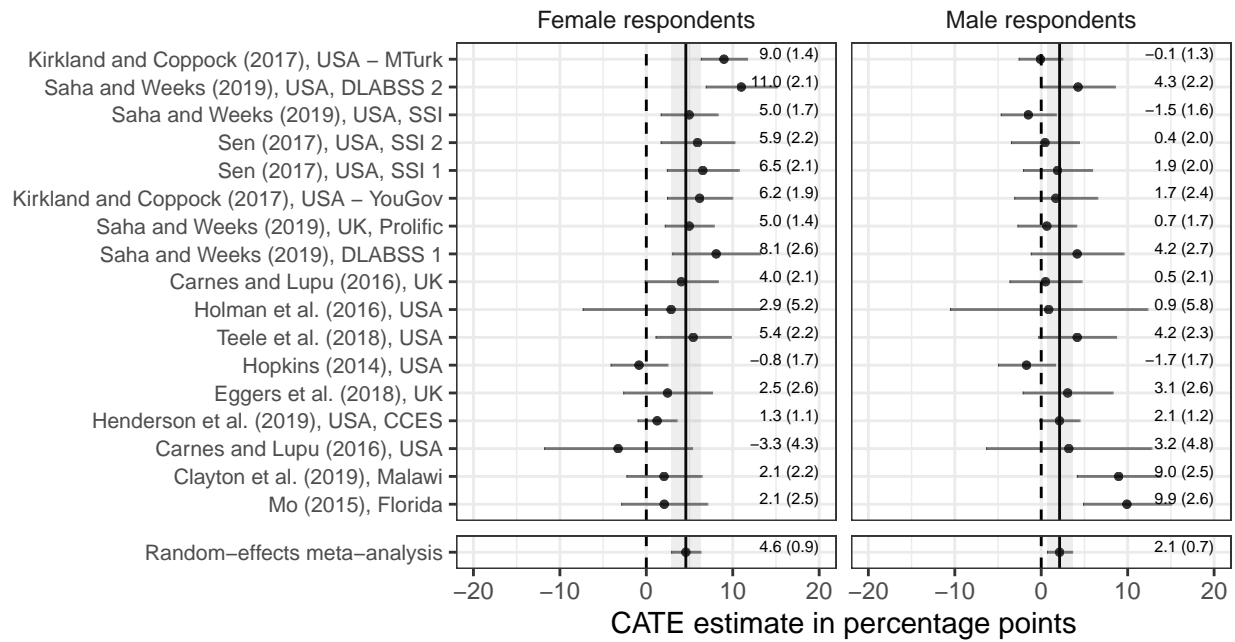Figure 2: 535 Estimates of the Effect of Candidate Gender, Conditional on Auxilliary Candidate Attributes

CATE estimate in percentage points

Figure 3: Conditional Average Effect of Candidate Gender, Conditional on Candidate Race

	Nonwhite candidate	White candidate
Carnes and Lupu (2016), USA	−8.1 (4.4)	7.5 (4.2)
Aguilar et al. (2015), Sao Paulo	1.9 (4.8)	10.3 (4.8)
Hainmueller et al. (2014), USA	−2.9 (2.1)	5.1 (3.1)
Sen (2017), USA, SSI 1	3.0 (1.7)	8.0 (2.8)
Kirkland and Coppock (2017), USA − YouGov	3.3 (1.7)	6.5 (2.9)
Sen (2017), USA, SSI 2	2.4 (1.7)	5.2 (2.9)
Hopkins (2014), USA	−2.1 (1.6)	−0.5 (1.7)
Henderson et al. (2019), USA, CCES	0.8 (1.1)	2.1 (1.1)
Kirkland and Coppock (2017), USA − MTurk	3.9 (1.1)	4.4 (1.9)
Bansak et al. (2018), USA − SSI	0.3 (0.6)	−0.3 (1.3)
Bansak et al. (2018), USA − MTurk	0.6 (0.3)	−0.2 (0.7)
Random−effects meta−analysis	1.0 (0.6)	2.9 (1.0)

CATE estimate in percentage points

(a) Respondent Gender

(b) Respondent Partisanship
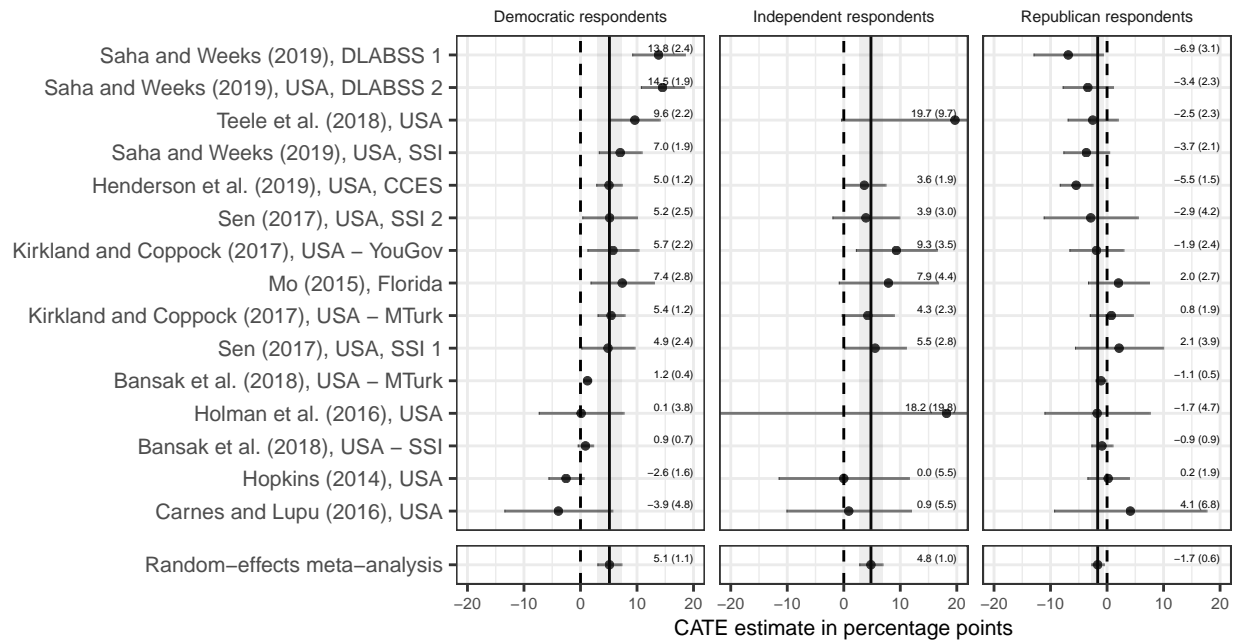
Figure 4: Conditional Average Treatment Effects of Candidate Gender, by Respondent Characteristics

External Validity Check: Colorado Judicial Elections 2016-2018

A persistent worry for hypothetical candidate choice experiments is that their artificial context renders their results inapplicable to politics more generally. The positive estimates could reflect true pro-female preferences, or they might simply be artifacts of the survey experimental design. One such artifact constitutes social desirability bias: respondents—contra their true preferences–may systematically choose female candidates because they believe that the survey enumerator will think highly of them for doing so. We agree with Clayton et al. (2019) that social desirability is not likely to be a consequential factor in conjoint experiments where the multiplicity of candidate attributes makes it difficult for the enumerator to determine which of the factors drove an individual's response (for a design that exploits this logic, see Horiuchi et al. (2018)). It is also possible that voters would prefer a hypothetical woman in principle, but would not prefer any actual real-world woman when they see her on the ballot. Finally, we might be worried that survey respondents do not pay serious attention in these experiments, but if they did, their biases against women would reveal themselves.

We were inspired by Hainmueller et al. (2015), who compared the results from conjoint experiments on attitudes towards immigrants with the results from actual referenda related to naturalization decisions in Switzerland and were able to replicate the findings from their survey experiments. We attempt a similar analysis, using low-salience elections in Colorado judicial retention elections. Such elections are a close real-world analogue to many of candidate choice experiments included here, since voters know next to nothing about the candidates beyond their names. Names carry many meanings, arguably the most salient of which is gender.

In the analysis that follows, we proceed under the assumptions that (1) the only information voters have about candidates is their name and (2) any additional information the name may carry is orthogonal to gender, i.e., even if voters infer race or class from names, these inferences are balanced across the male and female candidates. If these identifying assumptions hold, then we have a good real-world estimate of the effect of gender. If these assumptions do not hold, then our estimates are biased in the same ways as the observational studies discussed in the introduction.

Our data are precinct-level vote totals from Colorado judicial retention elections in 2016 and 2018. We subset to races with at least one male and one female candidate (as determined by our read of candidate names and confirmed by the pronouns associated with candidates on their websites). Our outcome variable is the vote share for each candidate, where vote share is calculated as the total number of votes cast for each candidate, divided by the total number

of voters, measured by the total votes cast in the highest-salience election (U.S. President in 2016 and U.S. Representative in 2018). We use this slightly nonstandard measure because in these retention elections, voters can vote for more than one candidate. We think assumption 1 is reasonable because of the low salience of the races for judgeships at the county, district, appeals, and juvenile court levels. Some empirical support for this logic comes from Canes-Wrone et al. (2018), who show that public opinion on low salience issues and judicial decisions are mostly uncorrelated. We think assumption 2 is plausible, though of course we cannot affirmatively demonstrate that it holds. Because race and ethnicity is also encoded in names, we assessed the putative race of all candidates through a visual inspection of their candidate profile photographs. Pooling across both years, 86.7% of the female candidates appeared to be white, compared with 87.3% of the male candidates, though more male candidates were white in 2016 and more female candidates were white in 2018. We reason on this basis that whatever racial or ethnic cues may be communicated by the names in addition to gender, they are similar across male and female candidates.

We estimate the effect of running as a woman separately for each race within the election before we pool the race-specific effects together using random-effects meta-analysis. This approach "controls for" any race-specific heterogeneities associated with having multiple male or multiple female candidates. While the gender effect is positive in some races and negative in others, we find that on average, the effects are very close to zero. In 2016, the meta-analytic estimate of running as a female candidate is -0.5 percentage points (SE: 0.6 points) and in 2018, the estimate is +1 percentage points (SE: 0.7 points). We present these estimates in Figure 5.

For us, the main takeaway from this analysis is that the average pro-female preference we observe in the experimental data is not wildly out of line with what we find in an observational analysis of data that are quite similar. These low-salience judicial candidates are *almost* like hypothetical candidates whose gender has been randomly assigned, so these meta-analytic estimates are *almost* credible causal effect estimates.

Discussion

We have summarized the statistical evidence from 42 candidate choice experiments on gender. Our main finding shows that survey subjects prefer female candidates by approximately 2 percentage points on average. We observe considerable study-to-study variation, though more than three-fourths of the studies show a net preference for women.
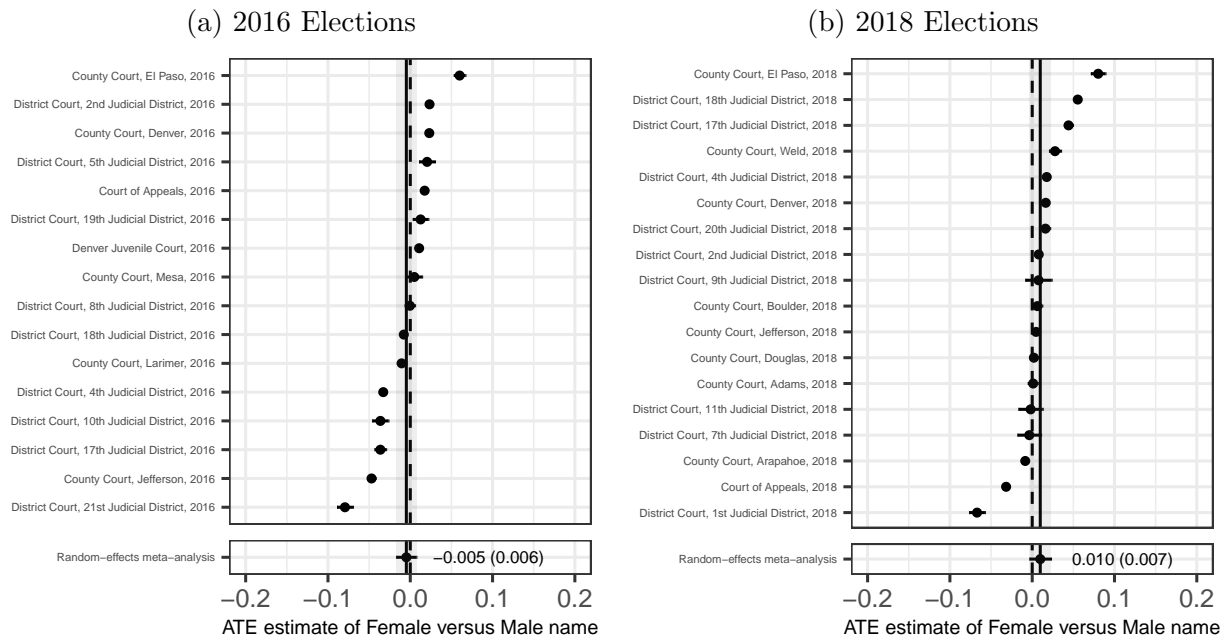
Figure 5: Observational Estimates of Average Treatment Effects of Female Candidate Names

We further investigated whether these average effects mask important heterogeneities. We find some evidence that the preference for women is stronger among white candidates than among non-white candidates. However, on the whole, our results do not depend on other candidate characteristics (such as experience, age, or occupation) and therefore (in line with Teele et al. (2018)) do not support the hypothesis that voters systematically apply double standards when they evaluate female candidates. We found some interactions of candidate gender with respondent characteristics, however. While both male and female respondents tend to prefer women, support for female candidates tends to be larger among women, lending some support to the gender affinity hypothesis. We also observed a stronger preference for women among Democrats and Independents compared with Republicans. It is unclear, however, whether this difference is due to a gender heuristic whereby partisans infer the sorts of policies women are likely to pursue when elected, or whether it arises from a taste-based preference among Democrats and Independents for female candidates in general.

Our meta-analytic results may come as a surprise to some, as they did to us when we began the project. Indeed, many of the authors of the experiments included in our meta-analysis expressed similar surprise at their own results as well. Clayton et al. (2019) write, "To our surprise, our experimental results did not reveal a generalized public distaste for women leaders." Kage et al. (2018) remark, "We were surprised to find, based on three

20

experimental surveys, that Japanese voters do not harbor particularly negative attitudes toward female politicians." Saha and Weeks (2019) conclude, "Contrary to the expectations of Hypothesis 1, we find no evidence that voters penalize ambitious women." Teele et al. (2018) summarize their main result as "In both surveys and among most subgroups we do not find evidence that women are discriminated against *as* women. In fact, in the baseline results from each survey, presented in Figure 1, female candidates actually get a *boost* over men" (emphasis in original). We hope that our meta-analysis convincingly demonstrates that those individual findings were not flukes but instead generalize quite well to many times and contexts.

Still, some may worry that these findings are simply an artifact of the survey methodology itself, masking experimenter demand effects or participants' low cognitive effort in a low-stakes environment. However, in the real world, low-profile, non-partisan elections provide a relatively close analogue to many of the survey-based candidate choice experiments included here, assuming that voters do not know anything about the candidates except the name on the ballot. We therefore used election return data from nonpartisan judicial retention elections in Colorado to show that–in line with the experimental findings–voters do not appear to discriminate against female candidates.

Overall, our findings offer evidence against demand-side explanations of the gender gap in politics. Rather than discriminating against women who run for office, voters appear to prefer female to male candidates on average. What then explains the persistent underprovision of women in politics across the globe? For us, the findings we discussed here suggest that supply side factors that include party structures, donor preferences, candidate recruitment, and differences in opportunity costs are correctly coming under deeper scrutiny by political scientists (e.g., Crowder-Meyer 2013; Lawless and Fox 2010; Silbermann 2015; Preece et al. 2016; Thomsen and Swers 2017). Of course, evaluating the causal influence of such supply side factors on female representation is inherently more difficult as candidate selection is a complex, often opaque process. Nevertheless, recent scholarship has demonstrated the potential for interesting field experimental work in this area. Foos and Gilardi (2019) shows that a randomized invitation to meet with female politicians did not increase self-reported political ambition among female university students in Switzerland. Karpowitz et al. (2017) randomly induced leaders of precinct level caucus meetings to read statements encouraging their membership to elect more women delegates to the statewide convention, increasing the fraction of precincts electing at least one woman by more than five percentage points. We hope that future work will continue to push forward this promising research agenda.

References

Aguilar, Rosario, Saul Cunow and Scott Desposato. 2015. "Choice Sets, Gender, and Candidate Choice in Brazil." *Electoral Studies* 39(Supplement C):230 – 242.

Alexander, Deborah and Kristi Andersen. 1993. "Gender as a Factor in the Attribution of Leadership Traits." *Political Research Quarterly* 46(3):527–545.

Anzia, Sarah and Christopher Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55(3):478–493.

Armendariz, Karen Caballero, Ben Farrer and Monica Martinez. 2018. "Badge of Courage or Sign of Criminality? Experimental Evidence For How Voters Respond to Candidates With A Political Arrest In Their Past.".

Atkeson, Lonna Rae and Brian T. Hamel. 2018. "Fit for the Job: Candidate Qualifications and Vote Choice in Low Information Elections." *Political Behavior; New York* pp. 1–24.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2018. "Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments." *Political Analysis* 26:112–119.

Bermeo, Nancy and Jasmine Bhatia. 2017. Strongmen or Technocrats? Experimental Evidence Testing Leadership Preferences in Afghanistan. Technical report International Growth Center.

Besley, Timothy and Stephen Coate. 1997. "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics* 112(1):85–114.

Bittner, Amanda and Elizabeth Goodyear-Grant. 2017. "Sex isn't Gender: Reforming Concepts and Measurements in the Study of Public Opinion." *Political Behavior* 39.

Blair, Graeme, Jasper Cooper, Alexander Coppock, Macartan Humphreys and Luke Sonnet. 2018. *estimatr: Fast Estimators for Design-Based Inference.*

Brescoll, Victoria L. and Eric Luis Uhlmann. 2008. "Can an Angry Woman Get Ahead?: Status Conferral, Gender, and Expression of Emotion in the Workplace." *Psychological Science* 19(3):268–275.

Brescoll, Victoria L. and Tyler G. Okimoto. 2010. "The Price of Power: Power Seeking and Backlash Against Female Politicians." *Personality and Social Psychology Bulletin* 36(7):923–936.

Brollo, Fernanda and Ugo Troiano. 2016. "What Happens When a Woman Wins an Election? Evidence from Close Races in Brazil." *Journal of Development Economics* 122:28–45.

Brooks, Deborah Jordan. 2011. "Testing the Double Standard for Candidate Emotionality: Voter Reactions to the Tears and Anger of Male and Female Politicians." *The Journal of Politics* 73:597–615.

Bucchianeri, Peter. 2017. "Is Running Enough? Reconsidering the Conventional Wisdom about Women Candidates." *Political Behavior* pp. 1–32.

Burleigh, Nina. 2016. "The presidential election was a referendum on gender and women lost.".
URL: *https://www.newsweek.com/2016/11/18/hillary-clinton-presidential-election-voter-gender-gap-520579.html*

Campbell, Rosie, Philip Cowley, Nick Vivyan and Markus Wagner. 2016. "Legislator Dissent as a Valence Signal." *British Journal of Political Science* pp. 1–24.

Canes-Wrone, Brandice, Tom S Clark and Amy Semet. 2018. "Judicial Elections, Public Opinion, and Decisions on Lower-Salience Issues." *Journal of Empirical Legal Studies* 15(4):672–707.

Carnes, Nicholas and Noam Lupu. 2016. "Do Voters Dislike Working-Class Candidates? Voter Biases and the Descriptive Underrepresentation of the Working Class." *American Political Science Review* 110(4):832–844.

Clayton, Amanda, Amanda Lea Robinson, Martha C Johnson and Ragnhild Muriaas. 2019. "(How) Do Voters Discriminate Against Women Candidates? Experimental and Qualitative Evidence From Malawi." *Comparative Political Studies* p. 0010414019858960.

Converse, Philip E., Angus Campbell, Warren E. Miller and Donald E. Stokes. 1961. "Stability and Change in 1960: A Reinstating Election." *American Political Science Review* 55(2):269–280.

Crenshaw, Kimberle. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review* 43(6):1241–1299.

Crockett, Emily. 2016. "Why misogyny won.".
URL: *https://www.vox.com/identities/2016/11/15/13571478/trump-president-sexual-assault-sexism-misogyny-won*

Crowder-Meyer, Melody. 2013. "Gendered Recruitment without Trying: How Local Party Recruiters Affect Women's Representation." *Politics & Gender* 9(4):390–413.

Crowder-Meyer, Melody, Shana Kushner Gadarian, Jessica Trounstine and Kau Vue. 2015. "Complex Interactions: Candidate Race, Sex, Electoral Institutions and Voter Choice.".

Darcy, R. and Sarah Slavin Schramm. 1977. "When Women Run Against Men." *Public Opinion Quarterly* 41(1):1–12.

De Quidt, Jonathan, Johannes Haushofer and Christopher Roth. 2017. Measuring and Bounding Experimenter Demand. Technical report National Bureau of Economic Research.

Dittmar, Kelly. 2017. "Candidates Matter: Gender Differences in Election 2016.".
URL: *https://www.cawp.rutgers.edu/sites/default/files/resources/closer_look_candidates_matter_2.14.17.pdf*

Dolan, Kathleen. 1998. "Voting for Women in the "Year of the Woman"." *American Journal of Political Science* 42(1):272–293.

Dolan, Kathleen. 2008. "Is There a "Gender Affinity Effect" in American Politics?" *Political Research Quarterly* 61(1):79–89.

Dolan, Kathleen and Timothy Lynch. 2014. "It Takes a Survey: Understanding Gender Stereotypes, Abstract Attitudes, and Voting for Women Candidates." *American Politics Research* 42(4):656–676.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper.

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

Folke, Olle and Johanna Rickne. 2016. "Electoral Competition and Gender Differences in Political Careers." *Quarterly Journal of Political Science* 11(1):59–102.

Folke, Olle and Johanna Rickne. 2019. "All the Single Ladies: Job Promotions and the Durability of Marriage." *American Economic Journal: Applied Economics* . Forthcoming.

Foos, Florian and Fabrizio Gilardi. 2019. "Does Exposure to Gender Role Models Increase Women's Political Ambition? A Field Experiment with Politicians." *Journal of Experimental Political Science* . Forthcoming.

Fox, Richard L. 2005. Congressional Elections: Where Are We on the Road to Gender Parity? In *Gender and Elections: Shaping the Future of American Politics*, ed. Susan J. Carroll and Richard L. Fox.

Fox, Richard L. and Eric R. A. N. Smith. 1998. "The Role of Candidate Sex in Voter Decision-Making." *Political Psychology* 19(2):405–419.

Fulton, Sarah A. 2011. "Running Backwards and in High Heels." *Political Research Quarterly* 65(2):303–314.

Gurin, Patricia. 1985. "Women's Gender Consciousness." *Public Opinion Quarterly* 49(2):143–163.

Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1):1–30.

Hainmueller, Jens, Dominik Hangartner and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments Against Real-world Behavior." *Proceedings of the National Academy of Sciences* 112(8):2395–2400.

Hardy-Fanta, Carol. 2013. *Intersectionality and Politics: Recent Research on Gender, Race, and Political Representation in the United States.* United Kingdom: Routledge.

Hayes, Danny and Jennifer L. Lawless. 2015. "A Non-Gendered Lens? Media, Voters, and Female Candidates in Contemporary Congressional Elections." *Perspectives on Politics* 13(1):95–118.

Henderson, John A., Logan Dancey, Stephen N. Goggin, Geoffrey Sheagley and Alexander G. Theodoridis. 2019. "Issues or Affinity? How Voters Decide in Primary Elections." *Unpublished manuscript* .

Hildreth, Anne and Ellen M. Dran. 1994. "Explaining Womens' Differences in Abortion Opinion:." *Women & Politics* 14(1):35–51.

Hills, Jill. 1981. "Candidates: The Impact of Gender." *Parliamentary Affairs* 34(2):221–228.

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

Holman, Mirya R and Monica C Schneider. 2018. "Gender, Race, and Political Ambition: How Intersectionality and Frames Influence Interest in Political Office." *Politics, Groups, and Identities* 6(2):264–280.

Hooks, Bell. 1982. "Ain't I A Woman?" *Das Argument; Berlin* 24(131):134.

Hopkins, Daniel J. 2014. "Do Voting Criteria Differ across Levels of the U.S. Federal System? A Conjoint Experiment to Understand Voter Decision-making.".

Horiuchi, Yusaku, Daniel M. Smith and Teppei Yamamoto. 2017. "Identifying Voter Preferences for Politicians' Personal Attributes: A Conjoint Experiment in Japan." *Unpublished Manuscript* .

Horiuchi, Yusaku, Zachary D Markovich and Teppei Yamamoto. 2018. "Can Conjoint Analysis Mitigate Social Desirability Bias?" *Unpublished Manuscript* .

Huddy, Leonie and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37(1):119–147.

Inter-Parliamentary Union. 2018. "Women in Parliaments: World and Regional Averages.".

Jeydel, Alana and Andrew J. Taylor. 2003. "Are Women Legislators Less Effective? Evidence from the U.S. House in the 103rd-105th Congress." *Political Research Quarterly* 56(1):19–27.

Kage, Rieko, Frances Rosenbluth and Seiki Tanaka. 2018. "What Explains Low Female Political Representation? Evidence from Survey Experiments in Japan." *Politics and Gender* Forthcoming.

Kang, Woo Chang, Jill Sheppard, Feodor Snagovsky and Nicholas Biddle. 2018. "Ethnic Identities and Descriptive Representation: A Conjoint Analysis of Vote Choice." *Unpublished manuscript* .

Kanthak, Kristin and Jonathan Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." *American Journal of Political Science* 59(3):595–612.

Karpowitz, Christopher F., J. Quin Monson and Jessica Robinson Preece. 2017. "How to Elect More Women: Gender and Candidate Success in a Field Experiment." *American Journal of Political Science* 61(4):927–943.

Kelley, Jonathan and Ian McAllister. 1984. "Ballot Paper Cues and the Vote in Australia and Britain: Alphabetic Voting, Sex, and Title." *The Public Opinion Quarterly; Princeton, N.J.* 48(2):452.

King, Amy and Andrew Leigh. 2010. "Bias at the Ballot Box? Testing Whether Candidates' Gender Affects Their Vote." *Social Science Quarterly* 91(2):324–343.

Kirkland, Patricia A. and Alexander Coppock. 2018. "Candidate Choice Without Party Labels:." *Political Behavior* . Forthcoming.

Klatch, Rebecca E. 2001. "The Formation of Feminist Consciousness among Left- and Right-Wing Activists of the 1960s." *Gender & Society* 15(6):791–815.

Krupnikov, Yanna, Spencer Piston and Nichole M. Bauer. 2016. "Saving Face: Identifying Voter Responses to Black Candidates and Female Candidates." *Political Psychology* 37(2):253–273.

Lawless, Jennifer L. and Kathryn Pearson. 2008. "The Primary Reason for Women's Underrepresentation? Reevaluating the Conventional Wisdom." *The Journal of Politics* 70(1):67–82.

Lawless, Jennifer L. and Richard L. Fox. 2010. *It Still Takes a Candidate: Why Women Don't Run for Office.* New York: Cambridge University Press.

McDermott, Monika L. 1997. "Voting Cues in Low-Information Elections: Candidate Gender as a Social Information Variable in Contemporary United States Elections." *American Journal of Political Science* 41(1):270–283.

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

McDonald, Jared. 2019. "Avoiding the Hypothetical: Why "Mirror Experiments" are an Essential Part of Survey Research." *International Journal of Public Opinion Research* . Forthcoming.

Mendoza, Saaid A. and Marissa G. DiMaria. 2019. "Not "With Her": How Gendered Political Slogans Affect Conservative Women's Perceptions of Female Leaders." *Sex Roles* 80(1-2):1–10.

Mo, Cecilia Hyunjung. 2015. "The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters." *Political Behavior* 37(2):357–395.

Mügge, Liza and Sara De Jong. 2013. "Intersectionalizing European Politics: Bridging Gender and Ethnicity." *Politics, Groups, and Identities* 1(3):380–389.

Mummolo, Jonathan and Erik Peterson. 2017. "Demand Effects in Survey Experiments: An Empirical Assessment.".

Ono, Yoshikuni and Barry C. Burden. 2018. "The Contingent Effects of Candidate Sex on Voter Choice." *Political Behavior; New York* pp. 1–25. Forthcoming.

Piliavin, Jane Allyn. 1987. "Age, Race, and Sex Similarity to Candidates and Voting Preference1." *Journal of Applied Social Psychology* 17(4):351–368.

Plutzer, Eric and John F. Zipp. 1996. "Identity Politics, Partisanship, and Voting for Women Candidates." *The Public Opinion Quarterly* 60(1):30–57.

Popkin, Samuel L. 1991. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns.* Chicago, IL: University of Chicago Press.

Preece, Jessica and Olga Stoddard. 2015. "Why Women Don't Run: Experimental Evidence on Gender Differences in Political Competition Aversion." *Journal of Economic Behavior & Organization* 117:296 – 308.

Preece, Jessica Robinson, Olga Bogach Stoddard and Rachel Fisher. 2016. "Run, Jane, Run! Gendered Responses to Political Party Recruitment." *Political Behavior* 38(3):561–577.

Saha, Sparsha and Ana Catalano Weeks. 2019. "Ambitious Women: Gender and Voter Perceptions of Candidate Ambition." *Unpublished Manuscript* .

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

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

Schreiber, Ronnee. 2002. "Injecting a Woman's Voice: Conservative Women's Organizations, Gender Consciousness, and the Expression of Women's Policy Preferences." *Sex Roles* 47(7-8):331–342.

Sen, Maya. 2017. "How Political Signals Affect Public Support for Judicial Nominations: Evidence from a Conjoint Experiment." *Political Research Quarterly* 70:374–393.

Sigelman, Lee and Carol Sigelman. 1982. "Sexism, Racism, and Ageism in Voting Behavior: An Experimental Analysis." *Social Psychology Quarterly* 45:263.

Silbermann, Rachel. 2015. "Gender Roles, Work-Life Balance, and Running for Office." *Quarterly Journal of Political Science* 10(2):123–153.

Smith, Jessi L., David Paul and Rachel Paul. 2007. "No Place for a Woman: Evidence for Gender Bias in Evaluations of Presidential Candidates." *Basic and Applied Social Psychology* 29(3):225–233.

Teele, Dawn Langan, Joshua Kalla and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112(3):525–541.

Thomsen, Danielle M. 2018. "Ideology and Gender in U.S. House Elections." *Political Behavior* . Forthcoming.

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

Tomz, Michael and Robert P. Van Houweling. 2016. "Political Repositioning: A Conjoint Analysis.".

Visconti Giancarlo. 2018. "After the Flood: Natural Disasters and Electoral Choices in Chile.".

Vivyan, Nick and Markus Wagner. 2015. "What Do Voters Want From Their Local MP?" *The Political Quarterly* 86(1):33–40.

Welch, Susan, Janet Clark and R. Darcy. 1987. *Women, Elections, and Representation*. New York, NY: Longman.

Welch, Susan, Margery M. Ambrosius, Janet Clark and Robert Darcy. 1985. "The Effect of Candidate Gender on Electoral Outcomes in State Legislative Races." *Western Political Quarterly* 38(3):464–475.

White, Ariel, Anton Strezhnev, Christopher Lucas, Dominika Kruszewska and Connor Huff. 2016. "Investigator Characteristics and Respondent Behavior in Online Surveys.".

Wuest, Reto and Jonas Pontusson. 2017. "Do Citizens Prefer Affluent Representatives? Evidence from a Survey Experiment in Switzerland.".

Zipp, John F. and Eric Plutzer. 1985. "Gender Differences in Voting for Female Candidates: Evidence from the 1982 Election." *Public Opinion Quarterly* 49(2):179–197.

Zizzo, Daniel John. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics* 13(1):75–98.