What Have We Learned About Gender From Candidate Choice Experiments?
A Meta-analysis of 30 Factorial Survey Experiments

Susanne Schwarz[*], William Hunt[†], and Alexander Coppock[‡]

March 30, 2018
Prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, April 5 - 8, 2018.

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. These experiments have been used to study many candidate attributes, such as policy position, race, age, political or career experience, attractiveness, and political party. By our count, the most common attribute studied in these experiments is candidate gender. We collect 30 such experiments and reanalyze them using a standardized statistical approach. Holding other candidate features fixed by design, female candidates are on average preferred by respondents by approximately 2 percentage points. We further investigate how this preference varies with respondent gender and partisanship and other candidate characteristics. We find limited evidence of heteogeneity, though the female preference appears to be somewhat smaller for black (versus white) candidates and among Republican (versus Democratic) respondents. We conclude by attempting to reconcile the disjunction between the unambiguous survey experimental results and the drastic underprovision of female elected representatives.

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

Candidate choice survey experiments have become a fertile testing ground for learning about the separate impacts of candidate characteristics on vote choice. Real-world candidates for elected office appear on the ballot as a bundle of intercorrelated characteristics and are the result of an endogenous process of candidate selection. By contrast, the characteristics of hypothetical candidates in factorial, vignette, or conjoint candidate choice experiments are randomized, to the better understand how each factor shapes voter preferences.

Hypothetical candidate choice experiments have been used to study the effects of candidates' policy positions, race, age, political or career experience, attractiveness, and party affiliation on vote choice. No attribute of hypothetical candidates has been more frequently randomized than gender.[1] We aim in this article to document and summarize in one place what has been learned about how gender influences candidate preference from these experiments. To that end, we make use of the increasing number of conjoint or factorial candidate choice experiments, which randomly vary candidate gender.

What emerges from our analysis is puzzling and seemingly at odds with our understanding of the gender dynamics of recruitment, campaigns, and voter preference. Whereas in the context of real elections, the balance of evidence points to electoral penalties for female candidates, the evidence from candidate choice experiments points in the opposite direction.

Theory

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 the persistent gender gap in electoral politics into factors that shape the supply and demand for politicians who are female (Karpowitz et al. 2017).

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 (Burns et al. 2001; Lawless and Fox 2010). 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

[1] 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). We argue that in these experiments the two are perfectly collinear in the sense that when a respondent is informed that a candidate is female, they infer both her gender and her sex.

than men (Preece and Stoddard 2015). In addition, they are less likely to be recruited by party gatekeepers to run for office (Crowder-Meyer 2013).

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 be 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).

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 in general. Next, we ask whether or not individuals discriminate against *specific types* of female candidates. Lastly, we test whether certain *groups* of respondents display a stronger (or weaker) preference for female candidates.

To preview our findings, we find an overall preference for female candidates of two points, and this preference does not appear to depend greatly on the other characteristics of the candidate, with the possible exception of candidate race. We also find that this gender preference holds for many demographic subsets, the major exception being Republicans in the US, who show a clear preference for men over women.

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).

Recent studies have suggested that outright discrimination against female political candidates may not be as prevalent as it once was. A study of candidates for US Congress did not find evidence of differential candidate evaluations by gender after conditioning on partisanship (Dolan and Lynch 2014). 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). As we summarize in greater detail below, candidate choice experiments also fail to find a preference of male over female candidates (Taylor et al. 2008; Teele et al. 2017; Crowder-Meyer et al. 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 differentially. 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. 2017). Experimental research has demonstrated penalties for women who overtly "seek power" or who are as assertive as men (Brescoll and Okimoto 2010). Respondents react negatively to women who show emotions like anger (Brescoll and Uhlmann 2008; Brooks 2011). In a conjoint experiment, Teele et al. (2017) found that women faced bigger electoral penalties than men only when they had a larger family. For all other characteristics – age, marital status, experience in politics, and previous occupation – men and women were not rewarded or penalized differentially.

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. 2017). 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 2018; 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 party ID, 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).

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, times, 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 a binary vote choice for or against the candidate

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). Studies were included regardless of the sample on which they were conducted. Some use convenience samples like Mechanical Turk (MTurk), Survey Sampling International (SSI), or student samples. Others use samples that are nominally representative of the US population at the time the survey was conducted. Overall, our database of studies includes 30 experiments from four continents across three and a half decades.

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. Several excellent studies that randomized gender but measured favorability, likelihood of support, or perceptions of competence instead of vote choice were excluded.

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. In total, we located 30 experiments from 21 papers. In 17 of the 30 cases, we were able to obtain replication data either through private communication or publicly-available repositories. In the remaining 13 cases, we attempted to recover the necessary statistics from the text of the articles.

We will present three types of estimates. The first is the estimate of the Sample Average Treatment Effect (SATE) of being a female versus male candidate. This estimand ignores all other candidate characteristics, but because we require gender to be independently randomized, the difference-in-means will be an unbiased estimator of the SATE. 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 applicable. When subjects rate multiple candidate profiles, we cluster our standard errors by respondent. Where raw data were not available, we searched the original write ups for estimates of the SATE as well as uncertainty estimates. Frequently, this process involved digitally measuring the coefficient plots for both point estimates and 95% confidence intervals. While this process is relatively precise, we acknowledge that our dataset likely includes non-statistical sources of measurement error.

Our second analysis will present estimates of the Conditional Average Treatment Effect

Table 1: Study Manifest

	N subjects	N ratings	Raw Data
Aguilar, Cunow, and Desposato (2015), Brazil	3,908	27,076	Yes
Aguilar, Cunow, and Desposato (2015), Sao Paulo	608	608	Yes
Bansak, Hainmueller, Hopkins, and Yamamoto (2018), USA - MTurk	2,411	144,494	Yes
Bansak, Hainmueller, Hopkins, and Yamamoto (2018), USA - SSI	643	38,482	Yes
Campbell, Cowley, Vivyan, and Wagner (2016), UK - Frequency of MP Dissent	1,899	18,990	Yes
Campbell, Cowley, Vivyan, and Wagner (2016), UK - Type of MP Dissent	1,919	19,190	Yes
Carnes and Lupu (2016), Argentina	1,000	2,000	Yes
Carnes and Lupu (2016), UK	1,149	2,298	Yes
Carnes and Lupu (2016), USA	5,548	11,096	Yes
Eggers, Vivyan, and Wagner (2017), UK	1,367	2,806	Yes
Hainmueller, Hopkins, and Yamamoto (2013), USA	311	3,466	Yes
Holman, Merolla, and Zechmeister (2016), USA	1,001	1,001	Yes
Kirkland and Coppock (2017), USA - MTurk	1,204	12,032	Yes
Kirkland and Coppock (2017), USA - YouGov	1,200	11,432	Yes
Mo (2015), Florida	407	5,700	Yes
Teele, Kalla, and Rosenbluth (2015), USA	1,052	6,312	Yes
Visconti (2017), Chile	210	3,360	Yes
Bermeo and Bhatia (2017), Afghanistan	2,485	7,455	No
Crowder-Meyer, Gadarian, Trounstine, and Vue (2015), MTurk	430	1,290	No
Crowder-Meyer, Gadarian, Trounstine, and Vue (2015), UC Merced	350	1,050	No
Fox and Smith (1998), UCSB	650	2,600	No
Fox and Smith (1998), University of Wyoming	990	3,960	No
Horiuchi, Smith, and Yamamoto (2016), USA	2,200	22,000	No
Kage, Rosenbluth, and Tanaka (2017), Japan	1,611	9,666	No
Nick Vivyan, Markus Wagner (2015), UK, YouGov - 2012	1,899	1,899	No
Nick Vivyan, Markus Wagner (2015), UK, YouGov - 2013	1,919	1,919	No
Piliavin (1987), USA	245	245	No
Sigelman and Sigelman, (1982), USA	227	227	No
Tomz and Van Houweling (2016), USA	4,200	25,200	No
Wuest and Pontusson (2017), Switzerland	4,500	9,000	No
Total	47,543	396,854	17

(CATE) of candidate gender depending on other candidate characteristics. For example, to estimate the CATE given that candidates are 55 years old, we condition the dataset to only include 55 year old 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 final analysis estimates the CATE of candidate gender, conditional on respondent characteristics. The covariate information provided in the replication datasets varies from study to study. We limit our investigation to respondent gender, race, education, partisanship, or ideology.

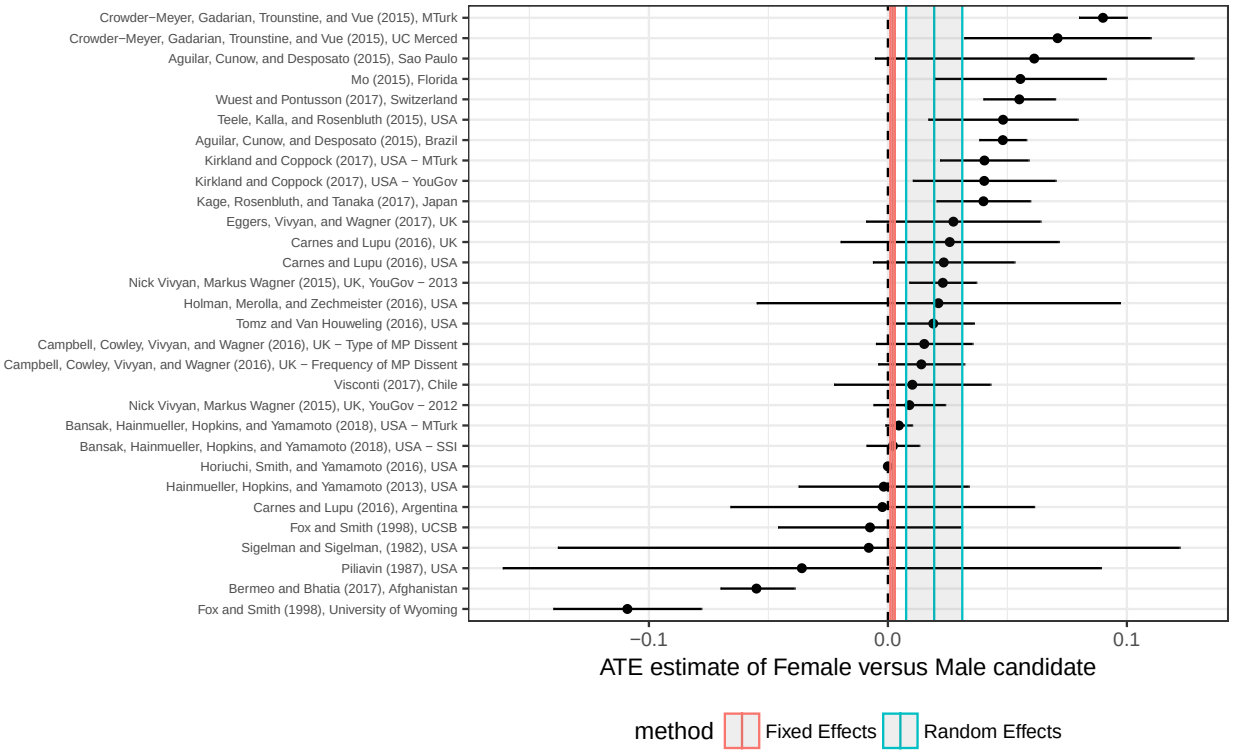
Results

Our main result is presented in Figure 1. Using random-effects meta-analysis, the average SATE is 1.9 percentage points, with a confidence interval stretching form 0.8 to 3.1 points. Fixed-effects meta-analysis puts the average much lower, at 0.2 points, with a tight confidence interval from 0.1 to 0.3 points. Regardless of which averaging technique we employ, we see a very consistent pattern in favor of women: 22 estimates are positive and 8 are negative. Only two experiments show statistically significant negative effects, one conducted at the University of Wyoming in 1998 (Fox and Smith 1998) and the other conducted in Afghanistan in 2017 (Bermeo and Bhatia 2017). Of the 30 experiments, 11 show statistically significant positive results and 17 produce estimates that cannot be distinguished from zero. While there are some exceptions and some underpowered studies, the results of these 30 studies indicate an overall preference for female candidates over male candidates.

Effects Conditional on Candidate Characteristics

We turn next to the effects of gender, conditional on other candidate characteristics. Figures 2 and 3 show our study-by-study results. The overwhelming pattern is one of treatment effect *homogeneity*, that is, the effect of candidate gender does not appear to depend much on the levels of the other candidate characteristics. The one exception that stands out is candidate race. In the four studies that also randomize candidate race (Aguilar et al. 2015; Carnes 2016; Hainmueller et al. 2014; Kirkland and Coppock 2018), the preference subjects exhibit for women is larger among white candidates than black candidates. This finding is consistent

Figure 1: Results of 30 Candidate Choice Experiments on the Effect of Candidate Gender



with intersectional theories of gender and race, which predict that black, female candidates are doubly-bound by racial and gender discrimination (Hooks 1982; Crenshaw 1991). On the whole, however, candidate choice experiments do not appear to give strong support to the hypothesis that gendered preferences for candidates depend on other candidate attributes.

We include the results of Mo (2015) separately in Figure 4 for two reasons. The first (and mundane) reason is that the CATE estimates are so much larger in magnitude that they require a different graphical scale. The second more important reason is that the candidate characteristic that is of most interest is whether or not, in a particular contest, the male or the female candidate is of obviously higher quality. The subjects in this study exhibit a clear dispreference for low quality candidates. When the candidates are both strong candidates (or have quality drawn from the population distribution), subjects exhibit the usual preference for women.

Effects Conditional on Respondent Characteristics

Lastly, we consider the effects of gender, conditional on the respondents' gender and political affiliations. Turning first to respondents' gender, we find nearly every pattern in individual studies, but no consistent trend across datasets. In some cases (e.g., Kirkland and Coppock 2018) we find that, as would be predicted by the gender affinity hypothesis, women exhibit a stronger preference for women than do men. But we also find the opposite pattern (Mo 2015), or no difference at all (Aguilar et al. 2015; Carnes 2016; Eggers et al. 2018; Holman et al. 2016; Teele et al. 2017).

The pattern with respect to partisanship is somewhat clearer. Democrats appear to exhibit a greater preference for female candidates than do Republicans in a number of studies (Kirkland and Coppock 2018; Mo 2015; Teele et al. 2017). In two cases (Carnes 2016; Holman et al. 2016), the partisan differences are too imprecisely estimated to make confident claims. Overall, it is unclear, however, whether partisan differences are attributable to a gender heuristic whereby partisans infer the sorts of policies women are likely to pursue when elected, or whether they reflect a taste-based preference among Republicans for male candidates in general.

Figure 2: Average Effect of Female Versus Male, Conditional on Candidate Characteristics

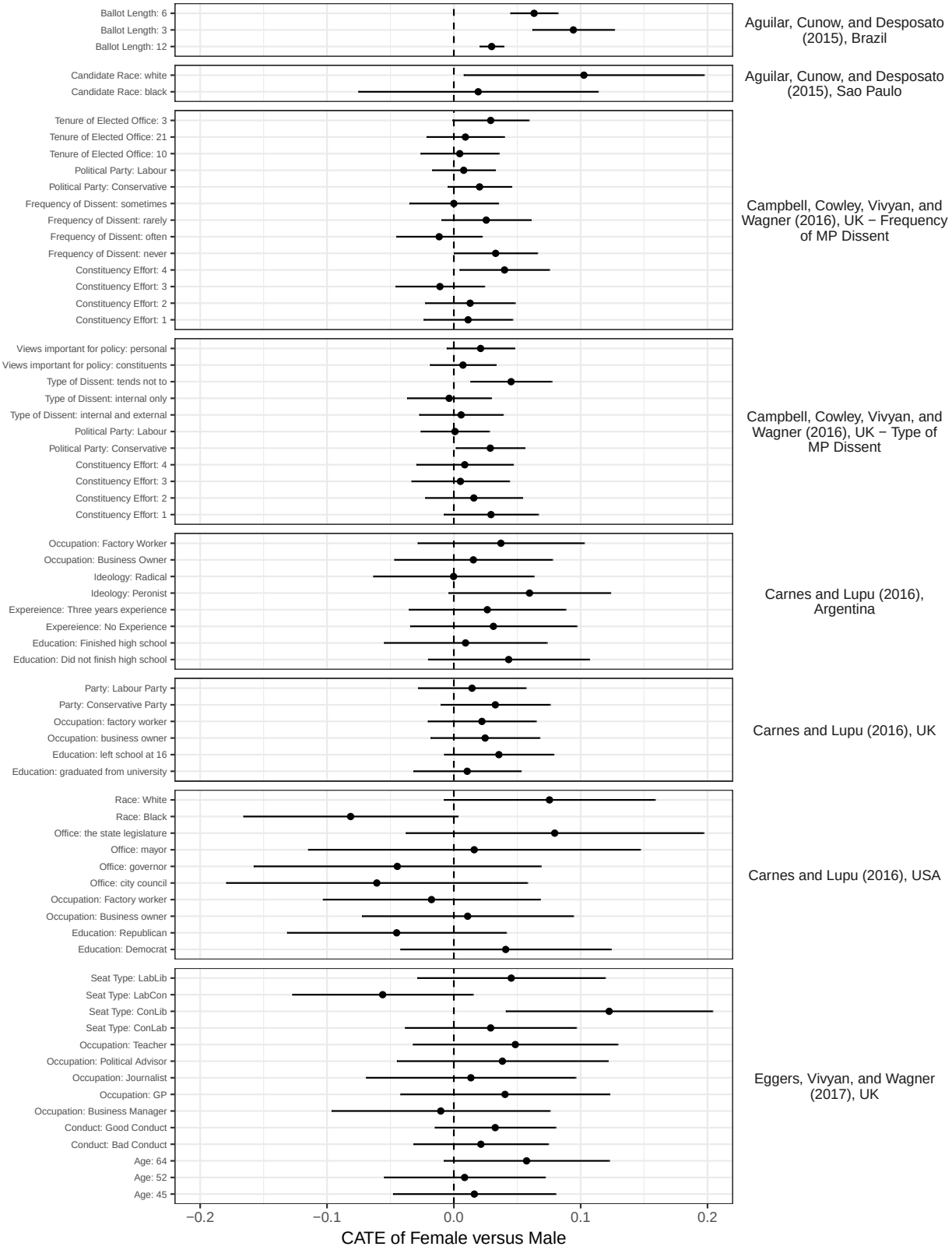


Figure 3: Average Effect of Female Versus Male, Conditional on Candidate Characteristics

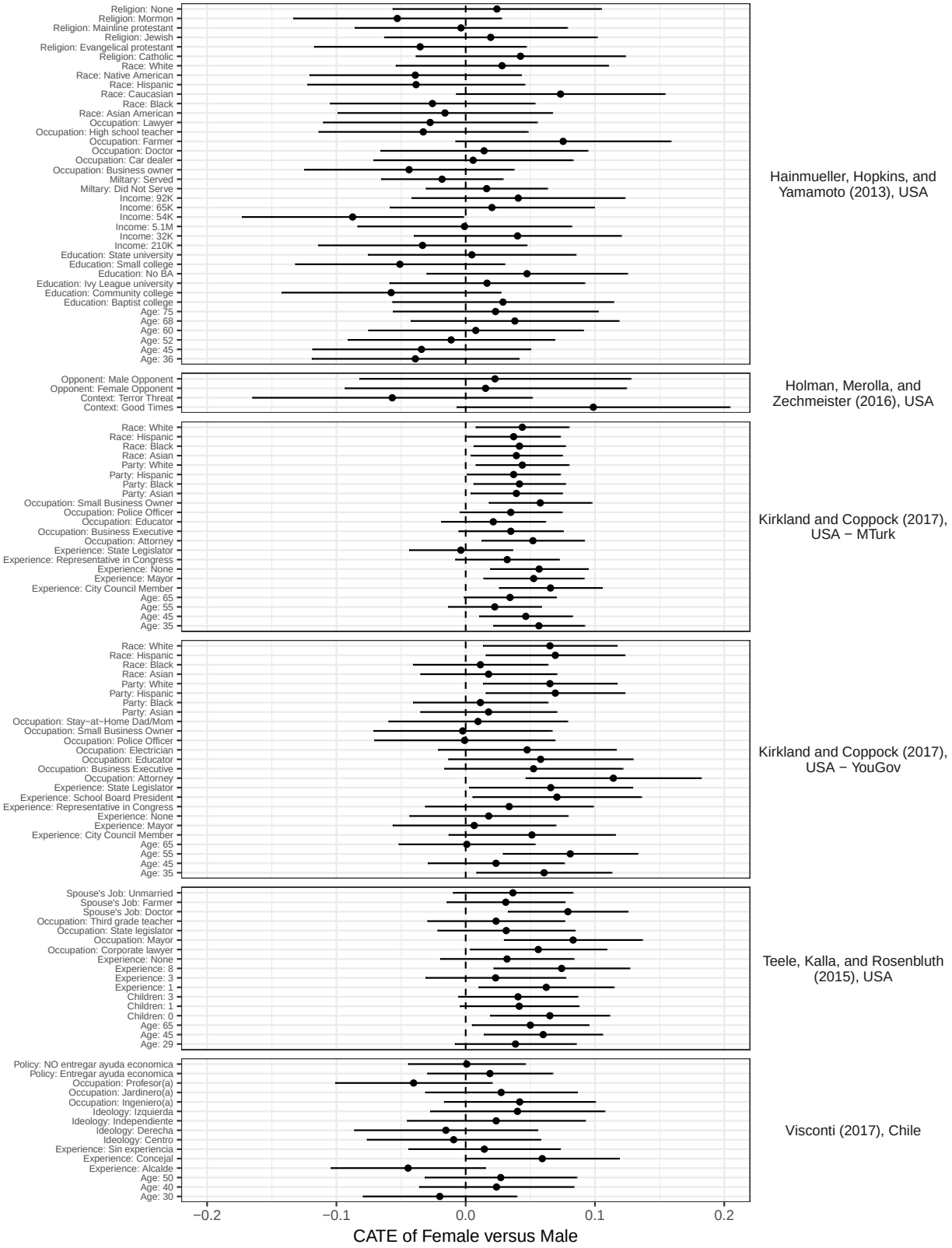


Figure 4: Average Effect of Female Versus Male, Conditional on Candidate Characteristics

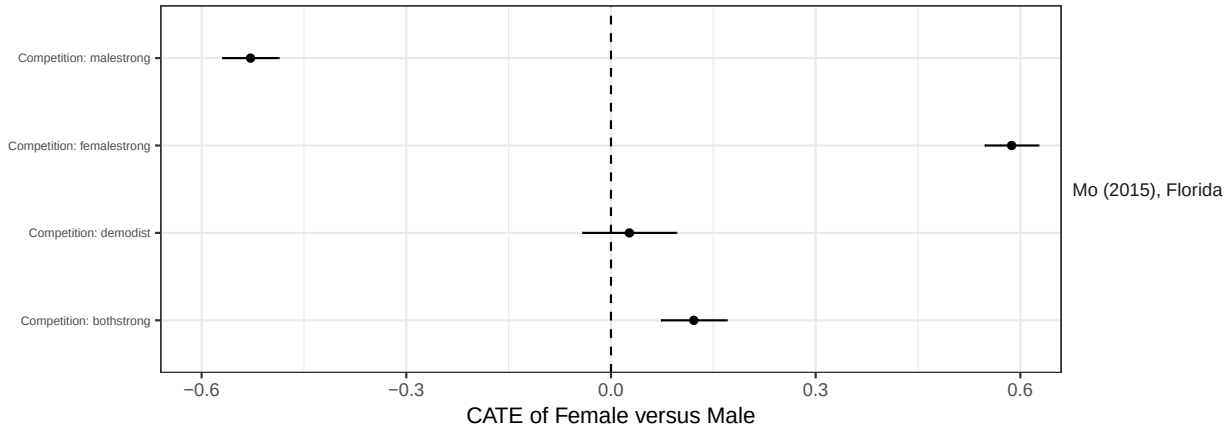
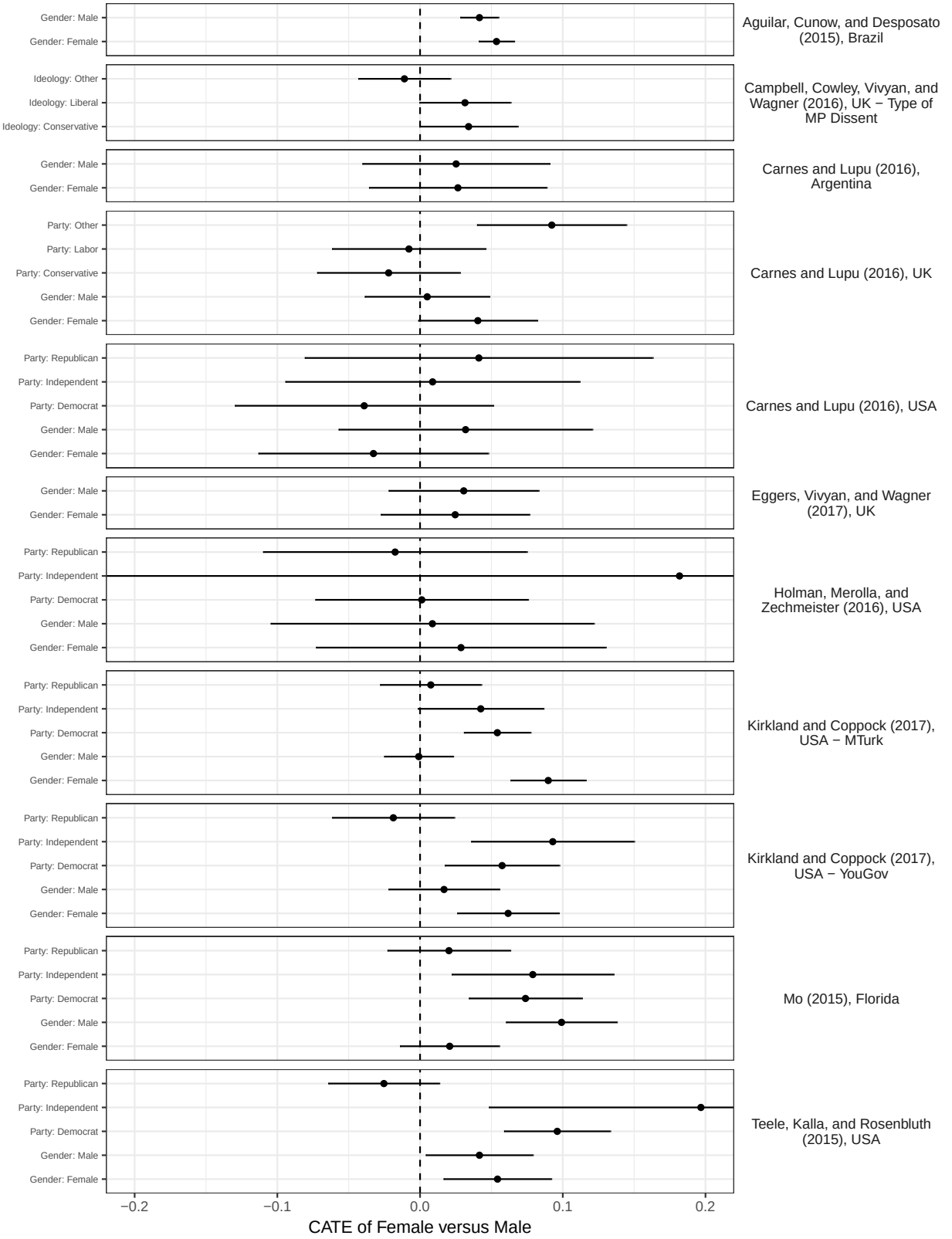


Figure 5: Average Effect of Female Versus Male, Conditional on Respondent Characteristics



Discussion

We have summarized the statistical evidence from 30 candidate choice experiments on gender. Our main finding suggests that survey subjects appear to prefer female candidates by approximately 2 percentage points on average. We observe considerable study-to-study variation, though more than two-thirds of the studies show a net preference for women. To the extent that we believe that candidate choice experiments are good proxies for how voters would behave at the ballot box, these results strongly confirm the growing consensus that the under-provision of women in elected office is due more to supply-side factors than demand-side factors.

We further investigated whether these average effects mask important heterogeneities. We find some suggestive evidence that the preference for women is stronger for white candidates than for black candidates. However, on the whole, our results do not depend on other candidate characteristics (such as experience, age, or occupation) and therefore do not support the hypothesis that voters systematically apply double standards when they evaluate female candidates.

The interactions of candidate gender with respondent characteristics are similarly small and inconsistent, though the stronger preference for women among Democrats compared with Republicans appears relatively robust. 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 Republicans for male candidates in general.

These surprising results raise a number of methodological concerns. Specifically, we may worry that our results reflect methodological limitations of candidate choice experiments, rather than actual preferences of voters, some of which we consider here.

First and foremost, we may be concerned that the observed response patterns are due to *experimenter demand effects* (Zizzo 2010). 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 behave 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).[2] For example, response patterns did not differ systematically when

[2]After all, 23 out of the 30 studies we consider here were administered on online samples.

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).

A related methodological objection is that findings from candidate choice experiments may not reflect voter preferences accurately, because the surveys do not accurately reflect *actual voters*. Most survey experiments rely on online convenience samples provided by platforms such as Amazon's Mechanical Turk, whose subject pool tends to be younger, unmarried, and more liberal (Levay et al. 2016). But even nationally representative samples may not be representative of the actual *electorate*, which is the true population of interest in the context of candidate choice experiments.

To assess the viability of this explanation, we post-stratified the estimate of the gender effect in the Kirkland and Coppock (2018) study to the 2016 electorate as estimated by the national exit poll.[3] Unadjusted, the estimate is 3.9 points with a standard error of 0.9 points; post-stratified, the estimate descends to 3.3 points (SE 1.3 points). The disjuncture between candidate choice experiments and the electoral success of women in the real world does not appear to be explained by differences between the national population and the 2016 electorate.

Even if bias due to experimenter demand or non-representativeness is relatively mild, candidate choice experiments may be artificial in ways that compromise the conclusions we draw from them. We view this as the most salient critique of our findings, and of conjoint analyses of voter preferences in general. Conjoint and factorial experiments typically present generic descriptions of hypothetical political candidates rather than profiles of real-world candidates running in the actual districts in which respondents reside. Hainmueller et al. (2015) 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. However, no such studies exist in the context of *candidate* choice experiments.

In particular, voters may respond to a hypothetical female candidate differently than they

[3]For now, we focused our attention on Kirkland and Coppock (2018) simply because the data for this study was collected in the most recent presidential election year and provided us with sufficient demographic details on the respondents, which made the post-stratification possible. In the future, we hope to perform similar analyses on other datasets considered in this paper to find further support for the argument.

would to a real-world female candidate. Just as generic horserace polls can overstate support for less well-known candidates (McGhee and Baldassare 2004), support for female candidates may deflate over the course of a campaign. Unfortunately, the experiments provided in this analysis cannot shed light on whether the particular versus generic distinction explains the divergence between survey experimental results and real-world underrepresentation of women. We leave the evaluation of this hypothesis to future research.

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.

Bermeo, Nancy and Jasmine Bhatia. 2017. "Strongmen or Technocrats? Experimental evidence testing leadership preferences in Afghanistan." *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*. http://estimatr.declaredesign.org, https://github.com/DeclareDesign/estimatr.

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.

Burns, Nancy, Kay Lehman Schlozman and Sidney Verba. 2001. *The private roots of public action*. Harvard University Press.

Carnes, Nicholas. 2016. "Keeping Workers Off the Ballot: Electoral gatekeepers and the shortage of candidates from the working class.".

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.

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.".

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

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; Salt Lake City* 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.

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.

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.

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.

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

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

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

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.

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.

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.

Levay, Kevin E., Jeremy Freese and James N. Druckman. 2016. "The Demographic and Political Composition of Mechanical Turk Samples." *SAGE Open* 6(1):2158244016636433.

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.

McGhee, Eric and Mark Baldassare. 2004. "Estimating Vote Choice: A New Look at the Generic Congressional Ballot." *Public Opinion Quarterly* 68(1):81–93.

Mendoza, Saaid A. and Marissa G. DiMaria. 2018. "Not "With Her": How Gendered Political Slogans Affect Conservative Women's Perceptions of Female Leaders." *Sex Roles* .

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.

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

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 / Samuel L. Popkin.* Chicago: 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.

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.

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.

Taylor, Paul, Rich Morin, D'vera Cohn, April Clark and Wendy Wang. 2008. A Paradox in Public Attitudes. Men or Women: Who's the Better Leader. Technical report Pew Research Center.

Teele, Dawn, Joshua L. Kalla and Frances McCall Rosenbluth. 2017. "Faces of Bias in Politics: Evidence from Elite and Voter Conjoint Experiments on Gender.".

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.".

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics* 13(1):75–98.