

A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns

Much candidate evaluation research has shown that voters care a great deal about candidates' personal characteristics, and that they often weigh personality traits more heavily than other politically relevant information like policy stances or ideology (Bittner 2011, Holian and Prysby 2014; Bartels 2002; Goren 2002; Funk 1999; Prysby 2008). Use of these personal judgments by voters is rational in that they require little information or effort and are therefore a cognitively "cheap" means of determining how a candidate will perform while in office (Holian and Prysby 2014; Shabad and Anderson 1979; Rahn, Aldrich, Borgida and Sullivan 1990). Indeed, perceptions of traits such as leadership, competence, integrity, and empathy have been shown to predict vote outcomes for presidential, House, and Senate candidates (Holian and Prysby 2014; Funk 1999; Kinder 1986; Markus 1982; Miller 1990; Miller and Miller 1976), and perceptions of a candidate's competence, in particular, are often considered to be particularly important in predicting how a candidate will fare (Holian and Prysby 2014; Fridkin and Kenney 2011; Kinder 1986; Markus 1982; Miller and Miller 1976; Shanks and Miller 1990, Mondak 1995; Funk 1999).

It is perhaps not surprising that voters care a great deal about how competent their political leaders are. What is potentially more problematic is that candidate gender seems to influence perceptions of competence. A well-established literature in political science has examined the role of gender-based stereotypes in candidate evaluation and voting behavior, and much evidence suggests that women candidates are perceived differently than their male counterparts on a number of fronts, including that women candidates are presumed to be less competent than men (e.g. Huddy and Terkildsen 1993; Sigelman, Sigelman, Walkosz and Nitz 1995). Because of the seeming importance of competence judgments in voters' decision-making calculus, this particular stereotype may put female candidates at a major disadvantage.

At the same time, the role of all gender-based stereotypes in actual elections is unclear. Many studies have found that overt prejudice is no longer an obstacle for female candidates, and even that

“when women run, women win” (Burrell 1994, Seltzer, Newman and Leighton 1997, Darcy, et al 1997, Woods 2000, Dolan 2004). Further, several recent studies have called into question whether gender-based stereotypes are applied to women candidates as readily as they once were (Dolan 2014; Brooks 2013; Hayes 2011).

While evidence for the role of gender-based stereotypes in voting behavior is not as straightforward as it may have once seemed, there is reason to believe that stereotypic evaluations still occur, and that they may have a more nuanced role in affecting vote choice and candidate evaluations than previously thought. For example, in a Dynamic Process Tracing (DTPE) experiment, Ditonto, Hamilton and Redlawsk (2014) find that, rather than directly influencing vote choice, stereotypic evaluations influence the types of information that subjects search for related to female candidates. Specifically, and among other things, they find that subjects seek out more information related to a candidate’s competence and experience when they are presented with a female candidate than when they see a male candidate. The results of their study suggest that voters may weight competence judgments more heavily when they are considering whether or not to cast a vote for a woman than when they are considering a man.

While Ditonto, et al (2014) find that subjects seek out an inordinate amount of competence-related information for women candidates, their analysis does not examine how the nature of the information they encounter influences subsequent evaluations and vote choice. Social psychological research on stereotypes and information processing suggests that subjects may react differently to competence-related information when it is applied to female candidates than when it is applied to male candidates due to the stereotypes subjects may hold. Because of assumptions that women are less competent, stereotype-confirming information that casts doubt on a candidate’s competence may be more damaging for women than for men. In other words, subjects may be more likely to believe that a bad debate performance or negative newspaper editorial is evidence that a candidate is incompetent when that candidate is a woman, while they may be more willing to give male candidates the benefit of the doubt. At the same time, though, a female candidate who does particularly well in a debate or receives a ringing

endorsement from a newspaper may never appear to be a victim of negative stereotyping because the positive nature of the substantive information about her makes competence stereotypes less salient.

This paper seeks to determine whether subjects react differently to competence-related information when it is applied to a woman candidate vs. when it is applied to a man. I use two DPTE-based experiments¹ that place subjects in a simulated campaign environment in which they can search for a large amount of information about the candidates running for office. In both experiments, subjects are asked to take part in a simulated presidential campaign and election in which the candidate running in their party of choice could be either a man or a woman. Among the information available about the candidates, subjects can access a number of competence-related information items, such as the results of a debate and comments from a former staff member. These items either confirm the in-party candidate's competence or cast doubt on his/her political abilities, based upon the experimental condition the subject is assigned to.

Importantly though, candidate gender is not the only characteristic that has been shown to cue competence or incompetence in voter assessments, nor is textual, substantive information the only type of “information” that voters utilize. An emerging line of study has found that individuals make quick judgments about others' personality traits based on nothing more than a brief exposure to their faces (e.g. Hall et al 2009; Hassin and Trope 2000). This observation has been applied to political candidate evaluation in a series of studies whose findings indicate that the results of actual elections can, partially, be explained by spontaneous judgments about how competent the candidates look (see Todorov, Olivola, Dotsch, and Mende-Siedleke 2015 for an overview of this literature).

Because gender is at least partially an appearance-based heuristic (Lau and Redlawsk 2001), it may “compete” or interact with other aspects of a candidate's facial appearance to influence inferences of competence. Further, and based on the literature mentioned above, this type of information may be just as influential on voters' decision-making as substantive information. Because of this, Experiment 2 also

¹ With funding provided by the National Science Foundation.

systematically varies whether a candidate has a face that has been judged competent- or incompetent-looking in order to determine whether a particularly competent or incompetent appearance matters differently for female candidates, and whether facial appearance can change the dynamic between candidate gender and the importance of substantive competence-related information.

In sum, I find that substantive competence-related information is far more important for subjects who are asked to evaluate female candidates than for those who are presented with male candidates. Across both experiments, women candidates who are paired with substantive information that casts doubt on their competence receive significantly lower feeling thermometer scores and, in Experiment 2, they are less likely to receive a subject's vote than women in the competent information condition. Women paired with competent information, however, do just as well or better than men. This suggests that, even if gender-based stereotypes are not always directly affecting vote choice, they may be playing a subtler role in the processes of information acquisition and opinion formation, which do in turn influence vote choice. My findings related to facial competence are less clear. When considering the entire sample, an incompetent appearance does not appear to have much of an effect on candidate evaluations or vote choice for male or female candidates. I do, however, find limited evidence that facial incompetence may be a problem for female candidates who run as Republicans, but not Democrats.

Gender Stereotypes and Candidate Evaluation

Much political science evidence suggests that women candidates are subject to stereotypes based on personality traits, issue competencies, and ideology, among other things. Female candidates are often assumed to have more feminine and communal characteristics—they are seen as more compassionate, gentle, warm, cautious, and emotional, for example (Huddy and Terkildsen 1993; Kahn 1996; Leeper 1991). Also, they are often seen as more trustworthy and honest than male candidates (Kahn 1996). At the same time, they are stereotyped as less agentic—less competent, less able to handle the emotional demands of high office, and lacking in masculine traits like “toughness” (Huddy and Terkildsen 1993; Carroll and Dittmar 2010). Likely because of these assumptions about women's personality traits, voters also often assume that women have different areas of policy expertise than men, with particular

proficiency in “compassion issues” like education, healthcare, poverty, and child care often attributed to women candidates. At the same time, more “masculine” issues like crime, the military, and the economy are seen as the arena of male politicians (Alexander and Andersen 1993; Cook, Thomas and Wilcox 1994; Dolan 2004; Rosenwasser and Seale 1988, Leeper 1991; McDermott 1998).

Importantly though, some recent work has cast doubt on the influence of gender-based stereotypes on election outcomes at all (Dolan 2014; Brooks 2013; Hayes 2011), suggesting that more expressly political variables like party affiliation are much more influential than any effects of gender. This may suggest that gender-based evaluations are less salient in politics than they once were, or it may be a sign that gender is influential only in certain contexts or in subtle, more indirect ways. Indeed, recent research suggests that gender stereotypes may not automatically be applied to women candidates in all election scenarios, but can be influential in certain circumstances, such as when stereotypes are activated by campaign messages (Bauer 2015a), when considering certain types of voters (Bauer 2015b), in times of terrorist and/or national security threat (Dolan 2004, Lawless 2004; Holman, Merolla and Zechmeister 2011), when cognitive demand is low (Higgle, Miller, Shields and Johnson 2008), and in low-information elections (Matson and Fine 2006). A female candidate’s political party can also influence whether and how stereotypes are applied (Sanbonmatsu and Dolan 2009; Plutzer and Zipp 1996; Schneider and Bos 2011; Huddy and Capelos 2002; Hayes 2005, 2011).

There is also a fair amount of evidence that stereotypes related to competence, per se, may continue to be a problem for women candidates, even when other kinds of stereotypes may not be. Schneider and Bos (2014), for example, find that women politicians actually comprise a *subtype* of women, and are therefore subject to stereotypes based on that subtype, rather than those applied to women more generally. In particular, women candidates are seen as deficient in masculine qualities associated with leadership, such as competence. This coincides with findings from psychological studies of gender and leadership that find a widespread perception that female gender roles are fundamentally incompatible with leadership roles (Eagly and Karau 2002) and that women are often assumed to lack agentic leadership-relevant traits such as competence (see reviews by Carli and Eagly 1999 and Foschi 2000).

Finally, Ditonto, et al (2014) find that subjects search for more information related to competence for female candidates than for male candidates, essentially “checking up” on female candidates’ qualifications and credentials.

Stereotypes and Information Processing

The conflicting nature of previous findings on gender stereotypes and voting behavior may make more sense if one takes into account the process of learning that must take place over the course of a campaign. Applying an information processing approach to the study of vote choice, Lau and Redlawsk (2006) posit that deciding whom to vote for is a lengthy process with the vote choice itself as only the last in a series of steps that voters must undergo. In this model, the information that voters acquire along the way to making a vote choice is a crucial intervening variable in determining how various subject- or candidate-level variables (like gender cues and stereotypes) influence candidate evaluation and vote choice. Rather than influencing evaluations and vote choice directly, the competence-based stereotypes that may be applied to women in leadership positions may cause voters to seek out different information for male and female political candidates (as Ditonto, et al find in their 2014 piece), to care more about certain kinds of information when it relates to female candidates (as Ditonto, et al also find), and/or to react differently to certain information when it relates to a female candidate instead of a male candidate (as I seek to show in this study).

One way to think about stereotypes—based on gender or anything else—is as an information shortcut, which helps individuals to make decisions in the absence of other kinds of information (e.g. Lau and Redlawsk 2001). As more individuating information about a person or candidate becomes available, people may rely less on stereotypic information (Locksley, Borgida, Brekke, and Hepburn 1980; Borgida, Locksley and Brekke 1981; Heilman, 1984; Locksley, Hepburn and Ortiz 1982; Ashmore 1981; Eagly and Wood 1982). Political campaigns—especially those at the national level—often provide voters with quite a lot of information about the candidates in the race, It may be, then, that stereotypes simply matter little in high-profile elections where much individuating information is accessible.

However, much evidence also suggests that stereotypes are not necessarily automatically applied when one encounters a member of a stereotyped group (Devine 1989; Blair and Banaji 1996; Blair 2002), but rather, that they need to be *activated* by stereotype-consistent information before they will be applied to an individual (Lepore and Brown 1997; Sinclair and Kunda 1999; Kunda and Spencer 2003). This seems to be true of women candidates—not only have recent studies found no effects of baseline gender stereotypes on candidate evaluations and/or vote choice (e.g. Brooks 2014; Dolan 2014), but Bauer (2015a) finds that gender-based stereotypes are not applied to women candidates at all *unless* they are activated through stereotypic campaign messages. The extent to which competence-based stereotypes influence ultimate evaluations of women candidates and vote choice, then, likely depends on the nature of the information that is encountered. Stereotypes that women candidates are less competent than men may or may not directly affect a particular candidate depending upon whether voters are presented with information that activates stereotypes about women’s competence (or lack thereof), such as a negative debate performance or critique on her competence or qualifications from a newspaper editorial. This stereotype activation could lead to less positive evaluations and a lower likelihood of electoral success. However, substantive information provided by the media and/or a campaign may well present a picture of competence, in which case female candidates may not be at a disadvantage, after all.

Appearance-Based Competence Judgments

At the same time, psychologists have also found that individuals frequently make spontaneous trait inferences based on others’ facial characteristics (see Todorov, et al 2015 for an overview of this topic). These inferences are made very rapidly—possibly in as little as 34 milliseconds (Todorov and Uleman 2003; Olson and Marshuetz 2005; Todorov 2008) and a growing body of evidence has found that subjects’ inferences about a candidate’s traits—and competence, in particular—can correctly predict the outcomes of actual elections across different countries and levels of office (Todorov, Mandisotza, Goren and Hall 2005; Hall et al 2009; Ballew and Todorov 2007; Olivola and Todorov 2010a; Mattes, Spezio, Kim, Todorov, Adolphs, and Alvarez 2010; Laustsen 2014; Lenz and Lawson 2011; Sussman, et al 2013).

Though the exact features a face must possess in order to appear competent are still under investigation, some evidence suggests that such inferences seem to depend on aspects of an individual's face that signal maturity, such as distance between the eyes, roundness of the face and angularity of the jaw (Oliviola and Todorov 2010; Herrick, et al. 2012). Further, several studies have found that a candidate's gender and the gender typicality of a candidate's facial features can affect competence judgments and electoral success. Herrick et al (2012), for example, find that women's faces are systematically rated as less mature and less competent in both real and simulated elections. Also, the features that seem to cue competence—eyes set far apart, a prominent jawline, a generally mature appearance—are also correlated with a masculine appearance (Oliviola and Todorov 2010; Herrick, et al 2010). On the other hand, Hehman, Carpinella, Johnson, Leitner, and Freeman (2014) find that more masculine looking women (who may also be more traditionally competent looking) are actually at a disadvantage among voters. Carpinella and Johnson (2013a) further find that facial femininity leads to higher competence ratings for liberals and Democrats but to lower competence ratings for conservatives and Republicans, but there is also some evidence that gender-typical features (feminine features for women and masculine features for men) give Republican candidates an advantage, but not Democrats (Carpinella, Hehman, Freeman and Johnson 2015; Carpinella and Johnson 2013b).

Taken together, these findings suggest that the competence cues provided by a candidate's facial features may also affect women differently from men. This type of information is fundamentally different from substantive information provided by the media, political parties, and candidates themselves. While substantive, textual information is generally incorporated into judgments via deliberate, "System 2" processing, the information provided by a candidate's facial features likely influences evaluations via automatic, "System 1" processing, possibly without a voter's conscious awareness (see Lodge and Taber 2013 for a comprehensive overview of this literature). By manipulating the facial features thought to signify competence, then, along with candidate gender (in Study 2), this paper seeks to determine to what extent this sort of System 1, rapidly-processed competence information affects ultimate evaluations of women candidates even when substantive, politically relevant competence information is also available.

Hypotheses

Even though stereotypes that women in leadership positions are less competent than men certainly exist, substantive information that portrays a female candidate as competent should forestall any such stereotypes before they are activated and applied to a particular candidate. Therefore, and given the importance of candidate competence in general, I expect subjects in the “competent information” condition to like their in-party candidates more and be more likely to vote for them than those subjects in the “incompetent information condition,” regardless of candidate gender.

At the same time, the type of substantive, competence-related information available about the candidates should be particularly important for female candidates. Women candidates should be particularly vulnerable to substantive information that portrays them as incompetent, since it may activate competence-related stereotypes. For this reason, I expect that female candidates who are in the incompetent information condition will receive lower feeling thermometer ratings and have a lower likelihood of receiving a subject’s vote than will male candidates in the incompetent information condition.

Further, I expect that an incompetent facial appearance may also be able to activate competence-based stereotypes for women candidates. Therefore, women candidates who also have an “incompetent” facial appearance will do worse in terms of evaluation and vote choice than women who “look competent” or men in either appearance-based competence condition, regardless of the information available about them.

Method

The experiments in this paper utilize the Dynamic Process Tracing Environment (DPTE)² in order to expose research participants to simulated election campaigns. DPTE has been described in detail elsewhere (Lau and Redlawsk, 2001; 2006; Redlawsk & Lau, 2013) so I will be brief here. Process tracing presents subjects with an abundance of information related to a number of alternatives (in this

² See www.processtracing.org. A screenshot of a DPTE campaign and an open information box are provided in the Supporting Materials.

case, realistic yet fictitious candidates) and allows researchers to follow a subject's choices as s/he examines whatever information s/he chooses. These experiments are "dynamic" because participants face an ever-changing information environment that mimics the ebb and flow of a real-world election campaign.

DPTE campaigns include the kinds of information that voters might expect to learn about candidates and assign sets of these attributes to a series of invented candidates. These attributes often include personal traits, background information, a wide range of issue positions, polls, and endorsements. During a DPTE campaign, individual pieces of information scroll down the subject's computer screen, each remaining available for a period of time (in both of these experiments, 12 seconds). As one piece of information (an "information box") moves off of the bottom of the screen, it is replaced by a new piece of information at the top of the screen. Each scrolling information box contains a brief synopsis of the information provided inside the box (e.g. Patrick Turner's stance on Education), as well as a small picture of the candidate to whom the information refers, and a colored border corresponding to the party of the candidate to whom the information refers (red for Republicans, blue for Democrats). When subjects want to access the information available in the information box, they click on it, at which point the box expands to fill up the entire screen, and allows subjects to read the information available inside. While subjects read the information in a particular box, the other pieces of information continue to scroll behind it, so they must choose which information is most important for them to learn. Subjects can choose to learn as much, or as little, information as they wish. Because the candidates were invented, participants knew nothing about them at the beginning; anything they learned must have come from the information available to them in the campaign.

Experiment 1

The first experiment was conducted in the Spring and Summer of 2012 and was completed by 449 total subjects. 106 of those subjects were recruited from the Central New Jersey area and took the experiment in the lab at the Center for the Experimental Study of Politics and Psychology at Rutgers University. The remaining subjects were recruited through Amazon Mechanical Turk and took the study

online using their own computers.³ The sample is 58% female, 76% white, 9% African American, and 5% Latino. The median age is 32, 16% of the sample identified as Republican, 48% as Democrat, and 31% as independent.

Subjects were asked to experience a simulated campaign for a presidential election, answer questions about the candidates they saw, and cast a vote for their candidate of choice. The entire experiment lasted approximately one hour and consisted of both a primary and general election campaign. Manipulations differed from the primary to the general election portions of the study, and this paper focuses on manipulations presented in the general election, therefore I will not discuss the primary election here.⁴ Before beginning the campaign, subjects first saw a series of instruction screens, then were asked to complete a pre-election questionnaire, which collected information on subjects' issue positions, ideology, party identification, demographics, political participation, political knowledge, ratings of interest groups, and a number of psychological scales. They then participated in a 2-minute "practice campaign" in order to gain some experience using the software, after which they viewed an 11-minute primary campaign.⁵ They then moved on to the general election campaign, which lasted approximately 10 minutes and featured one candidate from each political party. After the general campaign, voters were asked to cast a vote for one of the two candidates and then to answer a series of questions about each of them.

Subjects could access information about the candidates' stances on a number of policy issues, their ideology, family, education, and prior experience in work and politics. Each piece of information appeared twice during the campaign and the software randomly decided the order in which the boxes

³ Very few differences were found between the two sub-samples or in their performance in the study. Details can be provided upon request.

⁴ For a discussion of the Primary Election, see the Online Appendix A.

⁵ Subjects were asked to choose which party's primary election they'd like to vote in and their "in-party candidate" is considered to be the candidate that ran in the party whose primary they chose. 4 Democrats in the study chose to vote in the Republican primary, while all self-identified Republicans voted in the Republican primary. Among independents, 3 Republican leaners voted in the Democratic primary and 6 Democratic leaners voted in the Republican primary. Among pure independents, 21 voted in the Democratic primary and 18 voted in the Republican primary. I control for this in my analyses by including a "primary partisan match" variable, which is explained in detail below.

were seen by each subject. In the general election, there were 35 unique items available about each candidate. Each piece of information was available twice, meaning that there were 140 pieces of information available, in total.

Subjects experienced a 2 (gender) x 2 (competent/incompetent information) manipulation, in which their in-party candidate (i.e. the candidate from their preferred political party) could either be male or female and portrayed as competent or incompetent in the substantive information available about him or her.⁶ Gender was cued via the candidate's name (Patrick vs. Patricia Turner, e.g.) and the picture associated with the candidate.⁷ The competence of information manipulation either portrayed the candidate as very politically competent or considerably less competent via 6 information items about the in-party candidate—a summary of the in-party candidate's debate performance, comments about the candidate from a newspaper editorial, a summary of the candidate's job performance in previous office, a description of the candidate by a political opponent, a summary of the candidate's past political experience, and a description of the candidate by a former staff member (see Online Appendix A for the exact text of these items). In the “competent information” condition, these items all portrayed the candidate as very experienced in politics and someone others viewed as a competent politician. In the “incompetent information” condition, these items cast more doubt on the candidate's prior experience and potential as a presidential contender (though they were still realistic depictions of the types of information

⁶ This study also included a race manipulation in which in-party candidates could also either be black or white. Because this manipulation was randomly assigned and did not affect the nature of the results for candidate gender, I leave it out of the analyses that follow.

⁷ Pictures were taken from the websites of various state legislatures and presented headshots of the candidates against a blank, single-colored background. See the supporting material to see images. Importantly, all in-party candidates in the general election were also “incompetent-looking.” That is, the images associated with the in-party candidate scored low in “competence” ratings in a pretest of candidate images. Pretest details are also available in Online Appendix A. The out-party candidate was always a competent-looking, white man from the out-party's primary. This is the result of a manipulation in the primary election portion of the study in which both a competent-looking and an incompetent-looking candidate ran for the party's nomination. In order to answer unrelated questions, the incompetent-looking in-party candidate always “won” and advanced to the general election. While this is a confound, because all in-party candidate images (both men and women) in the general election were incompetent-looking, an incompetent appearance is constant across all groups. I can therefore still compare female candidates to male candidates.

that might be available about a presidential candidate). The debate performance summary was a “forced” information item that appeared and opened on the screen at the same point during the campaign, and all subjects were forced to close it before moving on. This ensured that all subjects were exposed to the manipulation.

Experiment 2

Experiment 2 was conducted in December of 2013 and was completed by 377 total subjects, recruited from Amazon Mechanical Turk.⁸ The sample was evenly split by gender, was 83% white, 6.5% black, and 7.5% Latino, and the median age was 35 years old. 64% identified as Democrats (or Democratic leaners), 21% identified as Republican (also including leaners), and 15% identified as independent.

The design of Experiment 2 was similar to that of Experiment 1, but differed in several important ways. First, it consisted of a general election only, which means that the total time subjects spent taking the experiment was shorter than in Experiment 1—approximately 30 minutes, on average. Second, this study adds a manipulation of whether the candidate had a face that appears competent or incompetent. To that end, the pictures of the candidates in Experiment 2 were created using FaceGen, which is a program that allows users to create human faces with varying physical characteristics (see the Appendix for a more detailed explanation and to see the images). By using FaceGen, I was able to manipulate the physical features thought to signify competence/incompetence, while controlling for other aspects of the candidates’ appearance that inherently vary in pictures of real people. This allows me to have more

⁸ See Clifford, Jewel and Waggoner 2015; Berinsky, et al 2012; Weinberg et al, 2014; Buhrmester, et al; Paolacci, et al; Crump et al 2013 for analyses of how MTurk samples compare to other types of internet and in-person samples. Evidence suggests that findings from MTurk studies do not differ in important ways from those conducted on other kinds of samples. Concerns about MTurk sample demographics center around the fact that MTurkers tend to be more liberal than nationally representative samples (Berinsky, et al 2012; Huff and Tingley 2015), which can pose a problem for certain kinds of studies. In the case of gender stereotypes, a more liberal sample presents a tougher case than one that is more conservative, as conservatives are more likely to hold traditional views on gender, and Republican women tend to fare worse than Democrats (King and Matland 2003; Dolan 2010).

confidence that any differences in evaluations of the candidates are due to my manipulations and not other factors that are unique to pictures of actual candidates, such as hairstyle, make-up, etc.

Multiple male and female “candidates” were created, varying the aspects of physical appearance that are thought to signal competence, including eye spacing, jaw width, and maturity/“baby-facedness” (Oliviola and Todorov 2010). These images were then pretested, and those with the highest and lowest competence ratings were chosen for the experiment. While verging on realistic, the images that FaceGen creates do not look entirely “real,” so subjects were informed that they were seeing avatars of real candidates whose identities were being protected.

Aside from these differences, Experiment 2 largely mimicked Experiment 1. Subjects saw a 2 x 2 x 2 design in which candidates could either be male or female, competent- or incompetent-looking, and portrayed as either competent or incompetent in the information available about them. The out-party candidate, meanwhile, was always a competent-looking, white man who was also portrayed as competent in the information presented about him. The procedure was the same—subjects began by taking the same pre-election questionnaire that was used in Experiment 1, seeing a practice campaign, then experiencing a 10-minute general election campaign, after which they were asked to “vote” for the candidate of their choice and answer several questions about the candidates they saw.⁹ 40 items were presented (twice, in random order) about each of the candidates and, as in Experiment 1, there were 6 competence-related pieces of information that subjects could access, with one “forced” item ensuring that every subject was exposed to the manipulation. Table 1 lays out the differences between the two experiments, both in terms of sample characteristics and elements of the design.

[Table 1 about here]

⁹ While subjects in Experiment 1 were asked to choose which primary election they wished to vote in, which allowed true independents to choose which candidate would become their “in-party,” true independents in Experiment 2 (those who do not lean toward Democrats or Republicans) were randomly assigned to either the “Democrat” condition or the “Republican” condition. There were 55 independents to whom this applied, half of whom experienced manipulations to the Democratic candidate, half of whom experienced manipulations to the Republican candidate. I include a measure of the strength of a subject’s party ID in order to control for this.

Results¹⁰

In both studies, my dependent variables of interest are the subjects' feeling thermometer difference scores (the difference between the feeling thermometer scores for the subject's in-party and out-party candidates: in-party – out-party), as well as whether the subject voted for his or her in-party candidate. The feeling thermometer scores are measured on the traditional 0-100 point scale, though because I take the difference between two scores, values can range from -100 to 100. Vote choice is measured using a dichotomous variable for which a score of 1 indicates a vote for the in-party candidate and a score of 0 indicates a vote for the out-party candidate (subjects were forced to make a choice between the two).

I use Analysis of Variance (ANOVA) in order to compare mean feeling thermometer difference scores for subjects in each gender and competence condition (as well as interactions between them). For the vote choice, I use a series of two logistic regressions. In each study, Model 1 includes the primary manipulations (gender and competence of information in both studies, and competence of appearance in Study 2). Model 2 adds 2-way interactions between seeing a female candidate and being in the incompetent information condition.¹¹ Experiment 2 also includes 2-way interactions between seeing a female candidate and being in the incompetent appearance condition, as well as an interaction that indicates that a subject was in both the incompetent information and incompetent appearance condition (the three-way interaction between candidate gender, competence of appearance, and competence of information was never statistically significant).

In all analyses in Experiment 1, I also control for whether the subject chose the primary election that matched his or her partisan ID (Primary Partisan Match). Subjects who identified as a strong, weak or

¹⁰ Data and syntax used for these analyses are available on the *Dataverse* site for *Political Behavior* (<https://dataverse.harvard.edu/dataverse/polbehavior>).

¹¹ I also constructed models incorporating a number of covariates typical to vote choice models, including party ID, ideology, demographics and political sophistication, as well as controls for the amount of information accessed (both total information accessed and the number of competence-related items accessed) by each subject, and found that none of these influenced the pattern of results. I present these simplified models for parsimony. Full models can be provided upon request.

“leaner” and voted in their in-party’s primary were given a “1” on this measure. True independents and subjects who chose to vote in their out-party’s primary were given a “0.” In Experiment 2, I include a “folded” strength of party ID measure to control for the fact that true independents were randomly assigned to either the Democratic or Republican set of conditions. This measure varies from 0-4, where 0=true independent and 4=strong partisans of either party.

Table 2 shows the results of a series ANOVAs in which the dependent variable is the difference score between the subject’s in-party and out-party candidate. In both experiments, the competence of information manipulation creates statistically significant differences ($F=7.673$, $p<.01$ in Experiment 1 and 13.586 , $p<.001$ in Study 2). In Experiment 1, the mean feeling thermometer difference score between a subject’s in-party and out-party candidate in the competent information condition was 21.848 (i.e. subjects liked their in-party candidate an average of almost 22 points more than the out-party candidate). In the incompetent information condition, however, that difference fell to 14.194 points. In Experiment 2, those in the competent condition had a mean difference score of 12.099, while those in the incompetent condition actually had a negative difference score of -1.558. Not surprisingly, subjects preferred candidates who were associated with information that makes them seem like competent politicians, all else equal.

[Table 2 about here]

Interestingly, subjects also preferred male candidates, all else equal, in Experiment 1 but not in Experiment 2 ($F=13.231$, $p<.001$ in Experiment 1, $F=.390$, n.s. in Experiment 2). Subjects with a female in-party candidate had an average difference score of 12.992, while those with a male in-party candidate had a mean difference of 23.050. Given the mixed evidence on the effects of gender on candidate evaluation and recent evidence suggesting that women candidates are not necessarily at a disadvantage, this was somewhat unexpected. It is also surprising that female candidates did worse in Experiment 1 but not in Experiment 2. I discuss possible reasons for this inconsistency below.

It is also worth noting that, in Experiment 2, the appearance-based competence manipulation had no effect on mean feeling thermometer difference scores. Subjects’ affect toward their in-party candidates

seem to be unaffected by whether or not the candidate has a face that is considered to be competent/incompetent, all else equal. This is also a surprise, given the large literature that has found appearance-based trait judgments to be so influential in election outcomes. Again, possible reasons for this are discussed below.

[Figure 1 about here]

Adding the 2-way interaction between candidate gender and competence of information begins to tell a more nuanced story. In both studies, this interaction is statistically significant and affects average difference scores in the same way ($F=4.019$, $p<.05$ in Experiment 1 and $F=8.056$, $p<.05$ in Experiment 2). Figure 1 displays the mean scores for Experiment 1 and Figure 2 does the same for Experiment 2. In Experiment 1, subjects whose in-party candidate was a man that was paired with either competent or incompetent information and those who saw a competently-portrayed woman had statistically indistinguishable difference scores (24.233 for competent men, 16.622 for competent women, and 21.88 for incompetent men). Subjects who saw women paired with incompetent information, however, had an average difference score of only 6.224. This suggests that subjects liked their in-party candidate far less, relative to the out-party candidate, when she was a woman portrayed as incompetent. Interestingly, and perhaps even more surprising, men paired with incompetent information saw no disadvantage compared to either competent men or competent women.

[Figure 2 about here]

Figure 2 shows a similar pattern, but with some important differences. First, in Experiment 2, difference scores were smaller across the board, meaning subjects in this experiment rated in-party and out-party candidates more closely than did subjects in Experiment 1. This is likely a byproduct of the fact that subjects in Experiment 1 experienced a primary in their own party before participating in the general election, while those in Experiment 2 did not. Because two candidates ran in each party's primary in Experiment 1, there was some ideological variation between those candidates (see Online Appendix A; each candidate in the primaries was randomly assigned to be either moderate or more obviously liberal/conservative) and either a more moderate or more ideologically distinct out-party candidate could

have made it to the general election. In Experiment 2, on the other hand, subjects saw only general election candidates in each party that were less ideologically distinct. Half of the subjects in Experiment 1, then, likely perceived a larger ideological gap between their in-party and out-party candidates.

Second, while women candidates paired with incompetent information do the worst of all candidates in both studies, the only statistically significant difference in Experiment 2 is between women paired with incompetent information and women paired with competent information (-7.026 vs. 16.563). It is also striking that the difference score for women in the incompetent condition is actually *negative*, meaning that subjects in this condition actually liked the out-party candidate *more* than their in-party candidate, on average. Subjects who saw a female candidate who was portrayed as competent, however, liked her far more than their out-party candidate. Subjects who saw a male candidate, on the other hand, were not affected by the competence-related information, as their difference scores were positive and statistically identical in both conditions (6.959 vs. 4.401). This finding provides further evidence that women are disproportionately hurt by information that portrays them as incompetent. Interestingly, it also suggests that women who are portrayed as particularly competent may exceed expectations and benefit disproportionately, as well. This may be due to expectancy violation (e.g. Bettencourt, Dill, Greathouse, Charlton, and Mulholland 1996). In other words, competent women candidates may violate expectations that female leaders are incompetent, leading to a positive “backlash.” It is also possible that voters prefer to support female candidates, all else equal, so in lieu of evidence that a female candidate is incompetent, subjects are more enthusiastic about women candidates than men. Whatever the reason, this finding requires further research.

I turn now to an analysis of subjects’ vote choice (Tables 3 and 4). For each study, I conduct two logistic regression models predicting a vote for the subject’s in-party candidate. Model 1 shows the results of this analysis including the control variables and the main effects of the primary manipulations, only, while Model 2 adds the appropriate interaction terms. In Experiment 1, as in the candidate evaluation analyses, subjects who see a female candidate in their in-party are less likely to vote for their in-party candidate than those who see a male candidate ($b=-.451, p<.05$) and, again, subjects who are exposed to

the incompetent information condition are also less likely to vote for their in-party candidate than those exposed to the competent information condition ($b = -.591, p < .01$). This again shows that the nature of the information available in the campaign makes a big difference across the board.

[Table 3 about here]

Model 2 adds the interaction between candidate gender and the competence of information condition. Unlike in the analysis of feeling thermometer difference scores, the interaction does not have a statistically significant effect on vote choice, though the coefficient has the same sign ($B = -.319, n.s.$). This suggests that, though female candidates who are portrayed as incompetent are at a disadvantage when they are evaluated, this may not have an effect on ultimate vote choice, all else equal.

[Table 4 about here]

Table 4 shows the results of the logistic regression models predicting a subject's vote for their in-party candidate in Experiment 2. Without interaction terms, the only significant predictor of interest is incompetence of information, which has a negative effect on the likelihood of vote choice ($b = -.606, p < .01$). When the interactions are added to the model, we see the same pattern as in the analysis of the feeling thermometer scores. Unlike in Experiment 1, the interaction between incompetence of information and seeing a female candidate is negative and significant ($b = -1.302, p < .01$) here, while the marginal effect of seeing a female candidate who is competent is positive and significant ($b = .892, p < .05$). This provides further evidence that women are subject to an exaggerated effect from the competence-related information in both directions. The interaction between seeing a female candidate and a candidate with an incompetent appearance is negative and approaches significance here ($b = -.796, p < .1$), and this is the only time we see any effect of the competence of appearance manipulation on the dependent variables.

Interestingly, when separate analyses are conducted for Republicans and Democrats, there is some evidence that it is Republicans driving this trend. For both feeling thermometer scores and vote choice, Republican respondents seem to be just as supportive of incompetent-looking men as competent-looking men and women, but are less supportive of incompetent-looking women (mean feeling thermometer difference score for incompetent women is -7.294 , compared to 18.412 for competent

women, 14.016 for competent men, and 20.460 for incompetent men, $p < .1$, while the coefficient for vote choice is $-.3.278$ for an incompetent-looking female candidate, $p < .01$). This pattern does not emerge for Democrats. While this suggests that female candidates who look incompetent may be evaluated less positively and also be somewhat less likely to receive a subject's vote than other kinds of candidates, particularly if she runs as a Republican, I am hesitant to put too much stock in this finding since there were only 80 Republicans in the sample for Experiment 2. This area is certainly worthy of further study with a larger Republican sample.

Figure 3 shows the predicted probability of an in-party vote for subjects in particular combinations of gender and competence of information conditions. Interestingly, subjects who see a male candidate in their in-party are just as likely to vote for him whether he is portrayed as competent or incompetent (they have about a 56% chance of doing so)¹². Meanwhile, women in the competent condition have a 67% chance of receiving a subject's vote while female in-party candidates in the incompetent condition have only a 38% chance of receiving a subject's vote. This means that subjects who are presented with an incompetent female candidate in their own party are actually more likely to vote for the candidate in the *other* party. The results of this study suggest that the disproportionate effects of substantive portrayals of competence on female candidates go beyond just affect, but actually may influence their electoral fortunes, as well.

[Figure 3 about here]

Discussion and Conclusion

Clearly, the relationship between gender-based stereotypes and substantive information in political campaigns is a complex one. One pattern of results stands out across both experiments, however. Female candidates who are portrayed as incompetent in the information available about them are at a

¹² The likelihood of an in-party vote is lower than might be expected in a general election. This is likely due to the fact that independents are included in these analyses, as well as the fact that my manipulations were designed specifically to discourage in-party votes.

disadvantage in a way that male candidates paired with the same information simply are not. It is striking that subjects in the female/incompetent information condition seemed to like their out-party candidates as much as, or more than, their in-party candidate, and that they were *more likely* to vote for the out-party candidate in this condition in Experiment 2. The combination of a female candidate whose competence has been cast into doubt is such a potent combination of cues, that it can even trump subjects' party identification. At the same time, women in the competent information condition fared much better. In fact, women in Experiment 2 who were paired with competent information did just as well as male candidates who were paired with the same competent information. This suggests that, even for female candidates who may be subject to stereotypes related to competence, the actual information available about them matters a great deal. This is heartening, in that gender-based stereotypes related to competence seem not to be insurmountable for female candidates, as long as they are in fact qualified candidates. Perhaps we should be more surprised that in-party male candidates who are portrayed as politically incompetent are still so well-liked and supported by so many subjects!

There are also some important differences in the results from the two studies that are worth discussing. First, the main effects of candidate gender are significant in Experiment 1 but not in Experiment 2. Female candidates in Experiment 1 were at a disadvantage in terms of both feeling thermometer scores and vote choice, all else equal. It is difficult to explain this discrepancy, but it may again have to do with the fact that Experiment 1 participants experienced a primary election prior to the general and so something learned in the primary may have affected general election results. Another possibility is that it has to do with the particular image that was paired with the in-party female candidate in Experiment 1. Images of male and female in-party candidates were rated approximately equally on competence, likability and attractiveness, but it is possible that a different dimension of appearance was overlooked leading to differences in evaluation (and this is precisely why I chose to use FaceGen-generated images in Experiment 2). It is also possible that, since all of the pictures used in Experiment 1 were rated as relatively "incompetent-looking" (see FN 7), that this is a result of subjects reacting differently to an incompetent-looking woman vs. an incompetent-looking man. Because I find limited

evidence that an incompetent appearance may affect men and women differently in Experiment 2, it may certainly be part of the explanation for this finding in Experiment 1 and for the discrepancy between the two studies, as subjects saw both competent- and incompetent-looking women in the second study.

The other important difference in results is that the interaction between candidate gender and informational competence is a significant predictor of vote choice in Experiment 2 but not in Experiment 1 (while this interaction is a significant predictor of feeling thermometer difference scores in both studies). Again, this may be a result of the greater ideological distinction between in-party and out-party candidates for some subjects in Experiment 1 (thanks to the primary election that preceded it). Subjects may like their in-party candidates less when she is an incompetent-seeming woman, but her issue stances (and the issue stances of the out-party candidate) may still be enough to cause participants to vote along party lines more often regardless of candidate characteristics.

The results of Experiment 2 suggest that the extent to which a candidate's facial features appear competent or incompetent does not matter much when candidate gender and substantive information are taken into account. The findings for vote choice suggest that women candidates who have an incompetent appearance may be less likely to receive a subject's vote, but this only reaches significance at the $p < .1$ level, and the same pattern is not found for feeling thermometer scores. Previous findings which suggest that facial competence also seems to be bound up with masculine features (Olivola and Todorov 2010; Herrick, et al 2010) and that more masculine looking women (who may also be more traditionally competent looking) are actually at a disadvantage among voters (Hehman, et al 2014) may provide a partial explanation as to why this is the case. If competent-looking women are also more masculine looking, and voters prefer women who look feminine, this may be another type of "double bind" that women candidates face when running for office.

Further, the fact that Republicans in this sample seem to be more affected by facial competence than are Democrats points to a potentially important difference in how members of the two parties make judgments about candidates, and female candidates in particular. Several studies have already found differences in facial competence judgments between Democrats and Republicans (e.g. Carpinella, et al

2015; Carpinella and Johnson 2013a, b), and this may be further indication that Republicans either infer different traits from female faces than Democrats do, or that they incorporate these kinds of judgments more readily into evaluations of female candidates than do Democrats. Further research is needed.

Interestingly, my analyses include only an indicator for whether the subject was in the “incompetent information” condition, rather than a count of the number of competence items that the subject accessed. Inclusion of such a count variable does not alter the results, suggesting that it does not matter how many competence-related items subjects saw related to the candidate. Just one suggestion of incompetence had essentially the same effect on evaluations as up to 6. This makes sense given psychological research indicating that people are inclined to maintain stereotypes if possible, so it does not take much to convince us that our stereotypic assumptions are correct (e.g. Leyens and Yzerbyt 1992; Wojcieszki 1994). Even one or two pieces of evidence that a female candidate is less competent than her opponent may be enough to confirm voters’ expectations and result in more negative evaluations or a vote for the “other guy.”

Of course, like all studies, the experiments in this paper have limitations. Experimental research sacrifices external validity for the ability to make causal inferences and this inherently introduces an element of artificiality. While DPTE experiments attempt to more closely mimic the real world by introducing stimuli in a high-information environment, and using DPTE to test the effects of a stimulus on candidate evaluation is likely a closer approximation of the real world than survey experiments that manipulate candidate characteristics in isolation, they are far from a perfect representation. Subjects are exposed to a relatively large amount of information in a relatively short time frame—and certainly a much shorter time frame than occurs in actual elections. This means that the effects of the competence items could be diluted if we were to study them over a longer period of time. However, since one indication of incompetence had essentially the same effect on evaluations and vote choice as seeing multiple competence-related items, just one indication of possible incompetence seems to be enough to be damaging, even in the face of a substantial amount of competing information. This gives me greater confidence that these results could persist over time, as well. Conducting a similar experiment that takes

place over multiple days or weeks (see Mitchell 2012; 2014 e.g.) might give a better indication of the effects of competence evaluations over time and is certainly an area for further research.

Another limitation is the nature of the samples. Because they were drawn primarily from MTurk, they are clearly not representative of the population as a whole. In particular, Democrats outnumber Republicans by a substantial margin in both studies. While this is certainly not ideal, a sample that is skewed to the left is likely a “tough case” with which to test my theory. Because women are more likely to run and win as Democrats, and liberals are less likely to hold traditional views on gender (e.g. King and Matland 2003; Dolan 2010), it is possible that the results found in this study would be further exaggerated in a sample that more closely mimicked the population as a whole. However, the limited number of Republican subjects does limit my ability to discuss inter-party differences, particularly those related to candidate appearance in Experiment 2, and only a more evenly-split sample will allow for more comprehensive examination of these sorts of partisan differences.

Limitations aside, the results of these studies have important implications for scholars of women and politics in that they lend support to previous findings that suggest that gender-based stereotypes may be playing a nuanced role in candidate evaluation and vote choice. While evidence is mixed regarding whether gender stereotypes have direct effects on vote outcomes, an examination of the process of information acquisition suggests that voters do care about different things when they are evaluating women vs. men and that certain kinds of information can affect men and women differently. This suggests that gender stereotypes *can* have negative effects for female candidates, but that they do not have to. The nature of the other information available to subjects aside from gender seems to matter a great deal, but candidate gender influences how that information matters.

This finding has important real-world implications for female candidates, as well. The fact that voters are reacting differently to women portrayed as competent vs. those who are portrayed as incompetent is arguably a good thing—it seems that they are allowing substantive information to influence their ultimate evaluations and vote choice. This suggests that women who run for office and are portrayed as competent may not be disadvantaged compared to men. What is more problematic is that

women are more disadvantaged by negative information than are men. Of course, information about political candidates of either gender is obtained in large part from the news media, and this has proven to be a source of problems for women candidates. Since gender stereotypes are often perpetuated through the media and women candidates are often “trivialized” in the way the media talks about them (e.g. Braden 1996; Devitt 2002) unbiased portrayals of women as competent politicians are likely not easy to come by. Further, even if media coverage of female candidates is unbiased and free of gender-stereotypic content (e.g. Hayes, Lawless and Baitinger 2015; Hayes and Lawless 2015), voters seem to be reacting to ostensibly gender-neutral information differently depending upon whether it is applied to a woman vs. a man. Individual campaigns, then, would likely be well served by focusing on female candidates’ competence and qualifications, and they should also be aware of the importance of even just a few pieces of information that may cast doubt on their competence.

Acknowledgments

This research was funded in part by a Doctoral Dissertation Research Improvement Grant from the National Science Foundation. A previous version of the paper was presented at the New Research on Gender and Political Psychology conference at the College of Wooster in 2014. Many thanks to all the participants there for the inspiration and support, and especially to Kris Kanthak, Nichole Bauer, Amanda Johnston, and Rebecca Bigler for their very thoughtful feedback. Thanks also to Rick Lau, Dave Redlawsk, Kira Sanbonmatsu, Jennifer Merolla, Dave Andersen, Amy Erica Smith, and Robert Urbatsch for their comments on various versions of the paper.

Works Cited

- Alexander, Deborah and Kristi Andersen. 1993. Gender as a Factor in the Attribution of Leadership Traits. *Political Research Quarterly* 46 (3): 527-545.
- Ashmore, Richard D. 1981. Sex Stereotypes and Implicit Personality Theory in David L. Hamilton (ed) *Cognitive Processes in Stereotyping and Intergroup Behavior*. New York: Psychology Press, 37-64.
- Ballew, C. C., & Todorov, A. 2007. Predicting political elections from rapid and unreflective Face judgments. *Proceedings of the National Academy of Sciences of the USA*, 104, 17948–17953.
- Bartels, Larry. 2002. The Impact of Candidate Traits in American Presidential Elections in Anthony King(ed) *Leaders' Personalities and the Outcomes of Democratic Elections*. Oxford: Oxford University Press, 44-69.
- Bauer, Nichole M. 2015a. Emotional, Sensitive, and Unfit for Office? Gender Stereotype Activation and Support Female Candidates. *Political Psychology*, 36(6): 691-708.
- Bauer, Nichole M. 2015b. "Who stereotypes female candidates? Identifying individual differences in feminine stereotype reliance." *Politics, Groups, and Identities* 3(1): 94-110.
- Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351-388.
- Bettencourt, B. Ann, Karen E. Dill, Scott A. Greathouse, Kelly Charlton, and Amy Mulholland. 1997. Evaluations of InGroup and Outgroup Members: The Role of Category-Based Expectancy Violation. *Journal of Experimental Social Psychology* 33(3): 244-275.
- Bittner, Amanda. *Platform or personality? The role of party leaders in elections*. Oxford University Press: Oxford, 2011.
- Blair, Irene V. 2002. The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), pp.242-261.
- Blair, Irene V., and Mahzarin R. Banaji. 1996. "Automatic and controlled processes in stereotype priming." *Journal of personality and social psychology* 70(6): 1142-1163.
- Borgida, Eugene, Anne Locksley, and Nancy Brekke. 1981. Social Stereotypes and Social Judgment, *Personality, Cognition and Social Interaction*, Hillsdale, NJ: Erlbaum, 153-169.
- Braden, Maria. 1996. *Women, Politics and the Media*. Lexington: University Press of Kentucky.
- Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes do not Harm Women Candidates*. Princeton, NJ: Princeton University Press.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive Yet High Quality Data? *Perspectives on Psychological*

Science, 6(1), 3-5.

- Carli, Linda L., and Alice H. Eagly. 1999. "Gender effects on social influence and emergent leadership." In Powell, Gary N. (Ed.) *Handbook of Gender and Work*. Thousand Oaks, CA: Sage, 203-222.
- Carpinella, Colleen M., Eric Hehman, Jonathan B. Freeman, and Kerri L. Johnson. 2015. "The Gendered Face of Partisan Politics: Consequences of Facial Sex Typicality for Vote Choice." *Political Communication* 33(1): 1-18.
- Carpinella, Colleen M., and Kerri L. Johnson. 2013a "Politics of the face: The role of sex-typicality in trait assessments of politicians." *Social Cognition* 31(6): 770-779.
- Carpinella, Colleen M., and Kerri L. Johnson. 2013b. "Appearance-based politics: Sex-typed facial cues communicate political party affiliation." *Journal of Experimental Social Psychology* 49(1):156-160.
- Carroll, Susan J. and Kelly Dittmar. 2010. "The 2008 Candidacies of Hillary Clinton and Sarah Palin: Cracking the "Highest, Hardest Glass Ceiling" in *Gender and Elections: Shaping the Future of American Politics*. Susan J. Carroll and Richard L. Fox, eds. New York: Cambridge University Press.
- Clifford, Scott, Ryan M. Jewell, and Phillip D. Waggoner. 2015. Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology? *Research and Politics*. DOI: 10.1177/2053168015622072
- Cook, Elizabeth Adell, Sue Thomas, and Clyde Wilcox, eds. 1994. *The Year of the Woman: Myths and Realities*. Boulder: Westview Press. Darcy, R. S. and S. S. Schramm. 1977. "When Women Run Against Men." *Public Opinion Quarterly*, 41(1): 1-12.
- Crump, Matthew J.C., John V. McDonnell, and Todd M. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. PLoS ONE. DOI: 10.1371/journal.pone.0057410
- Darcy, R., S. Welch, and J. Clark. 1994. *Women, Elections, and Representation*. Lincoln: University Of Nebraska Press.
- Devine, Patricia G. 1989 "Stereotypes and prejudice: their automatic and controlled components." *Journal of personality and social psychology* 56(1): 5-18.
- Devitt, James. 2002. Framing gender on the campaign trail: Female gubernatorial candidates and the press. *Journalism & Mass Communication Quarterly*, 79(2), pp.445-463.
- Ditonto, Tessa, Allison Hamilton, and David Redlawsk. 2014. "Gender Stereotypes, Information Search and Voting Behavior in Political Campaigns." *Political Behavior*, 36: 335-358.
- Dolan, Kathleen. 2004. *Voting for Women How the Public Evaluates Women Candidates*. Boulder, CO: Westview Press.
- Dolan, Kathleen. 2010. The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior* 32(1), 69-88.

- Dolan, Kathleen. 2014. *When does Gender Matter? Women Candidates and Gender Stereotypes In American Elections*. New York: Oxford.
- Eagly, Alice H. and Karau, Steven J. 2002. Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3): 573-598.
- Eagly, Alice H. and Wendy Wood. 1982. Inferred Sex Differences in Status as a Determinant of Gender Stereotypes about Social Influence. *Journal of Personality and Social Psychology*, 43(5): 915-928.
- Foschi, Martha. 2000. "Double standards for competence: Theory and research." *Annual Review of Sociology* 26: 21-42.
- Fridkin, Kim L. and Patrick J. Kenney. 2011. The Role of Candidate Traits in Campaigns. *Journal of Politics*, 73(1): 61-73.
- Funk, Carolyn. 1999. Bringing the Candidate into Models of Candidate Evaluation. *Journal of Politics*, 61(3): 700-720.
- Goren, Paul. 2002. "Character weakness, partisan bias, and presidential evaluation." *American Journal of Political Science*. 46(3):627-641.
- Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (2009). Shallow cues with deep effects: Trait judgments from faces and voting decisions. In E. Borgida, J. L. Sullivan, & C. M. Federico (Eds.), *The political psychology of democratic citizenship* (pp. 73–99). New York: Oxford University Press.
- Hassin, Ran and Yaacov Trope. 2000. "Facing Faces: Studies on the Cognitive Aspects of Physiognomy." *Journal of Personality and Social Psychology*, 78(5): 837-852.
- Hayes, Danny. 2005. Candidate Qualities through a Partisan Lens: A Theory of Trait Ownership. *American Journal of Political Science*, 49(4): 908-923.
- Hayes, Danny. 2011. When Gender and Party Collide: Stereotyping in Candidate Trait Attribution. *Politics and Gender*, 7(2): 133-165.
- Hayes, Danny and Jennifer L. Lawless. 2015. 2015. A Non-Gendered Lens? Media, Voters, and Female Candidates in Contemporary Congressional Elections. *Perspectives on Politics*, 13(01), pp.95-118.
- Hayes, Danny, Jennifer L. Lawless, and Gail Baitinger. 2014. Who cares what they wear? Media, gender, and the influence of candidate appearance. *Social Science Quarterly*, 95(5), pp.1194-1212.
- Hehman, Eric, Colleen M. Carpinella, Kerri L. Johnson, Jordan B. Leitner, and Jonathan B. Freeman. 2014. "Early processing of gendered facial cues predicts the electoral success of female politicians." *Social Psychological and Personality Science* 5(7): 815-824.
- Herrick, Rebekah, Jeanette Mendez, Sue Thomas, Amanda Wilkerson. 2012. Gender and Perceptions of Candidate Competency. *Journal of Women Politics and Policy*, 33(2): 126-150.

- Higgle, Ellen, Penny M. Miller, Todd G. Shields and Mitzi M. S. Johnson. 1997 "Gender Stereotypes and Decision Context in the Evaluation of Political Candidates." *Women and Politics*, 17(3): 69-88.
- Holian, David B. and Charles L. Prysby. 2014. *Candidate Character Traits in Presidential Elections*, Vol. 18. New York: Routledge.
- Holman, Mirya R., Merolla, Jennifer L. and Elizabeth J. Zechmeister. 2011. Sex, stereotypes, and security: A study of the effects of terrorist threat on assessments of female leadership. *Journal of Women, Politics & Policy*, 32(3), 173-192.
- Huddy, Leonie and Theresa Capelos. 2002. Gender Stereotyping and Candidate Evaluation: Good News And Bad News for Women Politicians. In *The Social Psychology of Politics*. Springer US: 29-53.
- Huddy, Leonie and Nayda Terkildsen. 1993a. Gender Stereotypes and Perceptions of Male and Female Candidates. *American Journal of Political Science*, 37: 119-147.
- Huddy, Leonie and Nayda Terkildsen. 1993b. The Consequences of Gender Stereotypes for Women Candidates and Different Levels and Types of Office. *Political Research Quarterly*, 46(3): 503-525.
- Kahn, Kim Fridkin. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York: Columbia.
- King, David C. and Richard E. Matland. Sex and the Grand Old Party: An Experimental Investigation of the Effect of Candidate Sex on Support for a Republican Candidate. *American Politics Research*, 31(6), 595-612.
- Kunda, Ziva, and Steven J. Spencer. 2003. "When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application." *Psychological bulletin* 129(4): 522-544.
- Lau, Richard R. and David P. Redlawsk. 2006. *How Voters Decide: Information Processing During Political Campaigns*. New York: Cambridge University Press.
- Lau, Richard R. and David P. Redlawsk. 2001. Advantages and Disadvantages of Cognitive Heuristics in Political Decision-Making. *American Journal of Political Science*, 45: 951-971.
- Laustsen, Lasse. 2014. Decomposing the Relationship Between Candidates' Facial Appearance and Electoral Success. *Political Behavior* 36(4): 777-791.
- Lawless, Jennifer L. 2004. "Women, War, and Winning Elections: Gender Stereotyping in the Post-September 11th Era." *Political Research Quarterly* 57 (3): 479-490.
- Lenz, Gabriel S. and Chappell Lawson. 2007. Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance. *American Journal of Political Science*.
- Lepore, Lorella and Rupert Brown. 1997. Category and stereotype activation: Is prejudice inevitable?. *Journal of personality and social psychology*, 72(2), p.275.

- Locksley, Anne, Eugene Borgida, Nancy Brekke, and Christine Hepburn. 1980. Sex Stereotypes And Social Judgment. *Journal of Personality and Social Psychology*, 39(5): 821-831.
- Locksley, Anne, Christine Hepburn, and Vilma Ortiz. 1982. Social Stereotypes and Judgments of Individuals: An Instance of Base-Rate Fallacy. *Journal of Experimental Social Psychology*, 18(1): 23-42.
- Lodge, Milton, and Charles S. Taber. 2013. *The rationalizing voter*. New York: Cambridge University Press.
- Matson, Marsha and Terri Susan Fine. 2006. "Gender, Ethnicity, and Ballot Information: Ballot Cues In Low-Information elections," *State Politics and Policy Quarterly*, 6(1): 49-72.
- Mattes, Kyle, Michael Spezio, Hackjin Kim, Alexander Todorov, Ralph Adolphs, and R. Michael Alvarez. 2010. *Political Psychology* 31(1): 41-58.
- McDermott, Monika. 1998. Race and Gender Cues in Low-Information Elections. *Political Research Quarterly* 51 (4): 895-918.
- Mitchell, Dona-Gene. 2012. "It's about time: The lifespan of information effects in a multiweek campaign." *American Journal of Political Science* 56(2): 298-311.
- Mitchell, Dona-Gene. 2014. "Here today, gone tomorrow? Assessing how timing and repetition of scandal information affects candidate evaluations." *Political Psychology* 35(5): 679-701.
- Mondak, Jeffrey J. 1995. "Competence, Integrity, and the Electoral Success of Congressional Incumbents." *The Journal of Politics*, 57(4): 1043-1069.
- Oliviola, Christopher and Alexander Todorov. 2010a. Elected in 100 Milliseconds: Appearance-Based Trait Inferences and Voting. *The Journal of Nonverbal Behavior*, 34: 83-110.
- Oliviola, Christopher and Alexander Todorov. 2010b. Fooled by First Impressions? Reexamining the Diagnostic Value of Appearance-Based Inferences. *Journal of Experimental Psychology*, 40: 267-280.
- Paolacci, Gabriele and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk As a Participant Tool. *Current Directions in Psychological Science*, 23(3), 184-188.
- Prysbly, Charles. 2008. "Perceptions of candidate character traits and the presidential vote in 2004." *PS: Political Science & Politics* 41(1): 115-122.
- Rahn, Wendy M., John Borgida, Eugene Borgida, and John Sullivan. 1990. A Social-Cognitive Model Of Candidate Appraisal. In Ferejohn, John and James Kuklinski (eds) *Information and Democratic Processes*. Urbana, IL: University of Illinois Press. 136-159.
- Redlawsk, David and Richard Lau. 2006. "I Like Him But...: Vote Choice when Candidate Likability and Closeness on Issues Clash" in David Redlawsk (ed.) *Feeling Politics: Emotion in Political Information Processing*. New York: Palgrave MacMillan.
- Rosenwasser, Shirley M., and Jana Seale. 1988. Attitudes Towards a Hypothetical Male or Female Presidential Candidate—A Research Note." *Political Psychology* 9 (4): 591-598.

- Sanbonmatsu, Kira and Kathleen Dolan. 2009. Do Gender Stereotypes Transcend Party? *Political Research Quarterly*, 62(3): 485-494.
- Schneider, Monica, and Angela Bos. 2009. "It Don't Matter if You're Black or White: Exploring the Content of Stereotypes of Black Politicians." Paper presented at the Annual Meeting of the Midwest Political Science Association. Chicago.
- Schneider, Monica C. and Bos, Angela L. 2014. Measuring stereotypes of female politicians. *Political Psychology*, 35(2): 245-266.
- Seltzer, Richard, Jody Newman, and Melissa Voorhees Leighton. 1997. *Sex as a Political Variable: Women as Candidates and Voters in US Elections*. Boulder, CO: Lynne Rienner.
- Shabad, Goldie and Kristi Andersen. 1979. Candidate evaluations by men and women. *Public Opinion Quarterly*, 43(1), pp.18-35.
- Sigelman, Carol, Lee Sigelman, Barbara Walkosz, and Michael Nitz. 1995. Black Candidates, White Voters: Understanding Racial Bias in Political Perceptions. *American Journal of Political Science*, 39: 243-265.
- Sinclair, Lisa, and Ziva Kunda. 1999. "Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes." *Journal of personality and social psychology* 77(5): 885-904.
- Sussman, Abigail B., Kristina Petkova, and Alexander Todorov. 2013. "Competence Ratings in US Predict Presidential Election Outcomes in Bulgaria." *Journal of Experimental Social Psychology*, 49(4): 771-775.
- Todorov, Alexander, Christopher Y. Olivola, Ron Dotsch, and Peter Mende-Siedlecki. 2015. Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology* 66: 519-545.
- Todorov, A. 2008. Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In A. Kingstone & M. Miller (Eds.), *The Year in Cognitive Neuroscience 2008: Annals of the New York Academy of Sciences* (Vol. 1124, pp. 208–224).
- Todorov, A. 2010. Evaluating faces on social dimensions. In A. Todorov, S.T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind*. Oxford University Press.
- Todorov, A. and A. Engell. 2010. The Role of the Amygdala in Implicit Evaluation of Emotionally Neutral Faces. *Social Cognitive and Affective Neuroscience*, 3: 303-312.
- Todorov, A., & Uleman, J. S. 2002. Spontaneous trait inferences are bound to actor's faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83, 1051–1065.
- Todorov, A., & Uleman, J. S. 2003. The efficiency of binding spontaneous trait inferences to actor's faces. *Journal of Experimental Social Psychology*, 39, 549–562.

- Todorov, A., Chaiken, S., & Henderson, M. D. 2002. The heuristic-systematic model of social information processing. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 195–211). Thousand Oaks, CA: Sage.
- Todorov, A., Mandisotza, A. N., Goren, A., & Hall, C. 2005. Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. 2008. Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12, 455–460.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. 2009. Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27, 813–833.
- Weinberg, Jill D., Jeremy Freese, and David McElhattan. 2014. Comparing Data Characteristics And Results of an Online Ractorial Survey Between a Population-Based and a Crowdsourced-Recruited Sample. *Sociological Science*, 1, 292-310.

Table 1. Sample and Design Details, Study 1 and Study 2

	Study 1	Study 2
Total N	449	377
Source	106, in-person convenience sample 343, Amazon Mechanical Turk	Amazon Mechanical Turk
% Women	58%	50%
% White	76%	83%
% African American	9%	6.5%
% Latino	5%	7.5%
% Democrat	48%	64%
% Republican	16%	21%
% Independent	31%	15%
Median Age	32	35
Candidate Gender Manipulation	Yes	Yes
Informational Competence Manipulation	Yes, 6 information items, 1 forced	Yes, 6 information items, 1 forced
Appearance-Based Competence Manipulation	No	Yes
Candidate Images	Pictures of actual candidates taken state legislators' websites	Created using "FaceGen" software
Out-Party Candidate	Competent, white, male	Competent, white, male
Assignment of Independents	Self-selection	Randomly assigned

Table 2. ANOVA Results for Feeling Thermometer Difference Score, Experiments 1 and 2

Note: Values are F-statistics

	Candidate Gender	Competence of Information	Facial Competence	Gender X Competence of Information	Gender X Facial Competence
Experiment 1 (n=439)	13.231***	7.673*	--	4.019*	--
Experiment 2 (n=372)	.390	13.586***	2.420	8.056*	.606

*p<.05, **p<.01, ***p<.001

Table 3.
Logistic Regression Results for In-Party Vote, Experiment 1

	Model 1	Model 2
Primary Partisan Match	1.521*** (.312)	1.515*** (.312)
Female Candidate	-.451* (.218)	-.275 (.325)
Incompetent Info	-.591** (.219)	-.425 (2.948)
Female Candidate X Incompetent Info	--	-.319 (.438)
Constant	.117 (.324)	.030 (.343)
R-square	.109	.111
N	448	448

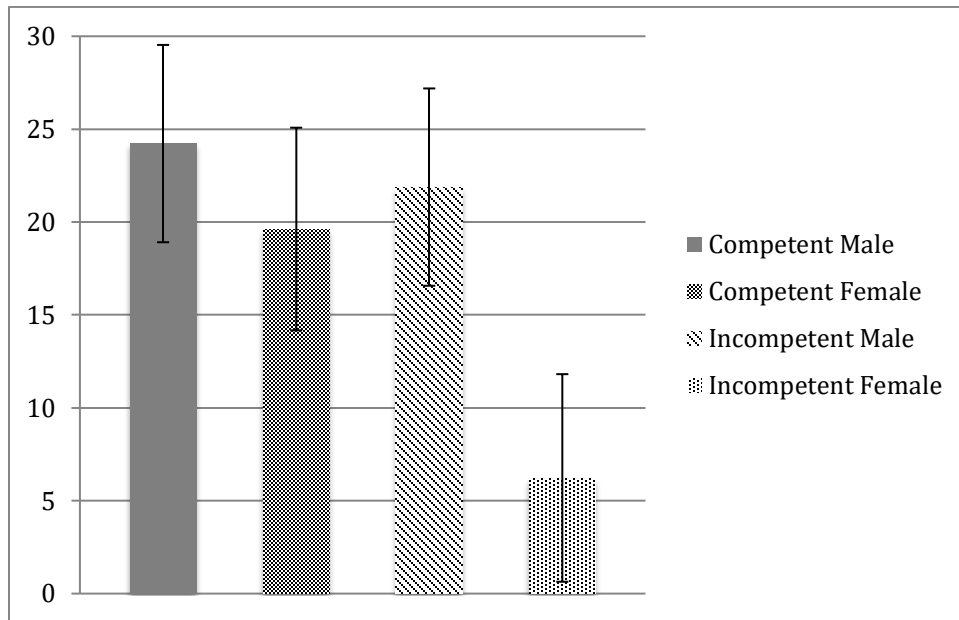
†p<.1, *p<.05, **p<.01, ***p<.001

Table 4.
Logistic Regression Results for In-Party Vote Choice, Experiment 2

	Model 1	Model 2
Strength of PID	.536*** (.112)	.552*** (.114)
Female Candidate	-.238 (.218)	.892* (.400)
Incompetent Info	-.606** (.220)	-.287 (.374)
Incompetent Face	-.037 (.218)	.627 (.543)
Female Candidate X Incompetent Info	--	-1.302** (.453)
Female Candidate X Incompetent Face	--	-.796 [†] (.451)
Incompetent Info X Incompetent Face	--	.627 (.453)
Constant	-.832* (.363)	-1.215** (.409)
Pseudo R-square	.112	.155
N	372	372

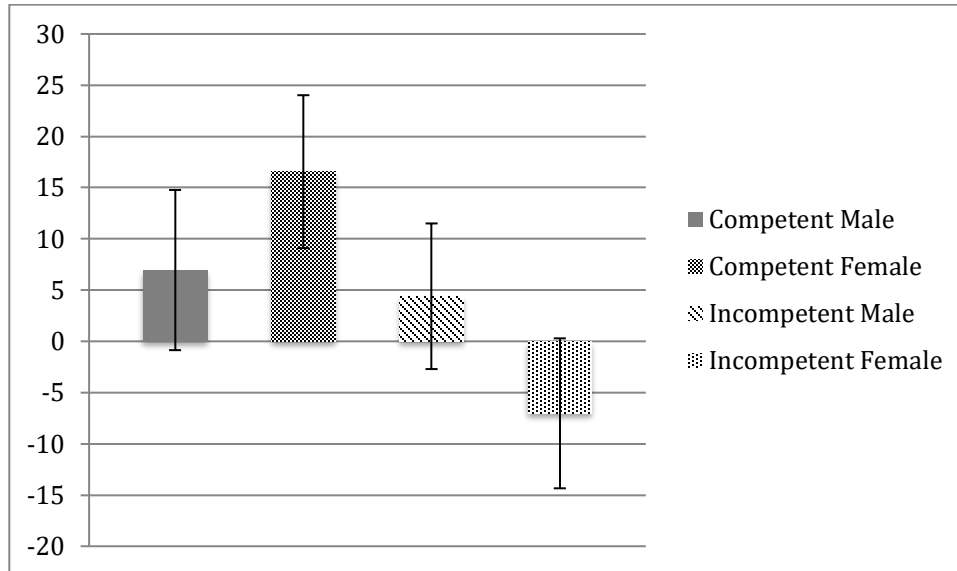
[†]p<.1, *p<.05, **p<.01, ***p<.001

Figure 1. Feeling Thermometer Difference Score Means, Experiment 1



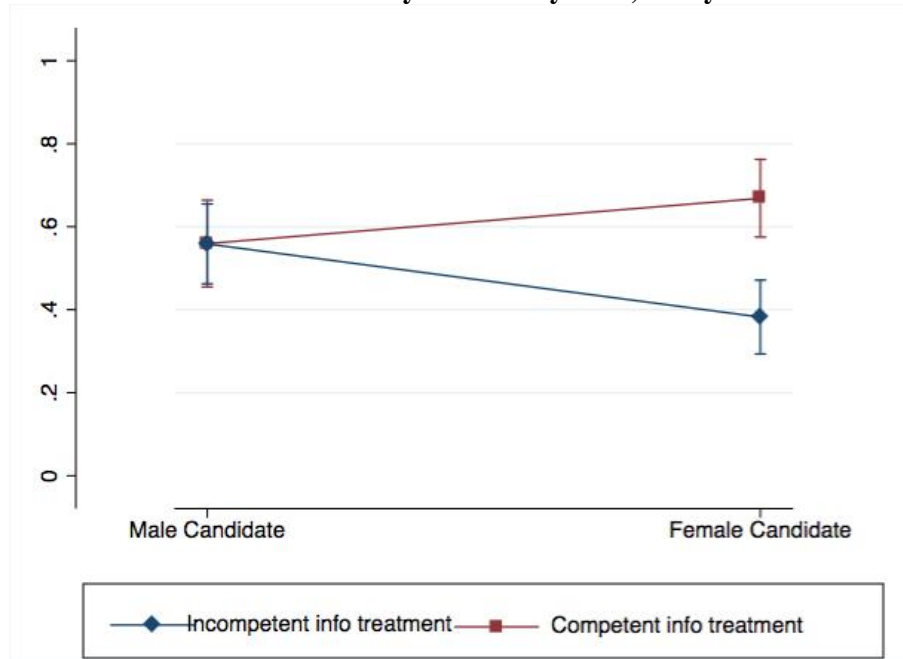
**Note: Statistical Test ANOVA. $F=4.019$, $p<.05$. $N= 448$
Error bars represent 95% confidence intervals.**

Figure 2. Feeling Thermometer Difference Score Means, Experiment 2



**Note: Note: Statistical Test is ANOVA. $F=8.056$, $p<.05$. $N= 372$
Error bars represent 95% confidence intervals.**

Figure 3.
Predicted Probability of In-Party Vote, Study 2




Note: Error bars indicate 95% confidence intervals, N=372


Online Appendix A: Experimental Design


Screenshot of a DPTE Campaign


Stage: 2012 Presidential Election
Sub-stage: Primary Election


Time Remaining: 28:07
8:22


 **Susan Turner's Religion**

 **Debra Johnson's Party Affiliation**

 **Lou Baker's Campaign Slogan**

 **Susan Turner's stance on Military Intervention**

 **Jim Davis's Stand on the Defense Budget**

 **Lou Baker's Military Experience**

DPTE: Dynamic Process Tracing Environment 2.5.1, Copyright (c) 2007 - 2012 The University of Iowa, all rights reserved.

Screenshot of Open Information Box in a DPTE Campaign

The screenshot shows a software interface for a DPTE campaign. At the top, a grey header bar contains the text "Stage: 2012 Presidential Election" on the left and "Time Remaining: 27:50" on the right. Below the stage text, it says "Sub-stage: Primary Election". Below the time, it says "8:05". The main content area is a large white rectangle with a blue border. Inside this rectangle, the title "Susan Turner's Stance on the Defense Budget" is displayed in bold. Below the title, a paragraph of text reads: "Susan Turner believes that defense spending in the US is out of proportion to the needs of the nation, and must be limited. Susan Turner would like to see reductions in the size of the military budget, mainly by reducing the number of bases operated domestically and abroad, and by limiting the research and development budget given to the department of defense." In the bottom right corner of the white rectangle, there is a small mouse cursor and a "Close" button. At the very bottom of the interface, a small grey bar contains the text: "DPTE: Dynamic Process Tracing Environment 2.5.1, Copyright (c) 2007 - 2012 The University of Iowa, all rights reserved."

Stage: 2012 Presidential Election
Sub-stage: Primary Election

Time Remaining: 27:50
8:05

Susan Turner's Stance on the Defense Budget

Susan Turner believes that defense spending in the US is out of proportion to the needs of the nation, and must be limited. Susan Turner would like to see reductions in the size of the military budget, mainly by reducing the number of bases operated domestically and abroad, and by limiting the research and development budget given to the department of defense.

Close

DPTE: Dynamic Process Tracing Environment 2.5.1, Copyright (c) 2007 - 2012 The University of Iowa, all rights reserved.

Image Pretest Details, Experiments 1 and 2

The images in Experiment 1 were taken from state legislatures' websites in multiple states and chosen because they showed officials' headshots against a single-colored background. 64 such images (including equal numbers of men, women, white, and black candidates) were pretested by 148 undergraduates at [redacted]. Subjects were asked to rate a subset of 16 of these pictures on a series of traits, including compassion, competence, strong leadership, and trustworthiness. Ratings were on a 1-4 scale (Banducci, Karp, Thrasher, and Rallings 2008) and photographs with the highest and lowest average competence scores were chosen for use in the primary election. Again, those with the lowest competence scores were used in the general election. Those images are available below.

In Experiment 2, I used images created by FaceGen, a face-generating software. Facegen allows the user to begin with a "neutral" face and alter its appearance in many various ways. 24 candidate faces were created—12 male faces and 12 female faces, varying the aspects of physical appearance that are thought to signal competence, including eye spacing, jaw width, and maturity/"baby-facedness". These images were then pretested in a manner similar to the images in Experiment 1. 120 subjects were recruited from Amazon Mechanical Turk and each subject was asked to rate a random subset of 12 of the 24 pictures on competence, compassion, attractiveness, likeability, and perceived age. The male and female pictures with the highest and lowest average competence scores were selected for use in the experiment. Further information on either pretest can be provided upon request.

Candidate Pictures, Experiment 1



FaceGen-generated Candidate Images, Experiment 2

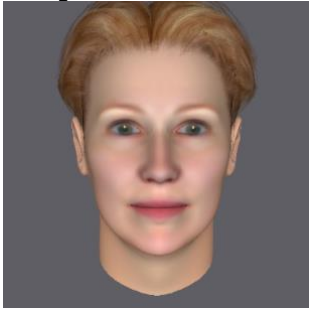
Competent Male Candidate



Incompetent Male Candidate



Competent Female Candidate



Incompetent Female Candidate



Competence Information Items, Experiments 1 and 2

Competent Information Condition:

In the presidential debate last night between Daniel Baker and [Candidate], [Candidate] maintained poise and command during tough questioning. By remaining steady and never appearing threatened, [Candidate] developed an image of strong leadership. It is generally acknowledged that [Candidate] won this debate, hands down.

The Cincinnati Enquirer proclaims, "[Candidate] continues to show the nation that [he/she] has what it takes--in experience, spirit and intellect--to win this election and lead this country to prosperity."

After nearly 2 decades of holding statewide office, [Candidate] is a growing presence in national politics. Rising to the highest level of state politics by consistently offering solid ideas and delivering on promises, [Candidate] has developed a reputation as a capable and creative politician.

"[Candidate] rose through state politics by making promises that were reasonable, and then carrying them out. On the national level, it will be much harder for [him/her] to do that. Still, I have a lot of respect for [Candidate]."

After being active in politics in college, [Candidate] ran for and was elected mayor of Sioux Falls when only 30 years old. After repeated terms there, [Candidate] moved to the House of Representatives for 8 years, before winning the Governorship of South Dakota. After 2 terms as Governor, [he/she] is looking to continue in politics at the national level.

"I thought [Candidate] was always pushing too hard for too much, until I realized that it was because everyone responded to it and we were able to accomplish far more than we ever expected. [Candidate] has the highest of expectations, but is rarely disappointed because [his/her] leadership always helped us to achieve our goals."

Incompetent Information Condition:

In the presidential debate last night between Daniel Baker and [Candidate], [Candidate] lost [his/her] poise and command almost from the beginning. [Candidate] appeared defensive and out of [his/her] league during much of the questioning and most agree that Daniel Baker came out on top.

The Cincinnati Enquirer proclaims, "[Candidate]'s actions continue to cast doubt on [his/her] own abilities; in terms of leadership, experience, and sheer capability, [he/she] has yet to convince the nation that [he/she] has what it takes.

After five years of holding statewide office, [Candidate] is a new presence in national politics. While gaining quick popularity among party loyalists, many residents of [his/her] state are less than impressed with [his/her] performance so far, citing numerous broken promises and failed ideas.

[Candidate] rose through state politics by making lots of promises to lots of different people. Carrying them out proved much harder, though, and the people of [his/her] state noticed that. I worry about the path that [Candidate] would lead our nation down if [he/she] were to be elected.

[Candidate] served two terms in the South Dakota state legislature and is currently serving in [his/her] first term as governor of that state.

"I thought [Candidate] was always pushing too hard for too much. [His/Her] expectations were often unrealistic and that led to a lot of disappointment on all our parts."

Experiment 1 Primary Election Description

In the primary, subjects experienced a 2x2x2 experimental manipulation, in which the picture of one of the candidates in their party's primary (Candidate A) varied by gender, race, and whether the candidate had a competent or incompetent appearance. Each party's primary included both a moderate candidate and either a liberal or conservative candidate, depending on the party (i.e. subjects in the Republican primary saw a moderate and a conservative candidate, while those in the Democratic primary saw both a moderate and a liberal candidate). Which candidate received each ideological position was randomly assigned so that Candidate A was moderate half of the time and Candidate B was moderate the other half of the time. This allowed me to control for subjects' ideological preferences. All of the information about the candidates was non-evaluative in tone.

Candidate A, then, could be male or female, black or white, and either competent-looking or incompetent-looking. Candidate B was always a white man and either looked competent or incompetent depending on whether Candidate A received a competent or incompetent picture. If Candidate A was competent-looking (regardless of the gender/race manipulations), Candidate B was always an incompetent-looking white man. If Candidate A was incompetent-looking (again, regardless of race or gender), Candidate B was always a competent-looking white man.

Subjects had the opportunity to access 25 unique pieces of information about each of the two candidates in their primary, as well as 25 pieces of information about the two candidates in the other party's primary. After the campaign, subjects were asked to vote for the candidate of their choice. The incompetent-looking candidate always progressed to the general election.

Online Appendix B: Methodology and Analysis

Table B1. Individual Cell Ns: Experiment 1

	Competent Information	Incompetent Information	Total
Male In-Party Candidate	117	117	234
Female In-Party Candidate	105	109	214
Total	222	226	448

Table B2. Individual Cell Ns: Experiment 2

	Competent Information	Incompetent Information	Total
Male In-Party Candidate	85	96	181
Female In-Party Candidate	86	110	196
Total	171	206	377

	Competent Face	Incompetent Face	Total
Male In-Party Candidate	100	81	181
Female In-Party Candidate	105	91	196
Total	205	172	377

Table B3. Analysis for Partisan Sub-Samples, Experiment 1

Candidate Gender X Competence of Information

	Democrats	Republicans
N	252	90
Feeling Therm	F=4.467*	F=1.217
Mean, Competent Female	24.938 (3.580)	20.810 (6.897)
Mean, Competent Male	32.938 (3.552)	12.481 (6.082)
Mean, Incompetent Female	7.233 (3.697)	1.167 (7.449)
Mean, Incompetent Male	30.508 (3.608)	7.864 (6.738)
Vote choice	B=-.669 (.675)	B=-.560 (.891)

**Table B4. Analysis for Partisan Sub-Samples, Experiment 2
Gender X Competence of Information**

	Democrats	Republicans
N	252	80
Feeling Therm	F=3.915*	F=1.334
Mean, Competent Female	20.254 (4.933)	12.791 (7.936)
Mean, Competent Male	8.647 (5.178)	13.647 (9.995)
Mean, Incompetent Female	-3.125 (4.494)	-1.674 (7.921)
Mean, Incompetent Male	3.906 (4.202)	20.829 (10.393)
Vote choice	B=-.956 [†] (.582)	B=-.459 (1.165)

**Table B5. Analysis for Partisan Sub-Samples, Experiment 2
Gender X Facial Competence**

	Democrats	Republicans
N	237	80

Feeling Therm	F=.260	F=2.964 [†]
Mean, Competent Female	13.419 (4.627)	18.412 (7.572)
Mean, Competent Male	8.755 (4.423)	14.016 (9.309)
Mean, Incompetent Female	3.710 (4.796)	-7.294 (8.351)
Mean, Incompetent Male	3.799 (4.959)	20.460 (11.010)
Vote choice	B=-.851 (.562)	B=-3.278** (.1163)

**Table B6. Analysis by gender – Experiment 1
Gender X Competence of Information**

	Male	Female
N	188	260
Feeling Therm	F=1.531**	F=4.207*
Mean, Competent Female	15.904 (4.353)	23.429 (3.834)
Mean, Competent Male	17.225 (4.963)	17.765 (3.335)
Mean, Incompetent Female	7.109 (4.628)	4.719 (3.800)
Mean, Incompetent Male	19.820 (4.439)	24.078 (3.586)
Vote choice	B=-.786 (.666)	B=-.035 (.614)

**Table B7. Analysis by gender – Exp. 2
Gender X Competence of Information**

	Male	Female
N	187	185
Feeling Therm	F=4.236*	F=4.697*

Mean, Competent Female	15.011 (5.425)	17.765 (5.449)
Mean, Competent Male	2.162 (6.541)	10.444 (4.918)
Mean, Incompetent Female	7.782 (5.345)	1.045 (7.449)
Mean, Incompetent Male	-2.320 (4.747)	-13.560 (6.738)
Vote choice	B=-1.318* (.655)	B=-1.420* (.673)

**Table B8. Analysis by gender – Exp. 2
Gender X Competence of Appearance**

	Male	Female
N	187	185
Feeling Therm	F=.359	F=.629
Vote choice	B=-1.043 (.645)	B=-.962 (.680)