

Parallel Conjoint Experiments for Measuring Gender Stereotypes and Analyzing Preferences*

Santiago López-Cariboni[†] Catherine Reyes-Housholder[‡]

Abstract

Identifying the effects of cognitive cues on political behavior is remarkably difficult because researchers cannot control subjects' assignment to stereotypes. Unlike easy-to-randomize traits such as candidates' race or gender, stereotypes are unknown quantities describing how voters associate these descriptive characteristics to a multivariate distribution of missing information about relevant candidate attributes. Because these cues are fundamentally unobservable and likely to vary across subjects, identification of their effects remains a challenge. This paper introduces a novel method based on parallel conjoint design that allows for measuring stereotypes and analyzing their effects on voters' preferences. We implemented original studies on political gender stereotypes in the US, Chile, and Uruguay. While a large stream of factorial candidate sex experiments has documented a small pro-female advantage, we find that feminine stereotypes have even larger positive effects on preferences. We also detect heterogeneity of preferences among different subgroups indicating that stereotypes operate in complex ways. Our method could generalize to other kinds of stereotypes based on race, ethnicity, and other political identities.

*Preliminary draft prepared for the International Methods Colloquium on March 18, 2022, sponsored by Wake Forest University. We thank the excellent research assistance by Luciana Cantera, comments and suggestions from participants at the 2019 EGEN conference at Yale University, and at the 2020 American Political Science Association meeting. Generous financial support was provided by the Centre for Social Conflict and Cohesion Studies (COES) through the ANID/FONDAP/15130009 grant as well as through Fondecyt grant #3190196.

[†]*Departamento de Economía, Universidad de la República, santiago.lopez@cienciassociales.edu.uy

[‡]*Instituto de Ciencia Política, Pontificia Universidad Católica de Chile*, creyeshousholder@uc.cl

Introduction

A common set of explanations for women’s underrepresentation in elected offices worldwide points to voter preferences. Studies from the U.S. suggest that while gender stereotypes often play a role in voter evaluations (Huddy and Terkildsen, 1993; Alexander and Andersen, 1993; Kahn, 1996) voters do not seem to penalize women candidates or hold them to double standards (Dolan 2014; Brooks 2013). Moreover, a meta-analysis of 67 conjoint experiments from 12 countries and at different levels of government reveals a small pro-female bias among voters (Schwarz and Coppock, 2022). Voters nevertheless may support male candidates strategically if they view female contenders as less electorally viable (Bauer, 2020*b*; Lawless, 2010; Oliver and Conroy, 2020; Sanbonmatsu, 2006).

What role, then, do gender stereotypes, defined as cognitive associations between candidates’ sex and attributes, play in voters’ assessments? Measuring multi-dimensional stereotypes and evaluating their electoral performance in a causal framework remains a methodologically challenging but crucial task for our understanding of gender and politics. Recent work shows that few gender stereotypes about “women in general” apply to the sub-type of “female politicians” (Schneider and Bos 2014). Research devoted to measuring associations between female politicians and a multitude of traits, however, cannot by itself tell us much about whether voters use these stereotypes to assess candidates’ viability and make decisions about which candidates to support. Moreover, while candidate sex experiments usefully identify causality (Schwarz and Coppock, 2022), they cannot reasonably interpret causal effects of candidate sex as due to gender stereotypes rather than, for example, a baseline preferences for greater or less descriptive representation of women. Masking problems inherent in candidate sex studies (Bansak, Hainmueller, Hopkins and Yamamoto, 2018; Dafoe, Zhang and Caughey, 2018) further limit our understanding of which gender stereotypes, if any, might causally affect vote choice.

This paper introduces a novel method to measure stereotypes and their causal effects. We define stereotypes as voters’ associations between candidate sex and other candidate attributes. In survey experiments, this resembles the masking process by which respondents given with a piece of information such as candidate sex impute missing data about candidate quality (Bansak et al., 2018; Dafoe, Zhang and Caughey, 2018). We exploit this phenomenon as a research opportunity to map the multivariate distribution of candidate traits on a latent dimension of gender stereotypes.

We develop a parallel conjoint design that allows for measuring stereotypes and analyzing their effects on voters’ preferences. The design, pre-registered in EGAP, consists of two

experiments conducted in parallel.¹ The key feature of our first conjoint experiment, which occurs with a "measurement sample," involves withholding information about candidates' sex and turning it into an outcome variable. This facilitates identification of the high-dimensional content of gender stereotypes and the AMCE's on candidate profiles' perceived probabilities of being male or female. We use results from the measurement sample to train a machine learning algorithm to classify profiles as either masculine or feminine. The second experiment uses an independent "testing sample" to conduct in parallel a standard candidate choice experiment. This enables the estimation of the causal effects of stereotypes on vote choice and viability perceptions. In this testing sample, we classify all candidates profiles as masculine or feminine using the estimates from the measurement sample. Finally, we provide guidance to applied researchers on how to implement stereotypes studies and a discussion about the key methodological decisions regarding their empirical design.

1. From Candidate Sex to Gender Stereotypes

We argue that research to date on candidate sex, gender stereotypes, and voter evaluations features three main limitations (1) candidate sex studies present interpretive challenges concerning a potential role of gender stereotypes; (2) studies measuring gender stereotypes have yet to estimate their multi-dimensional effects in an electoral context; (3) research in these veins usually is conducted in a single-country context. These limitations motivate our goal to identify the content of multi-dimensional gender stereotypes and estimate their causal effects in an electoral context across three countries.

1.1 Candidate sex studies

Candidate sex experiments have advanced scholarship by consistently and credibly estimating a small causal impact of candidate sex on vote choice (Schwarz and Coppock, 2022). Such results have been interpreted as evidence of little discrimination against female candidates in the aggregate. These designs estimating the average marginal component effects (AMCEs) of candidate sex nevertheless can reveal little about the complex role of gender stereotypes on voter evaluations of political candidates (Sanbonmatsu, 2002). First, it could be that candidate sex affects vote choice through specific gender stereotypes. Decades of gender stereotypes research shows that candidate sex can signal a plethora of secondary traits relevant for voter decision-making (Dolan and Sanbonmatsu, 2011),

¹Pre-registered in August 2020 in EGAP 20200824AA.

for example, leadership abilities (Alexander and Andersen, 1993); issue competencies (Huddy and Terkildsen, 1993; Holman, Merolla and Zechmeister, 2016); ideology (?) and moral integrity (Barnes and Beaulieu, 2014, 2019; Le Foulon and Reyes-Housholder, 2021). Serving as an informational cue, candidate sex can exert indirect effects on vote choice with gender stereotypes playing a potential mediating role in producing such outcomes. Voters therefore may, on average, prefer female candidates for instrumental reasons.

Alternatively, candidate sex could directly impact vote choice according to gender ideology: feminist/modern vs. masculinist/traditional (Duerst-Lahti and Kelly, 1995; Best and Puzio, 2019). Voters may prefer females or males not because of gender stereotypes—the traits associated with the sexes—but rather because of their sex. Feminist voters may possess a baseline preference for greater descriptive representation of women, which motivated them to express greater support for hypothetical female candidates (pro-female bias/feminist). These kinds of direct effects of candidate sex on vote choice can also be expressed as preferences for greater descriptive representation of women. Conversely, masculinist or tradition-oriented voters may not like the idea of more females in politics, motivating them to support male candidates (pro-male bias/sexist). Such direct effects of candidate sex on vote choice can also be expressed as preferences for the male-dominated status quo. Such direct effects motivated by gender ideology remain distinct from calculations based on instrumental reasons. Because candidate sex studies do not measure gender stereotypes, they cannot adjudicate among each of these possibilities.

Another way to express the limitations of candidate sex studies emerges from the explicit goal in conjoint designs of avoiding masking, which is minimized with all other attributes included in hypothetical candidate's profile. "Without information on the full set of relevant attributes, estimates of an AMCE of interest may be masking the effects of other, correlated attributes" (?) pg 1-2). Candidates' attributes listed in survey tables are often inter-related to candidate sex. Therefore, interpretations of candidate sex AMCEs as a quantity of interest in conjoint experiments potentially conflates an unknown fraction of the effect of gender stereotypes and the direct demand for descriptive representation regardless of any other relevant attribute. More explicitly, AMCE in conjoint experiments potentially are reporting a mixture of 1) the direct effect of candidate sex on vote choice (i.e. the existing effect politician's sex after all politically relevant individual characteristics are included and when they perfectly manipulate respondents), 2) a fraction of the effect of candidate sex that occurs through gender stereotypes resulting from either traits that are not included in the conjoint table or from traits that only imperfectly manipulate respondents beliefs. Importantly, in the case of demand for descriptive representation,

or a direct effect of candidate sex, respondents should perceive candidate sex as an as-if-random candidate trait, this is, an individual characteristic that is orthogonal to all other politically relevant characteristics. This is impossible to identify in the practice of survey experiments but is theoretically intuitive.

Thus, unless there is a positive demand for female descriptive representation (i.e. presence of feminist ideology), the effect of candidate sex should be necessarily zero because the effect of stereotypes would be muted under a context of perfect information (also called "eliminated effect" in (Acharya, Blackwell and Sen, 2018)). The estimated AMCE effect then is a fraction of the total effect of gender stereotypes plus the demand for female elected politicians. A limitation from this type of designs is not only that they report a partial effect of the role played by stereotypes but that, more fundamentally, such fraction is completely indeterminate: researchers cannot know what is both the unobserved role of stereotypes on voting due to omitted traits or due to imperfect manipulation of respondents beliefs by the included traits in the conjoint table.

Finally, gender stereotypes could still exist, but not affect vote choice (Dolan, 2014). For example, it could be that voters associate female candidates with a lower or a higher probability of having children (Teele, Kalla and Rosenbluth, 2018), but if voters care little about the quantity of candidates' children, such a gendered association may not directly affect vote choice. This could render the gender stereotype politically irrelevant in a strict sense, although such societal beliefs could matter in other stages of candidate selection processes, for example candidates' own decisions to run for office. Moreover, even if a specific gender stereotype does not affect baseline voter preferences, it could influence voters' views on whether the candidate meets traditional qualifications for office (Oliver and Conroy, 2020) or a candidate's probability of winning, that is, their viability.

In short, dozens of experiments estimate the causal effect of candidate sex on vote choice. This scholarship is converging on the idea that candidate sex tends to favor females, but overall matters little in terms of magnitude (Schwarz and Coppock, 2022), particularly compared to variables such as party ideology. While ground-breaking, such a conclusion could mistakenly downplay the relevance of gender in voters' decision-making (Dolan, 2014). These experimental studies have yet to fully grasp the complex, and nuanced ways in which candidate sex and gender stereotypes influence voter preferences. Such candidate sex studies end up telling us little about what people are thinking when they observe a male or female candidate in the abstract. Studies seeking to estimate the AMCE of candidate sex tend to gloss over masking problems, muddling interpretations.

Our work takes a fresh approach by providing first a new empirical strategy to measuring multi-dimensional gender stereotypes, and then, estimating their effects on vote choice and viability perceptions. This is particularly important because it remains unclear whether voters associate female presidential candidates with the same traits that they attribute to women in general (Schneider and Bos, 2014), a point further developed below.

1.2 Limitations of gender stereotype measurement studies

Political psychological studies, drawing from social role and personality theories, have developed myriad ways to classify traits associated with men or women. Each of these classifications converge on the notion that gender stereotypes are multiple. For example, Alice Eagly differentiates between assertive and communal traits, which have implications for stereotypically masculine and feminine leadership styles (Eagly and Karau, 2002; Eagly, Eagly and Carli, 2007). The assertive vs. communal distinction is widely used in politics and gender literature on the role of femininity on voter preferences (Bauer, 2015, 2020a).

The concepts of “masculinity” and “femininity” implicitly acknowledge that gender stereotypes are actually multi-dimensional: citizens tend to associate many characteristics with the male or female sex. Much work on the role of masculinity and femininity in politics draws on Bem’s Sex Role Inventory of idealized traits of males and females (Bem, 1974; Rosenwasser and Dean, 1989; McDermott, 2016; Bauer, 2020a). Building on Spence and Helmreich (1978), Oliver and Conroy (2020) conceptualize masculinity and femininity as sets of personality traits. Thus, it is not enough to study a gender stereotypes as stand-alone associations between candidates’ sex and candidates’ traits. It is more appropriate to conceptualize gender stereotypes in multi-dimensional terms, that is, as constellations of traits that may reinforce perceptions or interact with each other to influence perceptions of candidates’ masculinity or femininity.

Recent research from political psychology nevertheless suggests that while citizens attribute similar traits to men in general and men in politics, stereotypes for women in general do not overlap with stereotypes concerning women in politics (Schneider and Bos, 2014). Schneider and Bos’ (2014) ground-breaking study, again conducted in the U.S., suggests that female politicians may be losing on male stereotypical qualities while also not having any advantage on qualities typical of women, such as empathy and compassion. These last two traits are particularly relevant for studies of presidential candidates since empathy is routinely mentioned as a quality that voters desire in presidential candidates (Holian and Prysby, 2014; Kinder, Peters, Abelson and Fiske, 1980).

1.3 Limitations of single-country studies

Many gender stereotypes and masculinity-femininity expectations at the societal level remain consistent across much of the world (Sczesny, Bosak, Neff and Schyns, 2004). However, cross-national consistencies may not apply to multi-dimensional stereotypes of male and female political candidates. Gender stereotypes are fundamentally cultural constructions and hence context-dependent (Best and Puzio, 2019). Women's enormous, and yet variable, gains in political offices across the globe, in particular thanks to gender quotas (Franceschet, Krook and Piscopo, 2012), might lead to the expectation that gender stereotypes, as a result, have evolved over time but unevenly in different parts of the world.

Extant research on gender stereotypes for male and female candidates tell us little about whether and how they vary across countries. Much of the pioneering work cited above is based on the case of the United States, the paradigmatic presidentialist regime. This may produce a U.S. bias in our understanding of the role of candidate sex and gender stereotypes on viability and vote choice. The candidate sex studies included in the meta-analysis (2021) conducted outside of the United States also feature single-country contexts. One study not included in this meta-analysis that did apply the same conjoint design to multiple countries shows that the causal impact of candidate sex on vote choice does vary cross-nationally. Le Foulon and Reyes-Housholder (2021) found that while the AMCE for candidate sex in hypothetical mayor election marked by corruption accusations was significant in Uruguay, with Uruguayans preferring female candidates, it was not significant in Argentina or Chile. Again, these candidate sex studies do not measure gender stereotypes. Another recent study showed that stereotypes about female politicians in Brazil differed significantly from stereotypes about female politicians in the U.S. (Lucciola, N.d.). In particular, Brazilians seemed to associate female politicians rather than their male counterparts with several valued traits. No study has conducted the same experiment measuring gender stereotypes and their impact on presidential vote choice in multiple countries.

To sum up, no experimental study has yet sought to uncover the content of gender stereotypes and perceptions of viability for male and female candidates. We therefore pursue three empirical goals: (1) understand which campaign-relevant candidate attributes voters associate with male and female presidential candidates (multi-dimensional content of gender stereotypes); (2) evaluate the electoral performance of perceptions of masculinity and femininity of hypothetical candidate profiles; (3) compare these results cross-nationally. We explain below how our design enables us to pursue these objectives in our parallel

conjoint design.

2. High-Dimensional Gender Stereotypes and Voter Evaluations

This paper adopts a theoretical perspective in which voters evaluate candidates according to information provided about these candidates' traits. We maintain that voters, in deciding who to support, consider a variety of pieces of candidate information, which routinely include their ideology, experience, reputed leadership style, and select policy positions. Respondents may process information in possibly competing ways. One is the realistic Bayesian model where respondents react to a given survey manipulation by updating their beliefs about any attribute that in the real world is correlated with the provided information. As in the logic of statistical discrimination (Guryan and Charles, 2013), a political candidate described as a "male" is likely to convey information about background conditions and therefore induce respondents to update their background beliefs. An alternative process is the non-realistic model of information processing where stereotypes and heuristics do not necessarily relate to objective data (Kahneman and Tversky, 1973), which would also lead to updating beliefs on background characteristics. Dafoe, Zhang and Caughey (2018) explain how these two cognitive processes produce imbalance in other related background characteristics in survey experiments leading to the violation of "information equivalence." The consequences of this violation, regardless of the underlying information processing model, is analogous to confounding in observational studies.

We first distinguish among three core concepts (1) candidate sex; (2) single gender stereotypes, (3) and multi-dimensional gender stereotypes. We employ the term "candidate sex" to refer to information of whether a hypothetical candidate is "male" or "female." One-dimensional or single gender stereotypes refer to cognitive associations between "male" or "female" and specific attributes, such as "aggressive" or "compassionate." Crucially, the presence of gender stereotypes does not imply discrimination against female voters, as many traits associated with women in politics can also be desired by voters.

It is the third concept, multi-dimensional (or similarly "high-dimensional") gender stereotypes, that takes center stage in our study, its research design and data analysis. Political candidacies, particularly presidential ones, typically transmit more than one relevant attribute (Hainmueller, Hopkins and Yamamoto, 2014a). Gendered identities, similarly, remain multi-dimensional (Dolan and Sanbonmatsu, 2011). We therefore conceptualize

high-dimensional gender stereotypes as voters' perceptions of whether a particular candidate profile overall is more likely to be male or female. We also refer in this paper to high-dimensional gender stereotypes as measures of hypothetical candidates' "masculinity" and "femininity," as perceived by survey respondents. Single gender stereotypes about either men or women candidates could may interact with each other to move voters, to different degrees, to view hypothetical candidates as more likely to be either male or female. Our interest in the high-dimensionality of gender stereotypes thereby motivates our use of conjoint designs.

We expect male and female stereotypes in electoral contexts to differ not only in terms of content, which a vast literature cited in section 1 has argued for decades, but also in terms of precision. More specifically, we expect voters to stereotypically associate particular traits more strongly with female candidates than (albeit different) traits with male candidates. One reason for this is not that voters prefer male candidates, but rather that men's historical dominance in politics, particularly in the presidential arena, implies a different, indeed more demanding selection processes for women (Hinojosa, 2012). As a result, the women who emerge as presidential candidates and presidential winners appear as a far more homogenous group than their male counterparts. Historical male dominance—and the accompanying differences in (informal) selection processes—further means that voters in real life have observed a wider variety of male candidacies.²

Voters have observed not only fewer female candidates for president, but different, more demanding selection processes for women results in a relatively more homogenous group of real-life female contenders. This greater homogeneity could make it easier for voters to pin down the candidate traits associated with female contenders: some presidential traits are similarly very difficult to imagine for a female presidential candidate. For example, connections between the military and the presidency are one of the classic pieces of evidence that the presidency is a masculinist institution, which means—among many other things— that people associate authoritarian leadership with males and “good” presidents (Duerst-Lahti and Kelly, 1995). No woman from the military has ever run for president, but several military men have done so and won in the U.S. (i.e. George H.W. Bush) and

²It is not hard to call to mind a male presidential candidate who has served in the Senate (i.e. Barack Obama of the U.S.); another who is an academic (i.e. Fernando Henrique Cardoso of Brazil); and still another famous for his social activism (i.e. Luiz Inácio “Lula” da Silva of Brazil). It is also not hard to imagine a male candidate known for his authoritarian decision-making and another who is known for his skills at achieving consensus through dialog. Indeed, both forms of decision-making may be required for a successful presidency (?). Many voters could probably name a male presidential candidate who they have observed with each of the candidate attributes common to classic tables of conjoint designs. Thus, men's historic dominance in presidential politics moves us to expect less precise notions of which attributes, recurrent in classic conjoin designs, voters associate with male candidates.

Latin America (i.e. Jair Bolsonaro of Brazil; Hugo Chávez of Venezuela). Similarly, few female CEOs have run for president (with the notable exception of former H.P CEO Carly Fiorina who ran for the U.S. Republican Party nomination in 2016), but many men have, for example Donald Trump of the U.S. and Sebastián Piñera of Chile. Because of this, it is unlikely for voters to stereotypically associate female candidates with big business. Thus, female stereotypes are not only different from male stereotypes in terms of content, but we also expect more precise notions of which attributes, recurrent in classic conjoin designs, voters associate with female candidates.

High-dimensional gender stereotypes constitute a cultural construct, and thus can vary across countries. In some parts of the world, for example the U.S. and other industrially advanced democracies, female candidates may be associated with more progressive ideologies (Inglehart and Norris, 2003). The opposite could be true in some Latin American countries where public opinion polling suggests that women historically tend to be more conservative than men (Morgan 2015). Because descriptive stereotypes can vary across cultures, voters from different countries may process the same information about hypothetical candidates in contrasting ways.

We theorize the impact of high-dimensional gender stereotypes on two outcome variables: vote choice and viability perceptions. Voters not only use pieces of information about candidate attributes to assess whether to offer electoral support; they also may use the same or different pieces of information to assess how likely a candidate may be to win an election. Conceptually distinct from individual-level preferences, viability perceptions reflect how popular respondents believe a candidate profile to be among the broader voting population. Sometimes, the same traits that strengthen a candidate's appeal also bolster perceptions of a candidate's viability. This often occurs in the context of valence issues, defined as issues upon which virtually all voters agree such as the need for honest politicians and economic growth. The fact that a candidate is free of corruption could enhance voter support for the candidate and voter perceptions of the candidates' viability. Voters may reason that "I like this trait, and it is likely that others like this trait too." Conversely, some traits that sub-groups of voters do not like, for example "uses authority" may diminish these voters' support for the candidate, but not necessarily their perceptions of the candidates' viability. Such a divergence between some voters' individual preferences and these same voters' viability perceptions occurs when voters reason that "I like this trait, but most other people probably do not like this trait," thereby rating viability comparatively worse.

Voter preferences thus may dovetail or diverge from viability perceptions. These relationships may be gendered. Intuition-based ideas concerning whether female candidates are

qualified to run—or whether others will view them as such—can influence voters’ strategic decision-making and elections (Murray, 2010; Oliver and Conroy, 2020; Sanbonmatsu, 2006). Voters may express preference for a female, or gender-congruent feminine, candidate while acknowledging that another male, or masculine, candidate is more likely to win an election. Specifically, it could be that voters, on average, prefer traits associated with female candidates (femininity), but given men’s historical dominance in the presidency, they also may view these traits as less electorally viable. We therefore theorize electoral support and viability perceptions as distinct, potentially related, and deeply gendered outcomes.

Whether and to what degree gender stereotypes reduce, enhance or exert no effect on viability perceptions remains especially relevant in the settings of party primaries where voters must select a candidate to compete against other party-selected candidates in a general election. Even if voters prefer a particular female candidate or attributes associated with that the female sex, they may calculate that the female candidate is less likely to win. Voters who fear wasting their vote, particularly in tight elections, may decide to vote for a male candidate because of these viability concerns.

Our first hypothesis therefore seeks to confirm whether stereotypes in presidential politics exist at all. If voters come to associate candidate sex to electorally-relevant traits, such as those routinely included in extant conjoint designs, then relevant conclusions about candidate quality may result from cognitive cues from stereotypes of male and female politicians. Moreover, stereotypes hardly are limited to countries that have already experienced female presidents. We also expect voters to associate particular traits with either male or female hypothetical candidates even in countries, such as Uruguay, where no woman has ever run for president.

We build on the vast research discussed in Section 1, which converges on the notion that myriad traits can determine gender stereotypes, or societal notions of masculinity-femininity, in electoral contexts.³ We further expect certain types of interactions between different candidate traits to reinforce ideas about stereotypes, or to the contrary, undermine associations between an entire candidate profiles and candidate sex. Given that more than one attribute influences voters’ beliefs, individual candidate attributes will interact with each other to disproportionately augment or decrease perceptions that a particular candidate is male or female. For example, it could be that the combination of experience as

³Again, many traits associated with "women in general," for examples those associated with physical appearance may not apply to female presidential candidates. Thus, our results for the determinants of "femininity" apply only with regards to hypothetical candidate profiles for the presidency.

a primary school teacher and never being accused of corruption together augments voters' beliefs that a profile corresponds to a female presidential candidate more than teaching experience and a clean record by themselves. Similarly, it could be that "desire for power" combined with an authoritarian decision-making style augments voters' beliefs that a profile corresponds to a male presidential candidate more than "desire for power" and authoritarian decision-making by themselves. Finally, we expect voters to more precisely associate traits with female rather than male candidates, due to men's historical dominance.

H1: Candidate attributes and their interactions affect voters' beliefs of whether a particular candidate profile would correspond to a female or male presidential candidate.

Our second hypothesis is that gender stereotypes influence vote choice and viability perceptions.

H2: High-dimensional gender stereotypes influence vote choice and viability perceptions.

3. Methods and Research Design

Assumptions and quantities of interest

Our goal is to identify the effect of gender stereotypes $S \in \{\text{"masculine"}, \text{"feminine"}\}$ on a given outcome Y such as electoral support, viability perceptions, or other kinds of voter evaluations. Multi-dimensional gender stereotypes are in turn an unknown aggregation of beliefs about male and female candidates' background characteristics B . Thus for a sample of N subjects i we have,

$$Y_i = \theta S_i + \epsilon_i \tag{1}$$

This imposes two challenges. The first one is measuring the multi-dimensional content of stereotypes while second is ensuring exogeneity conditions to learn about their causal effects. A key difficulty of estimating these effects refers to the requirement of manipulating subjects' perceptions about someone's gender.

The definition of stereotypes outlined before refers to a specific form of *in-equivalence* of information about background features produced by a single informational trait such as candidate gender. This is likely to occur if individuals behave as Bayesian information

updaters.⁴ Existing literature shows why this process of "masking" is known to occur in survey experiments and imposes a challenge to causal inference through the violation of "Information Equivalence of Background Features" even when the informational treatment Z is randomly assigned (Dafoe, Zhang and Caughey, 2018):⁵

$$B_i(Z_i = 1) \neq B_i(Z_i = 0)$$

For the purposes of this paper, we show that a valid strategy to measuring gender stereotypes is precisely estimating the form of the violation of this assumption. This is the difference in background beliefs that results from the target information Z . Thus, stereotypes S can be characterized as follows:

$$S_i = B_i[Z_i = 1] - B_i[Z_i = 0]$$

Note that this definition is consistent with any form of Bayesian updating. Whether voters adjust beliefs in a realistic or non-realistic fashion is irrelevant. Moreover, the definition provides clear guidance about how to estimate the objective structure of *information inequivalence* on background characteristics that stereotypes fundamentally imply.⁶ Since we care about measurement, we can assign subjects to different realizations of candidate characteristics B and then collect reported beliefs of candidate gender.

⁴Respondents may assign probabilities to profiles given information about a single trait Z , such that $p(D = d_k, B = b_k | Z = z) = \frac{p(Z=z|D=d_k, B=b_k)p(D=d_k, B=b_k)}{p(Z=z)}$. For example, $p(Z = \text{"female"} | D = \text{female}, B = \text{seeks consensus})$ is the probability that an individual believes in a female politician who seeks consensus being described as a "female". Some individuals are thus likely to assign a high value to the preceding quantity, and a low value to $p(Z = \text{"female"} | D = \text{female}, B = \text{uses authority})$.

⁵Say that a causal factor D^* that affects the outcome Y^* in the real world. A researcher may manipulate assignment of every subject i to a textual element $Z_i \in \{0, 1\}$ intended to change subjects' beliefs $D_i \in \{0, 1\}$ about the causal factor of interest. If Z_i is randomly assigned imposing an exogenous variation in D_i , a difference in means across experimental conditions is an unbiased estimate of the "intent-to-treat" effect: $ITT = E[Y_i(Z_i = 1) - Y_i(Z_i = 0)]$. However, if the manipulation of Z also affects background beliefs B that, in turn, have a direct effect on the outcome, then the IV "exclusion restriction" by which Z only affects the outcome Y through its effect on D no longer holds. Thus, it is not sufficient to randomly assign subjects to different Z conditions in a survey experiment, subjects' beliefs about background characteristics should also be the same.

⁶A standard practice in political psychology is in fact to collect data about this difference for a number of different attributes. Respondents are for example asked to mark all the traits that they believe better describe a "man" and those that better describe a "woman" (Schneider and Bos, 2014; Lucciola, N.d.). Such strategy is, however, unable to map stereotypes systematically across a predefined population of profiles.

$$S_i = B_i[Z_i = 1] - B_i[Z_i = 0] = Z_i[B_i = 1] - Z_i[B_i = 0] \quad (2)$$

The underlying assumption is that the effect of candidate gender Z_i on beliefs about a background feature B_i is the same as the effect of that background feature on beliefs about candidate gender. This reflects the fact that heuristics are cognitive processes of associations between different elements. For the purposes of measurement both ways of mapping these associations are valid under unconfoundedness or exogeneity. We explain below why the second strategy is better suited for estimating causal effects of stereotypes on other political outcomes.

Parallel conjoint experiments

We propose a particular measurement experiment that maps candidate stereotypes $S_i \in \{\text{"masculine"}, \text{"feminine"}\}$ across a set of "background beliefs" of gender. These background beliefs B define a high-dimensional population of candidate profiles that withhold information about candidate sex or gender (man/woman). Instead, candidate gender is the outcome variable in the measurement experiment. As in any other conjoint design, this allows estimation of the effect of individual attributes on the change in perceptions of candidates' gender through estimation of average marginal component effects (Hainmueller, Hopkins and Yamamoto, 2014b). AMCEs may be therefore used to describe the *content* of stereotypes (in our study, the attributes associated with either male or female presidential candidates).

Under the aforementioned assumption that gender cues involve two-way associations rather than a single directional causal path between gender and background traits, we take the gender perceptions (experimental outcome) of candidate profiles (treatments) as the information that, if provided, would produce an equivalent change in background beliefs. This allows for an estimate of the structure of information inequivalence on background beliefs implied in the concept of gender stereotypes. The measurement design not only naturally relates to our definition of stereotypes but also allows for a systematic classification across a predefined population of candidates. In a parallel candidate choice experiment, we assign subjects to different stereotypical candidates by drawing profiles from the same population of hypothetical candidates used in the measurement experiment.

Consider for a sample of N units a factorial design of J factors where each factor

$j \in \{1, \dots, J\}$ has $L_j \geq 2$ levels. Thus the treatment for the unit i , denoted as B_i , is J -dimensional vector of random variables defining background beliefs of gender, each of which represents the assigned level of the corresponding factor variable. For instance, the j th element of this random vector $B_{ij} \in 0, 1, 2, \dots, L_j - 1$ represents the level of factor j . Hence, following Dasgupta, Pillai and Rubin (2015), the potential outcome in a fully randomized factorial design is defined as $S_i(\mathbf{b})$, where $\mathbf{b} \in B^*$ represents the realized treatment with B^* being the support of the randomization distribution for B_i . Thus, the observed outcome is given by $S_i = S_i(\mathbf{b})$.

Implementing uniform randomization of the treatment assignment implies,

$$\{S_i(\mathbf{b})\}_{\mathbf{b} \in B^*} \perp B_i \quad \forall i$$

Ensuring independence between potential outcomes of stereotypes and assignment to candidate characteristics rules out the possibility that pre-treatment covariates (i.e., respondent ideology) confound this relationship, and ultimately, induce a bias in how political candidates are classified in the candidate choice experiment. The design-based measure of stereotypes allows to directly relate the background characteristics to gender beliefs.

$$S_i = g(B_i) + v_i$$

Applying a function $g(\cdot)$ to the data provides a measure of *subjects' perception of a candidate's gender*, which fulfils the goal of mapping stereotypes across all candidate profiles. There are numerous statistical alternatives to estimate profile stereotypes. One could predict the outcome using the marginal means of each factor level (Leeper, Hobolt and Tilley, 2020). Since stereotypes may involve complex interactions between different traits, the latent utility model should allow for analyzing the effects of high-dimensional treatments. A growing literature in political science is adopting machine learning algorithms to flexible estimate the effects of factorial designs (Horiuchi, Smith and Yamamoto, 2018; Grimmer, Messing and Westwood, 2017; Ratkovic and Tingley, 2017; Egami and Imai, 2018; Bansak, Bechtel and Margalit, 2021; De La Cuesta, Egami and Imai, 2022; Goplerud, Imai and Pashley, 2022; Ham, Imai and Janson, 2022). For now, we employ regression forests in which prediction is the main goal.

Now we turn to outline how our design-based approach ensures exogeneity conditions of stereotypes to analyze their effects on political variables. A parallel conjoint design is implemented in a another independent sample, where the second experiment consists

of a standard candidate choice experiment. This "testing sample" of N' units enables the estimation of the causal effects of stereotypes on voters' evaluations. A main reason for working with different samples—one for stereotype measurement and another for causal effect testing—is that asking respondents about political support after they guess candidate sex induces a potentially large post-treatment bias (Montgomery, Nyhan and Torres, 2018; Chaudoin, Gaines and Livny, 2021). In expectation the two respondent samples are equivalent. This means that the estimated data generating process about from the measuring sample can be confidently used to predict gender stereotypes in the testing sample (standard candidate experiment). After data collection, we classify all candidate profiles in the testing sample as either "masculine" or "feminine" using the machine learning estimates from the measurement sample.

Armed with a predictive model of stereotypes, we evaluate it on the new data of profiles distribution corresponding to the testing sample where respondents participate of a candidate choice experiment. Since the stereotypes model predicts is a binary variable indicating candidate gender, the predicted classification for both training and testing samples, \hat{S}_i and \hat{S}'_i respectively, depends on the predicted latent variable S^* :

$$\hat{S}_i = \begin{cases} \text{"feminine"} = 1, & \text{if } S_i^* > 0.5 \\ \text{"masculine"} = 0, & \text{if } S_i^* < 0.5 \end{cases}$$

Provided that the $\mathbf{b}' \in \mathbf{B}^*$ now represents the realized treatment for the same support of the randomization distribution \mathbf{B}^* for \mathbf{B}'_i . This is, random assignment of the factors used to estimate stereotypes in the first experiment implies random assignment of the traits that describe perceptions about background beliefs of male and female candidates. Therefore, we can express the potential outcomes in the second experiment with respect to stereotypes of profiles as $Y_i(S'_i)$ with the observed outcome being $Y_i = Y_i(S'_i)$. Since we ensure assignment to stereotypes—that is, traits and their interactions which we learn are associated with either female or male presidential candidates—is independent from the potential outcomes,

$$Y_i(S_i) \perp S_i \quad \forall i$$

we estimate the following model assuming voters do not observe candidate traits used to estimate stereotypes,

$$Y_i = \theta \hat{S}'_i + \epsilon_i,$$

We estimate the effect of stereotypes as an average marginal component effect in conjoint designs. This involves adjusting standard errors by clustering at the respondent level. Also because stereotypes constitute a predicted variable, we implement bootstrapping such that the standard error of θ reflects prediction uncertainty.

Experiment 1: political gender stereotypes

We implemented the parallel conjoint design outlined above for measuring stereotypes and analyzing their effects on voters' preferences.⁷ The parallel design consists of two experiments, to which survey respondents are randomly assigned. The first occurs with a "measurement sample" of respondents which we use to train a machine learning algorithm to classify profiles as either masculine or feminine. Our design innovates here by withholding information about candidates' sex and turning it into an outcome variable. This identifies the high-dimensional content of gender stereotypes, that is, overall perceptions of the probability of an entire candidate profile corresponding to a male or female presidential candidate.

The outcome variable in the first experiment asks respondents to guess the candidate sex of hypothetical candidate profiles. We follow Schneider and Bos (2014) in asking respondents about their perceptions concerning what others in society would guess were male or female presidential candidates. This phrasing encourages participants to try to think beyond their own perceptions and consider what the broader society (as they view it) would think. The phrasing also helps reduce social desirability bias (Streb, Burrell, Frederick and Genovese, 2008).

Standard alternatives for asking about respondents' perceptions concerning whether profiles stereotypically correspond to men or women include the forced-choice and the non-forced questions. The forced choice version reads as follows: *Many times people in society have impressions of what men and women in politics are like. If you had to guess what other people would think, which of the two candidates would you say is a MAN/WOMAN*, where MAN/WOMAN are randomized with equal probabilities. The non-forced choice question is *Many times people in society [...]. If you had to guess what other people would think, what would you say is the gender of each candidate?* Both strategies offer advantages and disadvantages,

⁷Our survey received IRB (Ethics Committee) approval from the Universidad Diego Portales in February 2019 and then, spurred by changes in one of the author's affiliations, we obtained IRB approval from the Pontificia Universidad Católica in December 2020.

so we randomize forced choice and unforced choice questions. The results, however and remarkably similar using either question.

Respondents complete six tasks comparing pairs of candidate profiles that vary across the attributes presented in Table 1. While some attributes are often observed in the context of electoral campaigns, such as the age of candidates or their past experience, other kind of attributes related to candidate quality are more difficult to anticipate. Whether politicians, once in office, are going to incur in corruption or to adopt an authoritarian mandate style is something voters try to infer from other candidate traits such as their gender. The design allows for all traits to affect candidate sex perceptions. However, as we show below, researchers are free to choose what traits will be assumed to be "gendered" in the data analysis stage.

Table 1: Conjoint attributes

Attributes	Values
Age	36, 52, 68
Number of children	0, 1, 3
Social class	Lower-class, Middle-class, Upper-class
Experience	Business, Social activist, Union leader, Governor, US state representative, Primary school teacher, Academic
Corruption accusations	Never accused, Has been accused
Ideology	Left, Center, Right
Issue priority	Crime, Economy, Jobs, Health, Education, Poverty, Gender equality
The press mentions candidate's...:	Leadership, Moral integrity, Competence, Empathy, Desire for power
Self-declared "decision-making style"	Uses authority, Seeks consensus

Experiment 2: Standard candidate choice experiment

The second is a standard candidate choice experiment in which respondents observe the same traits as shown in Table 1 and then report both forced and non-forced choice questions of political support: *Which candidate would you vote for, or which do you prefer* (forced choice); and *On a scale of 1 to 7, where 1 indicates that you definitely would NOT support*

the candidate and 7 indicates that you definitely support this candidate, how would you rate each of the candidates described above? In this experiment we have also placed a second outcome question about viability perceptions. The viability question reads as follows: *Independent of your preferences, which candidate do you think is more likely to win a presidential election?* As in the first experiment, respondents evaluate six pairs of hypothetical candidate profiles.

4. Results

4.1 Gender Stereotypes for Hypothetical Candidates in the U.S.

Figure 1 shows results from the U.S. surveys, specifically the coefficients for the candidate attribute AMCEs on perceptions of whether a candidate is female. Here, we are, using only the measurement sample, measuring candidates' masculinity-femininity as a continuous probability with 0 indicating the maximum value of consensus that a profile corresponds to a male and 1 the maximum value of consensus that a profile corresponds to a female. The figure includes responses to the forced and non-forced versions of the question, which was prefaced by the statement designed for citizens to think about stereotypes rather than their own personal views: "Many times people in society have impressions of what men and women in politics are like. If you had to guess what other people would think..." Positive and significant AMCE coefficients mean that, on average, a particular attribute strengthens perceptions that a candidate is female; negative and significant coefficients strengthen perceptions that a candidate is male. The figure overall reveals that gender stereotypes for presidential candidates indeed exist and involve multiple attributes, as Hypothesis 1 expects.

Many of Figure 1's results appear intuitive in light of existing studies on gender stereotypes. For example, consistent with theories of gendered leadership styles and corruption research, the decision-making style of "uses authority" and corruption accusations diminish the probability of a candidate being viewed as female. Compared to the salient trait of "empathy" all other candidate traits—"leadership," "competence," "desire for power," and "moral integrity"—all reduce perceptions of candidates being female rather than male. Compared to crime, prioritizing the compassion issue of health and, to a greater magnitude, gender equality, augments this same probability. Our conceptualization of gender stereotypes as multi-dimensional helps motivate our exploration of possible interactive effects among attributes. In addition to estimating AMCEs, we leveraged the machine learning technique of Generalized Random Forests to identify possible causal interactions between candidate attributes, which improve precision of estimate of candidates

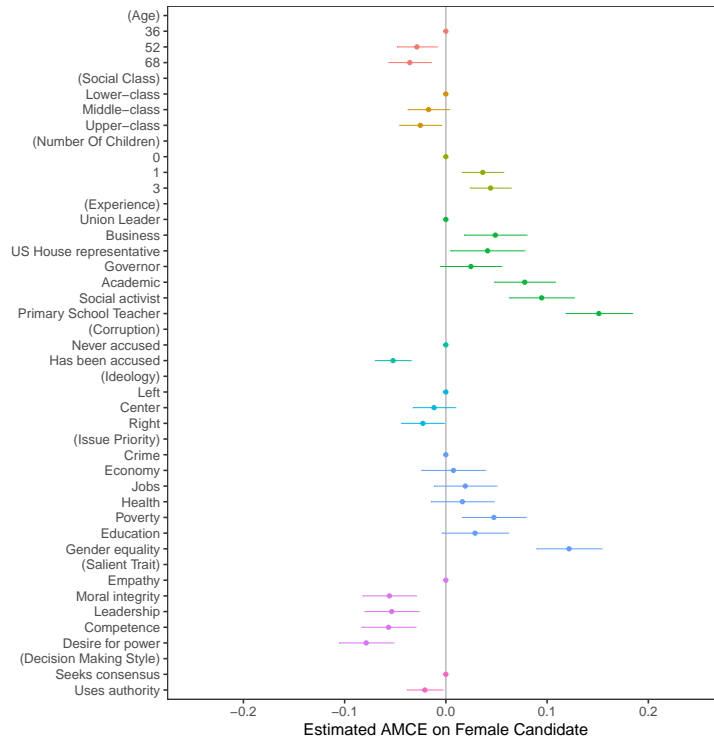


Figure 1: Stereotype Content: AMCEs on Perceptions of Candidates' Gender Based on Measurement Sample

masculinity-femininity.

Table 1 in the online appendix lists the top 10 most feminine candidate profiles and the top 10 most masculine ones. These results overall suggest that corruption served as the most important pieces of information that respondents use to stereotypically differentiate between female and male presidential candidates.

Do voters find it easier to identify which candidate is male or which candidate is female? To answer that question, we examine only results of the forced-choice questions concerning candidates' gender. "Many times people in society have impressions of what men and women in politics are like. If you had to guess what other people would think, which candidate do you think is a MAN/WOMAN." While the unforced question lets respondents choose freely to name, for example both candidates as male or both as female, the forced choice version offers a better view on how voters classify either males or females. The forced choice question primes respondents to think either about masculine or feminine cues.

Figure X shows the distribution of predicted probabilities of a candidate being classified as male (with a maximum value of 0) vs. female (maximum value of 1). Values closer to 0.5

mean greater uncertainty or less consensus about whether a candidate profile should be classified as male or female. Extreme values—those closer to 0 or closer to 1—reveal greater consensus concerning profiles' gender. Importantly, the distribution of Figure X is skewed to the right, meaning voters seemed to have greater consensus over the female-classified profiles than the male-classified profiles. This analysis is consistent with our theoretical expectations.

4.2 Stereotypes' Effects on Voter Evaluations

Figure 2 shows on the X-axis our rankings of candidate profiles' gender from 0 to 5000. 0 indicates the most masculine candidate profile and 5000 indicates the most feminine one. Here we classify the profiles using the machine learning (random forests) by considering the traits of corruption accusations, ideology, issue priority, salient traits and decision-making style.

The horizontal line of 0.5 marks our cutoff from classifying profiles as either stereotypically masculine or feminine. Because the blue line from 0 to about 3,500 lies below the cutoff, these profiles would be classified as masculine. However only profiles whose 95-percent confidence intervals lie below the cutoff line are classified with reasonable certainty as masculine. More profiles are classified with reasonable certainty as masculine than feminine. Less than about 1,000 profiles are classified with reasonable certainty as feminine. This is unsurprising given men's historical dominance in presidential politics.

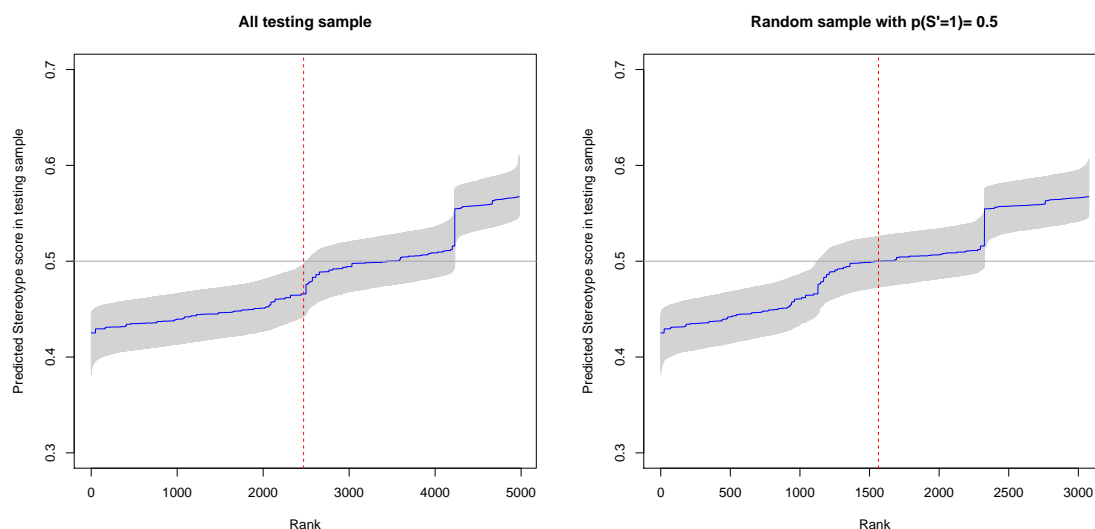


Figure 2: Stereotype predictions in the testing sample

Figure 3 shows on the left the distribution of feminine and, on the right, masculine profiles in terms of the Table 1 traits. Visualizing these data help bolster our claims to the informational inequivalence in assumption regarding background information on gendered traits. Some traits appear evenly distributed, suggesting that voters do not use them to guess the gender of candidate profiles. Other traits appear unevenly distributed, suggesting that voters rely on these traits to make their best guesses about candidate profiles' gender.

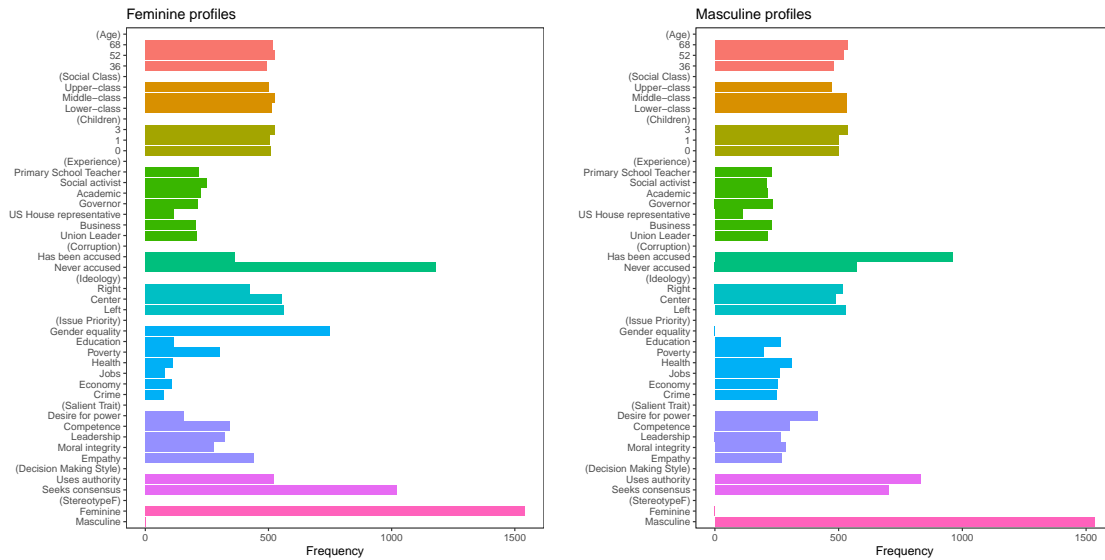


Figure 3: Informational inequivalence in background information on gendered traits (*Corruption, Ideology, Issue priority, Salient trait, Decision-making style*)

The feminine profiles feature a much higher quantity of profiles which have never been accused of corruption and a leadership style characterized by consensus rather than authority. A greater quantity of feminine-classified profiles also prioritize gender equality and poverty issues.

Figure 4 shows the estimated AMCEs of all candidate traits which were not used to predict profiles' masculinity-femininity as well as the overall AMCE of a profile's gender. Here the AMCE of femininity includes the traits of corruption accusations, leadership styles, and issue priorities.

The figure on the left shows a positive effect of about 0.07 of femininity on the probability of a voter supporting that profile. The figure on the right shows a non-significant effect of femininity on the probability of a voter viewing the profile as viable, relatively likely to win an election.

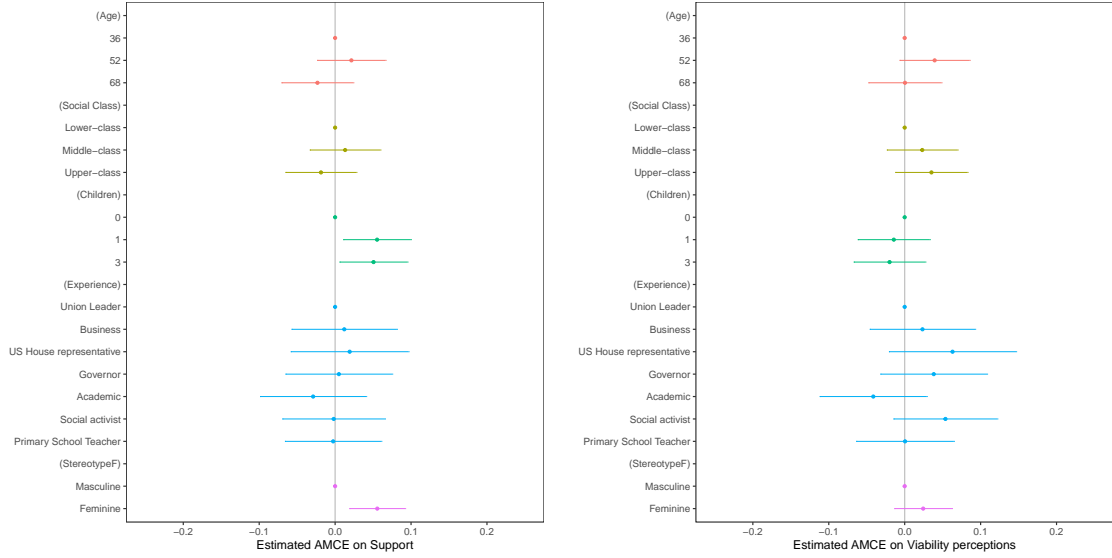


Figure 4: Effect of Stereotypes on Support and Viability perceptions (US sample)

5. Conclusions

Research estimating causal effects of candidate sex on vote choice suggests that citizens discriminate little against women in the aggregate. However, without gender stereotype measurements, we know little about why voters may prefer female or male candidates. We have argued throughout this paper that parallel conjoint designs can usefully uncover the content of high-dimensional gender stereotypes in an electoral context. They also permit analysis of the relationships between gender stereotypes and voters' complex evaluations of presidential candidates. Machine learning techniques can help identify relevant interactions, enhancing estimates of candidates' masculinity-femininity.

We also have argued, and have shown that voters tend to associate a greater number of candidate attributes with female rather than male presidential candidates. This is likely due to men's historical dominance in running for and winning presidential elections worldwide. Our AMCE results concerning these perceptions are generally consistent with the literature on gender stereotypes discussed in section 1. Notably, corruption accusations seem to serve as a key piece of information that voters use to stereotypically differentiate between female and male candidates. Corruption, as a valence issue, reduces support across the board for presidential candidates. Overall, more intense perceptions of candidates' masculinity reduces electoral support while more intense perceptions of candidates' femininity increases support.

Our design, which uncovers the high-dimensional content of gender stereotypes and

estimates their causal impact on voter evaluations, could generalize to study other kinds of stereotypes, for example those based on race, religion, or ethnicity.

References

- Acharya, Avidit, Matthew Blackwell and Maya Sen. 2018. "Analyzing Causal Mechanisms in Survey Experiments." *Political Analysis* 26(4):357–378.
- Alexander, Deborah and Kristi Andersen. 1993. "Gender as a Factor in the Attribution of Leadership Traits." *Political Research Quarterly* 46(3):527–545.
- Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26(1):112–119.
- Bansak, Kirk, Michael M. Bechtel and Yotam Margalit. 2021. "Why Austerity? The Mass Politics of a Contested Policy." *American Political Science Review* 115(2):486–505.
- Barnes, Tiffany D. and Emily Beaulieu. 2014. "Gender Stereotypes and Corruption: How Candidates Affect Perceptions of Election Fraud." *Politics & Gender* 10(03):365–391.
- Barnes, Tiffany D. and Emily Beaulieu. 2019. "Women Politicians, Institutions, and Perceptions of Corruption." *Comparative Political Studies* 52(1):134–167.
- Bauer, Nichole M. 2015. "Who stereotypes female candidates? Identifying individual differences in feminine stereotype reliance." *Politics, Groups, and Identities* 3(1):94–110.
- Bauer, Nichole M. 2020a. "A Feminine Advantage? Delineating the Effects of Feminine Trait and Feminine Issue Messages on Evaluations of Female Candidates." *Politics & Gender* 16(3):660–680. Publisher: Cambridge University Press.
- Bauer, Nichole M. 2020b. *The Qualifications Gap: Why Women Must Be Better than Men to Win Political Office*. Cambridge University Press. Google-Books-ID: 34LoDwAAQBAJ.
- Bem, Sandra L. 1974. "The measurement of psychological androgyny." *Journal of Consulting and Clinical Psychology* 42(2):155–162.
- Best, Deborah L. and Angelica R. Puzio. 2019. Gender and Culture. In *The Handbook of Culture and Psychology*. 2 ed. New York: Oxford University Press.
- Chaudoin, Stephen, Brian J. Gaines and Avital Livny. 2021. "Survey design, order effects, and causal mediation analysis." *Journal of Politics* 83(4):1851–1856.
- Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4):399–416.
- Dasgupta, Tirthankar, Natesh S Pillai and Donald B Rubin. 2015. "Causal inference from

- 2K factorial designs by using potential outcomes." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77(4):727–753.
- De La Cuesta, Brandon, Naoki Egami and Kosuke Imai. 2022. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution." *Political Analysis* 30(1):19–45.
- Dolan, Kathleen A. 2014. *When Does Gender Matter?: Women Candidates and Gender Stereotypes in American Elections*. New York: Oxford University Press. Google-Books-ID: a23rAwAAQBAJ.
- Dolan, Kathleen and Kira Sanbonmatsu. 2011. Candidate Gender and Experimental Political Science. In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. Cambridge ; New York: Cambridge University Press pp. 289–299.
- Duerst-Lahti, Georgia and Rita Mae Kelly. 1995. *Gender Power, Leadership, and Governance*. Ann Arbor: University of Michigan Press. ID: 32664978.
- Eagly, Alice H. and Steven J. Karau. 2002. "Role congruity theory of prejudice toward female leaders." *Psychological Review* 109(3):573–598.
- Eagly, Alice Hendrickson, Linda L. Carli Alice H. Eagly and Linda Lorene Carli. 2007. *Through the Labyrinth: The Truth about how Women Become Leaders*. Harvard Business Press. Google-Books-ID: b2kf_B_4f0kC.
- Egami, Naoki and Kosuke Imai. 2018. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." *Journal of the American Statistical Association* pp. 1–34.
- Franceschet, Susan, Mona Lena Krook and Jennifer M. Piscopo. 2012. *The Impact of Gender Quotas*. New York: Oxford University Press.
- Goplerud, Max, Kosuke Imai and Nicole E. Pashley. 2022. "Estimating Heterogeneous Causal Effects of High-Dimensional Treatments: Application to Conjoint Analysis.".
- Grimmer, Justin, Solomon Messing and Sean J. Westwood. 2017. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods." *Political Analysis* 25(4):413–434.
- Guryan, Jonathan and Kerwin Kofi Charles. 2013. "Taste-based or statistical discrimination: the economics of discrimination returns to its roots." *The Economic Journal* 123(572):F417–F432.

- Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014a. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1):1–30.
- Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014b. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1):1–30.
- Ham, Dae Woong, Kosuke Imai and Lucas Janson. 2022. "Using Machine Learning to Test Causal Hypotheses in Conjoint Analysis." pp. 1–37.
- Hinojosa, Magda. 2012. *Selecting women, electing women: Political representation and candidate selection in Latin America*. Temple University Press.
- Holian, David B. and Charles Prysby. 2014. "Candidate Character Traits in the 2012 Presidential Election." *Presidential Studies Quarterly* 44(3):484–505.
- 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.
- Horiuchi, Yusaku, Daniel M. Smith and Teppei Yamamoto. 2018. "Measuring Voters' Multidimensional Policy Preferences with Conjoint Analysis: Application to Japan's 2014 Election." *Political Analysis* 26(02):190–209.
- 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. ID: 4891449960.
- Inglehart, Ronald and Pippa Norris. 2003. *Rising tide: gender equality and cultural change around the world*. Cambridge, UK; New York: Cambridge University Press. ID: 50291196.
- Kahn, Kim Fridkin. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York: Columbia University Press.
- Kahneman, Daniel and Amos Tversky. 1973. "On the psychology of prediction." *Psychological review* 80(4):237.
- Kinder, Donald R., Mark D. Peters, Robert P. Abelson and Susan T. Fiske. 1980. "Presidential prototypes." *Political Behavior* 2(4):315–337.
- Lawless, Jennifer L., Fox Richard Logan. 2010. *It still takes a candidate : why women don't run for office*. New York: Cambridge University Press. ID: 467359385.

- Le Foulon, Carmen and Catherine Reyes-Housholder. 2021. "Candidate sex, corruption and vote choice." *Electoral Studies* 69(2):102270.
- Leeper, Thomas J., Sara B. Hobolt and James Tilley. 2020. "Measuring Subgroup Preferences in Conjoint Experiments." *Political Analysis* 28(2):207–221.
- Lucciola, Natalia. N.d. "The Perks of Being Female: Gender Stereotypes and Voters' Preferences in Brazil." *Politics & Gender*. Forthcoming.
- McDermott, Monika L. 2016. *Masculinity, Femininity, and American Political Behavior*. Oxford, New York: Oxford University Press.
- Montgomery, Jacob M., Brendan Nyhan and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3):760–775.
- Murray, Rainbow. 2010. *Cracking the highest glass ceiling : a global comparison of women's campaigns for executive office*. Santa Barbara, Calif.: Praeger. ID: 548555697.
- Oliver, Sarah and Meredith Conroy. 2020. *Who Runs? The Masculine Advantage in Candidate Emergence*. Ann Arbor: .
- Ratkovic, Marc and Dustin Tingley. 2017. "Sparse estimation and uncertainty with application to subgroup analysis." *Political Analysis* 25(1):1–40.
- Rosenwasser, Shirley Miller and Norma G. Dean. 1989. "Gender Role and Political Office: Effects of Perceived Masculinity/Femininity of Candidate and Political Office." *Psychology of Women Quarterly* 13(1):77–85.
- Sanbonmatsu, Kira. 2002. "Gender Stereotypes and Vote Choice." *American Journal of Political Science* 46(1):20–34.
- Sanbonmatsu, Kira. 2006. *Where Women Run: Gender and Party in the American States*. Ann Arbor: University of Michigan Press.
- Schneider, Monica C. and Angela L. Bos. 2014. "Measuring Stereotypes of Female Politicians." *Political Psychology* 35(2):245–266.
- Schwarz, Susanne and Alexander Coppock. 2022. "What Have We Learned About Gender From Candidate Choice Experiments? A Meta-analysis of 67 Factorial Survey Experiments." *Journal of Politics* 84(2):1–14.
- Sczesny, Sabine, Janine Bosak, Daniel Neff and Birgit Schyns. 2004. "Gender Stereotypes

and the Attribution of Leadership Traits: A Cross-Cultural Comparison." *Sex Roles* 51(11):631–645.

Streb, M. J., B. Burrell, B. Frederick and M. A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72(1):76–89.

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.