Differences in Appearance-Based Trait Inferences for Male and Female Political Candidates

To what extent does a political candidate's physical appearance affect his or her chances of winning an election? From Chris Christie's weight to Hillary Clinton's age, much media attention is often devoted to how political figures look, but do voters really incorporate these sorts of seemingly superficial considerations into their voting calculus on Election Day? For quite some time, studies have found a link between attractiveness and election results (e.g., Efron and Patterson 1974) and a growing body of research in political psychology has recently begun to consider the effects of other aspects of a candidate's physical appearance on voting behavior. In particular, a series of studies have persuasively demonstrated that appearance-based rapid trait inferences from very short exposures to images of candidates' faces can predict the results of actual elections. Evidence from these studies suggests that candidates who are rated as more competent by experimental subjects are more likely to win their races than their less competent-looking opponents (e.g. Todorov et al. 2005), and that competence judgments are actually more reliable predictors than attractiveness judgments. These reflexive competence judgments, which are made after nothing more than very short exposures to still images, have been found to predict election winners at multiple levels of government and in several different countries (see Todorov, Oliviola, Dotsch, and Mende-Siedleke 2015 for an overview of this literature).

While findings that automatic trait judgments predict election outcomes are robust and consistent across a large and growing number of studies, most of these

experiments do not systematically consider the gender of the candidates involved. This is likely because a separate line of research has found substantial evidence that women candidates are subject to a number of gender-based stereotypes that may affect the way traits are inferred from a female candidate's appearance vs. a man's (e.g. Huddy and Terkildsen 1993; Cook, Thomas and Wilcox 1994; Dolan 2004; Rosenwasser and Seale 1988). Relevant to this study, female candidates are often rated as less competent and less strong or "tough" than their male counterparts (e.g. Huddy and Terkildsen 1993; Kahn 1996). It is still unclear whether these assumptions extend to reflexive trait judgments as well, though one notable exception (Herrick, Thomas, Mendez, and Wilkerson 2012) suggests that they do; they find that automatic competence inferences, in particular, vary based on gender, and that women tend to be seen as less competent than men.

Further, judgments of attractiveness may matter more and differently for women candidates than for men. Evidence suggests that there is more of a focus on the physical appearance of women candidates than on that of men (e.g. Schubert and Curran 2001), and there is some evidence that physical attractiveness may positively influence women's electability, at least indirectly (Sigelman, Sigelman and Fowler 1987). For these reasons, attractiveness judgments may be more influential—and more predictive of election results—when women are running in a given race, as opposed to when two men are on the ballot. For the latter, studies have shown the effects of attractiveness to be ambiguous at best (e.g. Todorov et al. 2005).

Because of gender-based differences, it is likely that when we specifically examine races that include women, we will find more nuanced effects of automatic trait inferences, such that the relationship between specific traits and election outcomes will be different when female candidates are in the mix than when they are not. This paper seeks to determine if this is the case for judgments of competence, attractiveness, and perceived threat. Following procedures used in previous trait inference studies, we use data from three experiments in which subjects were asked to rate the personality traits of candidates running for election to state or local office. Subjects were shown pictures of a series of pairs of candidates who ran against each other for a particular office and asked to determine which of the pair was more competent-looking, more attractive, and more physically threatening. While most prior studies using this method have generally only included same-gender (and predominantly male) candidate pairs in their analyses, this paper includes, and focuses on, female-female and mixed-gender pairs of candidates that faced off in real-world electoral contests.

Our expectation that gender influences the relationship between automatic trait judgments and election outcomes is largely borne out in our analysis. We find that, for male-male races, our results generally mimic what previous studies have found. Competence predicts election winners, in that the more competent-looking candidates are more likely to have been victorious. Meanwhile, threat is associated with losing, and attractiveness has little predictive value.

On the other hand, it is *attractiveness,* and not competence, that is the strongest predictor of electoral success in races with two women. Further, we find

that the role of competence in mixed-gender races is more nuanced, in that competence is a particularly strong indicator of electoral success in these races *only* when the female candidate is judged as more competent-looking than her male opponent. Further, we also find that women in mixed-gender races who are perceived to be more threatening than their male counterparts are *more* likely to win election than those who are not. Altogether, the results of our analysis suggest that the traits inferred from faces have very different implications for election outcomes depending on the gender composition of a particular race.

Automatic Trait Assessments and Election Outcomes

Evidence from psychology suggests that, in all forms of human interaction, individuals make spontaneous assessments of others' traits based on their appearance—and on their faces, in particular (e.g. Hall et al. 2009; Hassin and Trope 2000; Oliviola and Todorov 2010a). Judgments made based on faces tend to be reliable, in that people infer the same sorts of traits from the same sorts of faces (Hassin and Trope 2000), but most evidence suggests that these inferred traits do not necessarily correlate with the actual traits possessed by individuals (Alley and Cunningham 1988; but see Berry and Zebrowitz 1988). Regardless of their accuracy (or lack thereof), characteristics derived from faces have been shown to alter people's evaluations of targets, even when other information is available, and subjects seem unable to ignore information about physical appearance, even when they are told to do so (Hassin and Trope 2000) or when other information is also available (Budesheim and DePaola 1994).

Inferences of traits from physical appearance are often involuntary and made very quickly—possibly in as little as 33 milliseconds (Todorov and Uleman 2003; Olson and Marshuetz 2005; Todorov 2008). Though the exact features a face must possess in order to appear competent or threatening is still an open question, 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 2010b; Herrick, et al. 2012). Other findings suggest that trait inferences depend on the extent to which someone's facial structure resembles people in various emotional states—faces that look happy, for instance, tend to be seen as more trustworthy (Montepare and Dobish 2003). Whatever the triggers, because these sorts of trait inferences are occurring frequently and rapidly, it is perhaps not surprising that they affect individuals' decision-making in many different areas, including selection of romantic partners (Oliviola et al. 2009), judicial decisions (Zebrowitz and McDonald 1991) and hiring choices (Naylor 2007).

Politics seems to be no exception. In a series of studies, Todorov and colleagues (Todorov, Mandisotza, Goren and Hall 2005; Hall et al. 2009; Ballew and Todorov 2007; Oliviola and Todorov 2010a; Todorov, Oliviola, Dotsch, and Mende-Siedleke 2015; Mattes, Spezio, Kim, Todorov, Adolphs, and Alvarez 2010; Laustsen 2014; Lenz and Lawson 2011; Sussman, et al. 2013) have found that subjects' spontaneous inferences about a candidate's traits—and competence, in particular—correctly predicted the outcomes of actual elections. Inferences were made only from still images of the candidates' faces, and those rated as more competent by the

subjects were more likely to have been the real election winner. While multiple trait ratings were collected, perceived competence was the only positive trait found to have this effect across multiple studies. Additionally, two of these studies also look at a negative trait assessment—personal threat. They find that candidates with more threatening faces are less likely to win their races (Mattes et al. 2010) and that these rapid evaluations of threat induced brain activation patterns similar to those from voting decisions (Spezio et al. 2008).

The role that attractiveness judgments play in candidate evaluation is less clear. Previous studies have found a "halo effect" for attractive people, in which individuals perceived as attractive are also more likely to be ascribed positive traits (Dion et al. 1972; Kanazawa and Kovar 2004), and there is ample evidence that attractive candidates may do better than unattractive candidates in political scenarios (Riggle, Miller, Shields, & Johnson, 1997; Rosenberg, Bohan, McCafferty, & Harris, 1986; Rosenberg, Kahn, & Tran, 1991). However, there is reason to believe that the link between spontaneous attractiveness judgments and election outcomes is more nuanced and perhaps contextual. For instance, Mattes and Milazzo (2014) find that in British elections, attractiveness judgments better predicted races in marginal constituencies, and competence judgments better predicted elections from less competitive constituencies. Also, while Banducci, et al. (2008) find that more attractive candidates are more likely to do better in electoral contests, this effect is no longer significant when other trait perceptions (e.g., competence) are added to the equation.

While the sum total of many of these trait inference studies suggests that perceptions of competence are particularly important in predicting candidate evaluation and vote choice, we argue that judgments related to threat and attractiveness also have a role to play. Importantly, previous studies of automatic trait inferences have generally controlled for the gender of the candidates that subjects were asked to evaluate, and the vast majority of the candidate pairs shown to subjects were comprised of two men. For reasons discussed below, we posit that the inclusion of female faces in trait inference studies is important and will lead to substantively different results.

Gender Stereotypes

Scholars of gender and politics have found much evidence that female candidates are often subject to particular gender-based trait stereotypes. Women, for example, are perceived as more compassionate, trustworthy, expressive, and honest than men, but less decisive, assertive, rational, experienced than men (Alexander and Andersen 1993; Burrell 2008; Huddy and Terkildsen 1993; Kahn 1996; King and Matland 2003; Lawless 2004; Leeper 1991; Paul and Smith 2008; Sapiro 1982). Of particular interest to this study, women have also been found to be perceived as less competent and experienced than male candidates, as well as less lacking in masculine traits like "toughness" (Huddy and Terkildsen 1993a; Kahn 1996; Lawless 2004; Carroll and Dittmar 2010).

The extent to which these stereotypes influence overall candidate evaluation and vote choice directly is unclear, however (Dolan 2014; Brooks 2013; Hayes 2011), and there is much evidence to suggest that gender stereotypes can be more

or less salient depending on other political factors, such as whether domestic policy issues or things like war or terrorism are at the top of the agenda (Cook, Wilcox and Thomas 1994, Dolan 2004, Lawless 2004), which office is being considered (Huddy and Terkildsen 1993), whether cognitive demand on voters is high or low (Higgle, Miller, Shields, and Johnson 2008; whether an election is high- or low-information (Matson and Fine 2006) and whether stereotypes are activated in campaign messages (Bauer 2015ab). Ditonto, Hamilton and Redlawsk (2014) also find evidence that gender may work indirectly to influence evaluations and voting via the amount and types of information that experimental subjects seek out about that candidate. Specifically, they find that subjects search for more information related to competence when a candidate is a woman.

Further, evidence suggests that women candidates may be judged more heavily on appearance than men and that media coverage of women candidates tends to focus on physical appearance more than coverage of men does (Bystrom, et al. 2001). While some studies have found that this focus on physical appearance is in and of itself detrimental to women candidates (Heflick and Goldenberg 2009; Heflick et al. 2011), there is at least some evidence that being perceived as attractive, per se, is helpful to women running for office. Sigelman, et al. (1987), for example, find that attractiveness indirectly affects female candidates' chances of winning via assessments of femininity, dynamism, niceness, and age, while Schubert and Curran (2001) find that attractiveness matters more for women candidates than for men.

Judgments based on gender are cognitively cheap—it does not take a lot of mental energy to look at someone and notice that he or she is a man or a woman—and evidence suggests that social categories, like gender, are processed easily and quickly when looking at a person's face (Mouchetant-Rostaing and Giard 2003; Mouchetant-Rostaing, Giard, Bentin, Aguera and Pernier 2000; Ito and Urland 2003). Further, because long-term memory is usually conceptualized as a series of interconnected nodes in which associated concepts are activated together, preconceptions about what traits women possess tend to automatically accompany those observations (Anderson 1983). Also, even people that know very little about politics are used to making person judgments in everyday life. For this reason, these sorts of appearance heuristics are particularly salient for most people, regardless of their level of political sophistication (Rahn, et al. 1990, Lau and Redlawsk 2001).

Taken together, the evidence suggests that gender, and the stereotypes and expectations tied to women candidates, should affect appearance-based trait inferences, which should in turn affect vote outcomes. We posit that the influence of appearance-based trait inferences may be another way in which gender can indirectly affect election outcomes. Not only may subjects infer different traits from male and female faces, but they may also care more about different traits given the gender composition of a given race. Indeed, in the only study to consider gender when examining the relationship between rapid trait inferences and election outcomes so far, Herrick, et al. (2012) find systematic differences in competence evaluations for men and women candidates. Women's faces are rated as less mature and less competent than men's faces in both simulated and real elections. Further,

and unlike in automatic trait inference studies that do not include women candidates, they find that, while subjects prefer male candidates in general, competence was *not* significantly related to electoral success. This suggests that the inclusion of women in these studies should, in fact, lead to systematically different results. We expect to find that this is true in our data, as well, and that gender differences also exist for ratings of attractiveness and threat.

Hypotheses

Our analysis compares the ability of three trait judgments (competence, attractiveness and threat) to predict election outcomes among pairs of candidates with different gender compositions. Among mixed-gender pairs, we also consider which of the two candidates was more likely to be judged as possessing each trait, and whether any differences in judgments by gender had implications for election outcomes.

Among two-male candidate pairs, we expect to be able to duplicate prior findings related to competence, attractiveness, and threat. That is, we expect competence judgments to positively predict electoral success, perceptions of threat to predict electoral failure, and attractiveness to have little predictive effect.

For two-woman and mixed-gender elections, our hypotheses differ, as we expect that gender-based considerations will influence the kinds of traits that are inferred from the candidates' faces and that those traits will matter differently to voters depending on the gender composition of a race. First, we expect to find that attractiveness is a positive predictor of electoral success for women facing women, due to an increased attention to physical appearance for female candidates and

evidence that physical attractiveness can help women's chances for election. Further, because women candidates are stereotyped as being less competent, generally, and Herrick, et al. (2012) find that competence assessments are indeed less influential in predicting races that include women, we also expect to find that competence inferences do not predict electoral success in two-woman races as well as attractiveness judgments do.

Extant theory does not give us much insight as to the relationship between threat assessments and electoral outcomes in races with two female candidates. It is possible that these inferences will work in much the same way for all same-gender pairs, whether those pairs include two men or two women, meaning that more threatening candidates will be at a disadvantage. On the other hand, gender-based stereotypes and expectations may mean that other characteristics are simply more important in predicting outcomes in two-woman races, rendering threat inferences less significant.

In mixed-gender races, it is possible that the inclusion of even one woman in a race may make attractiveness more salient than in two-male elections. As mentioned above, we expect female candidates to be rated more heavily on their attractiveness than their male counterparts. Also, the female candidates should be less likely to be rated as more competent, due to stereotypes that female candidates are less competent than male candidates. At the same time, we expect the male candidates to be rated as more threatening candidate more often, since female faces are often perceived to be more warm and approachable (Chiao et al. 2008; Friedman

and Zebrowitz 1992; Johns and Shephard 2007) and women candidates are often rated as less strong and "tough" than men.

However, we expect that when women *are* perceived as either more competent or more threatening than their male opponents, they ought to be more likely to win an election, due to the unexpected nature of observing a particularly competent- or threatening-looking woman. Expectancy violation (e.g. Bettencourt, Dill, Greathouse, Charlton and Mulholland 1997) is the phenomenon of rating a member of a stereotyped out-group more positively than a similar member of an in-group when the out-group member violates a stereotype in an unexpected and positive way. In other words, because women candidates are generally stereotyped as less competent and threatening when compared to men, we hypothesize that a particularly competent- or threatening-looking woman (or, perhaps, a particularly incompetent- or unthreatening-looking man) will stand out to voters and confer an electoral advantage on these women candidates. Similarly, a particularly attractive-looking man may gain an electoral advantage when running against a woman.

To summarize, our hypotheses are:

(H1) For male-male candidate pairs, candidates judged more *competent* are more likely to be election winners; candidates judged more *threatening* are more likely to be election losers; *attractiveness* judgments will have no predictive value.

(H2) For female-female candidate pairs, candidates judged more *attractive* are more likely to be election winners; *competence* judgments will have less predictive value. It is unclear whether threat judgments will be predictive.

(H3) For mixed-gender candidate pairs, when the female candidate is judged as more *competent* or more *threatening* than the male candidate, the female candidate is more likely to be an election winner; when the male candidate is judged as more *attractive* than the female candidate, the male candidate is more likely to be an election winner.

Using a Laboratory Survey to Assess First Impressions

To test this, we conducted three studies at [reference removed] university with a total of 128 paid undergraduate and graduate students. Stimuli were 182 black-and-white headshot images of real political candidates, paired according to the actual electoral races in which they ran against each other. In order to obtain both mixed gender and female-female races, we have combined data from three separate sessions where subjects viewed about one-third of the different image pairs.[1] The candidates competed in state and local elections, including races for the Oregon State Legislature, Arkansas State House and Senate, Washington State House, and California Assembly. Of the 91 elections included in the dataset, 53 were partisan elections contested by a Republican and a Democrat, 29 were non-partisan, and 9 were primaries.

Though we considered using more image pairs from these state elections, we limit our analysis to these 91 for several reasons. First, we restricted our study to elections in which the first and second place candidates were of the same race and

[1] None of the three datasets on their own contained a large enough sub-sample of either mixed-gender or same-gender pairs to conduct our analysis, since a gender analysis was not the primary objective of any of the studies. By combining all three, we gain sufficient power to test our hypotheses.

ethnicity. Second, we only considered elections for which the candidates' pictures were readily accessible from a voters' guide. Third, and most importantly, we paid careful attention to the quality and consistency of the candidates' images. We were concerned that our participants would be unduly influenced by the quality of the photographs and that this would affect their trait judgments. For instance, candidates with better resources might be able to look more attractive or competent simply by producing higher-quality images. Thus for a pair of photos to be included in our study, we required the photographs to be of similar resolution with approximately central presentation of the candidate, and with the candidates' faces taking up a similar amount of space. Also, since the representation of faces is viewpoint dependent (Desimone et al., 1984; Lee et al., 2006; Straube et al., 2009), we only chose images in which the candidates were frontal-facing. We presented all photos in black-and-white with a uniform, neutral grey background, and resized all photographs to be 110x150 pixels in size. Figure 1 displays a sample pair of images.

Stimuli were presented on an LCD monitor in a computer laboratory using Matlab software. Participants were asked to make judgements about the candidate images for three traits—attractiveness, competence, and threat. Specifically, they were asked to decide which candidate seems "more competent to hold political office", "more attractive to you," and "more likely to act in a physically threatening manner toward you." We followed the TED protocol (Kim et al., 2007), which shows the candidate pictures one at a time rather than contemporaneously. This forces an encoding of the face into working memory for the comparison. Our working assumption is that the two candidates are rarely seen for the first time

simultaneously—especially if one is an incumbent—so this is a (slightly) more valid simulation of how voters would have initially encountered images of the two opponents. Participants were shown each of a pair of pictures for one second each, alternating, with an inter-image interval of one second. Research participants indicated their choice—which of the two images better fit the trait being judged—by pressing the appropriate key. Image pairs continued repeating for up to sixty seconds, and the next image pair was not shown until the subject had chosen a picture from the previous pair. Participants were asked to judge every image pair on one question at a time before moving on to the next assessment block. The order of the blocks, pairs and the two images comprising each pair were counterbalanced among participants.

We expect the brief presentation time to be both sufficient and beneficial. Previous studies have shown that personality traits are perceived after only 50-millisecond exposure to an unfamiliar face (Borkenau et al., 2009), and that judgments of political candidates are not only made rapidly and effortlessly, but are also reliable after as little as 100 milliseconds (Oliviola and Todorov, 2010a). At the same time, using first-impression judgments helps to ensure, to the greatest extent possible, that we are gauging the effect of judgments that are not colored by any other confounding factor. People are generally unaware of the cues they use in these facial evaluations (Rule et al., 2008), so careful deliberation is both unnecessary and potentially disruptive (Levine et al., 1996).

Predicting Election Winners

In order to investigate our hypotheses, we conduct a series of logistic regressions predicting the winner of each election by using the trait judgments of our research participants. Table 1 presents the results of these regression models, in which the dependent variable—the real election winner—is modeled as a function of the attractiveness, competence, and threat judgments made by our subjects. Positive coefficients indicate that the candidate more often judged to exhibit a given trait is more likely to be the real election winner. The first three columns present the baseline models for observations from same–gender candidates pairs only (Model 1), followed by mixed-gender pairs (Model 2), and all pairs combined (Model 3).

Model 1 uses the three judgments to predict election outcomes from only the sixty races with two candidates of the same gender, thereby replicating earlier studies. As in those studies, we find that competence judgments are a strong predictor of electoral success. It is important, however, to note that our study broadens the applicability of this competence result, as we introduce a previously unused set of candidate images from state-level legislature elections. As such, our results here expand upon those previous studies, which were typically done using candidates from upper-ballot elections, such as governors or national legislatures.

At the same time, our findings from same-gender races also support those of a smaller group of studies in showing that candidates who are judged as "more likely to act in a physically threatening manner toward you" are more likely to be real election losers. Finally, in these elections, while candidate appearance is clearly

important, not all aspects of appearance are equally important, as evidenced by the lack of relationship between attractiveness judgments and actual election outcomes.

Model 3 includes judgments from all 91 of the elections and produces results similar to those of Model 1, in that competence judgments predict election winners and threat judgments predict losers. On the other hand, for the competence result, much of the prediction effectiveness stems from the impressive success rate of competence judgments in same-gender pairs. Indeed Model 2, which uses only the mixed-gender pairs, suggests that competence judgments from these pairs are not contributing, a point to which we return in more detail shortly.

How Are Races with Two Female Candidates Different?

As mentioned in our theoretical discussion above, we expect that the gender of the candidates may, in part, affect whether and how competence, attractiveness and threat judgments predict electoral outcomes. Therefore, in the remaining models, we combine the dataset and introduce interactions between the trait judgments and the gender composition of the two participating candidates—i.e., whether the election featured two male candidates (our baseline), two female candidates, or one of each.[2] Using this more nuanced approach we find more convincing support for our hypotheses.

Columns 4 and 5 present the results of our interactive models. Model 4 accounts for the gender composition of the candidate pairs. For male-male pairs, the pattern of results stays the same. For races that include women, however, the effects

[2] We also add dummy variables for mixed-gender and female-only elections; for these, negative coefficients simply mean that the model is better at predicting winners in these elections, while positive indicate that it is less effective.

of competence and attractiveness are straightforward, consistent, and yet quite different than they are for male-only pairs. Most notably, for female-female pairs, *attractiveness is the only positive predictor of electoral success*. While not a significant factor in male-male races, attractiveness appears to confer an advantage when two women appear together on the ballot. On the other hand, in our two-woman races, the coefficient estimate of the interaction term for competence is negative and significant, indicating that female candidates judged as more competent in these races were actually less likely to be election winners.

Expectancy Violation in Mixed-Gender Elections

Model 4 does not give us much insight as to the predictive value of judgments in the mixed-gender pairs. However, the caveat is that we have yet to integrate our hypotheses about conforming to expectations (expectancy violation)—that judgments in these elections will be particularly salient whenever the women appeared more competent, less attractive or more threatening than their male opponents. This relationship can be seen in Model 5 which includes interaction variables for whether the female candidate was chosen over the male candidate in judgments of mixed-gender pairs.

We find support for two of these hypotheses. First, when a woman was judged as more competent than a male opponent, she was much more likely to be the real election winner in that electoral contest. We should note that, unexpectedly, in terms of competence evaluations, there was no statistically significant difference by gender. The female candidate was rated as more competent 49% of the time, which runs counter to gender-based stereotypes. As Model 4 shows, however, these

competence judgments were spot-on, in that whenever the woman was chosen as more competent by our subjects, the voters in that particular election tended to agree.

Second, women who were rated as more threatening than their male opponents were also more likely to be the real election winners, a striking departure from our finding in same-gender races, in which threatening candidates tended to be election losers. Here, as one might expect, the female candidate was considered more threatening than the male candidate only 28% of the time ($p<.001$). Taken together, this signals that women who come across as "tougher" may have an advantage when they run against male opponents.

Next, to illustrate the substantive effect of trait judgments on election outcome predictions, we use the estimates in the full model to calculate the marginal probability that a respondent will correctly predict the winner given a change in her perceptions of a given candidate's traits. Figure 2 presents these estimates. In two-male contests, a candidate judged as more competent has a 7% higher probability of being the actual election winner than when judged as less competent. On the other hand, in contests between two women, the more competent-looking candidate is 5 percentage points less likely to win and, strikingly, in mixed-gender contests, when a female candidate is judged as more competent, she has a 15% higher probability of winning. Attractiveness, on the other hand, is a powerful predictor of electoral success in two-woman races, with the more attractive-looking woman having a 14% higher probability of winning. In mixed-

gender races, women judged as more attractive than their male opponent have a slightly lower (5%) probability of victory.

Discussion and Conclusion

While previous studies have found ratings of competence to positively predict winning actual elections, we have provided evidence that this effect seems to apply best to races contested by two men. For races that include women, the effects of competence, threat and attractiveness are more nuanced and, in some cases, the opposite of what they are for two-man races. Importantly, our findings provide evidence that, not only does a candidate's gender matter, but so does the overall gender composition of a race. The traits that are inferred from a candidate's image seem to vary by gender, and different traits predict electoral success for races than include women than for those that do not. For instance, when a race includes two women, it is the most attractive candidate that has the advantage. Why competence judgments have a muted effect in these elections is not clear. Perhaps, since the facial features that seem to signify competence—eyes set far apart, prominent jawline, etc.—also signify masculinity (Oliviola and Todorov 2010b; Herrick, et al. 2012), it is possible that more competent-looking women are also more masculine-looking. More masculine-looking women have been found to be disadvantaged among voters (Hehman, Carpinella, Johnson, Leitner, and Freeman 2014), so it is possible that the effects of competence seen in this analysis are partially a function of facial masculinity. Interestingly, Carpinella and Johnson (2013) find that facial femininity is associated with higher competence ratings for

female candidates among Democrats/liberals, but with lower competence ratings for Republicans/conservatives.

Of course, the dynamic seems to change somewhat when women candidates run against men. Competence inferences only predicted election outcomes in mixed-gender races when the woman is the one who is perceived to be more competent. Again, this could be an indication that women who look particularly competent benefit from the fact that they violate voters' stereotypic expectations. Similarly, though women were usually seen as less threatening in mixed-gender pairs, those who were seen as more threatening were actually more likely to win their races. While judgments of threat and of competence work against female candidates when they run against other women, then, having features that signal a certain amount of "toughness" may work to a woman's advantage when she runs against a man.

Clearly, there is much work still to be done here and future studies should expand on our analysis in a number of ways. First, if gender matters for the types of traits that predict election outcomes, other social group memberships with visual markers (such as race and ethnicity) probably also make a difference. Including mixed-race/ethnicity pairs in these sorts of studies could yield still different results. Second, because women are subject to stereotypes based on a number of traits other than competence, threat and attractiveness, future studies should include subject ratings of traits such as compassion and leadership in order to see if these assessments also matter for female candidates. Finally, it would be instructive if detailed subject-level data were gathered in future studies of these trait inferences. Question batteries measuring gender based prejudice, for example, would allow us

to determine if certain groups of subjects are inferring different traits from female faces than from male faces.

Similarly, one limitation of this study is that we do not have subject-level data available on partisanship or ideology. It is possible (and perhaps likely, given results from Carpinella and Johnson 2013) that trait perceptions vary by these politically-salient individual characteristics, which would add even more nuance to our understanding of the relationship between gender, automatic trait inferences, and electoral outcomes. It is worth noting, however, that 38 of our 91 races were primary races or otherwise non-partisan in nature, and all candidate images were presented without any reference to party identification. Also, as the purpose of our study is to examine how trait inferences affect real-world elections, the fact that our subjects' ratings of the candidates are predictive of actual election outcomes—even without taking partisanship or ideology into account—suggests that these sorts of trait inferences have a meaningful effect beyond the partisan voting that so profoundly affects general elections.

Despite their limitations, our findings may have important implications, both for scholars studying the role of automatic trait assessments and/or gender in voting behavior, and for women running for office. That gender has such serious consequences for a process as fundamental and automatic as these sorts of trait inferences suggests that many aspects of political behavior are likely similarly influenced by gender. For political scientists, our findings add more evidence to a growing consensus in the literature that the effects of a candidate's gender on political outcomes are contextual and more nuanced than was once thought (e.g. Fox

and Lawless 2014; Dunaway, Lawrence, Rose, an Weber 2013; Bauer 2015a; Ditonto, et al 2014). The role of gender stereotypes, in particular, has been called into question by recent studies such as Brooks 2013, Dolan 2014 and Hayes 2011. However, our results suggest that gender-based stereotypes can and do matter to election outcomes when studied in conjunction with appearance-based trait judgments. It seems that stereotypes can have an effect on women candidates, but not necessarily in the straightforward manner other studies have looked for. Our results suggest that gender scholars may benefit from taking into account other aspects of candidate appearance, as well as the overall gender composition of a race, when examining the effects of gender on vote choice and candidate evaluation, as both seem to change the landscape for women candidates.

In terms of real-world politics, this study may provide valuable information for female candidates and their campaign staff. Women are still seriously underrepresented in American political institutions, so a better understanding of the dynamics at play when voters evaluate female candidates could help to even the playing field in a campaign scenario. While it is perhaps unfortunate that physical appearance has seemingly substantial consequences for candidate appraisal, especially for female candidates, an awareness of the nature of how appearance matters may allow women candidates to more effectively craft an image that takes into account the effects of how women's appearances are evaluated.

References

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

Alley, Thomas R. and Michael R. Cunningham. 1988. Averaged Faces are Attractive but Attractive Faces are not Average. *Psychological Science,* 2(2): 123-125.

Anderson, John R. 1983. *The Architecture of Cognition.* Cambridge: Harvard University Press.

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.

Banducci, Susan, Jeffrey Karp, Michael Thrasher, and Colin Rallings. 2008. Ballot Photographs as Cues in Low-Information Elections. *Political Psychology*, 29(6): 903-917.

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.

Berry, Diane and Leslie Zebrowitz-MacArthur. 1988. What's in a Face? Facial Maturity and the Attribution of Legal Responsibility. *Personality'and'Social' Psychology'Bulletin*, 14(1): 22-23.

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.

Borkenau, Peter, Steffi Brecke, Christine Mottig, Marko Paeleck. 2009. Extraversion is Accurately Perceived After a 50-ms Exposure to a Face. *Journal of Research in Personality*, 43(4): 703-706.

Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes do not Harm Women Candidates.* Princeton, NJ: Princeton University Press.

Budesheim, Thomas L. and Stephen J. DePaola. 1994. Beauty or the beast? The effects of appearance, personality, and issue information on evaluations of political candidates. *Personality and Social Psychology Bulletin*, 20(4), pp.339-348.

Burrell, Barbara. 2008. Political parties, fund-raising, and sex. *Legislative women: Getting*

elected, getting ahead. Boulder, CO: Lynne Rienner.

Bystrom, Dianne G. , Terry A. Robertson, Mary Christine Banwart. 2001. Framing the Fight: An Analysis of Media Coverage of Female and Male Candidates in Primary Races for Governor and US Senate in 2000. *American Behavioral Scientist*, 44(12): 1999-2013.

Carpinella, Colleen and Kerri Johnson. 2013. 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.

Chiao, Joan Y., Tetsuya Iidaka, Heather L. Gordon, Junpei Nogawa, Moshe Bar, Elissa Aminoff, Norihiro Sadato, Nalini Ambady. 2008. Cultural Specificity in Amygdala Response to Fear Faces. *Journal of Cognitive Neuroscience*, 20(12): 2167-2174.

Cook, Elizabeth Adell, Sue Thomas, and Clyde Wilcox, eds. 1994. *The Year of the Woman: Myths and Realities.* Boulder: Westview Press.

Desimone, Robert, Thomas D. Albright, Charles G. Gross, Charles Bruce. 1984. Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque. *Journal of Neuroscience*, 4(8): 88-121.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 285-290.

Ditonto, Tessa M., Allison J. Hamilton, and David P. Redlawsk. 2014. Gender Stereotypes, Information Search, and Voting Behavior in Political Campaigns. *Political Behavior* 36(2): 335-358.

Dolan, Kathleen. 2004. *Voting for Women How the Public Evaluates Women Candidates.* Boulder, CO: Westview Press.

Dolan, Kathleen. 2014. *When does Gender Matter? Women Candidates and Gender Stereotypes In American Elections.* New York: Oxford.

Dunaway, Johanna, Regina G. Lawrence, Melody Rose, and Christopher R. Weber. 2013. Traits versus Issues: How Female Candidates Shape Coverage of Senate and Gubernatorial Races. *Political Research Quarterly* 66(3):715-726.

Efron, Michael G., and E.W.J. Patterson. 1974. Voters Vote Beautifully. *Canadian Journal of Behavioural Science* 6(4): 352-56.

Fox, Richard L., and Jennifer L. Lawless. 2014. Uncovering the Origins of the Gender Gap in Political Ambition. *American Political Science Review* 108(3): 499-519.

Friedman, Heidi and Leslie A. Zebrowitz. 1992. The Contribution of Typical Sex Differences In Facial Maturity to Sex Role Stereotypes. *Personality and Social Psychology Bulletin*, 18(4): 430-438.

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. 2011. When Gender and Party Collide: Stereotyping in Candidate Trait Attribution. *Politics and Gender*, 7(2): 133-165.

Heflick, Nathan A. and Jamie L. Goldenberg. 2009. Objectifying Sarah Palin: Evidence that Objectification Causes Women to be Perceived as Less Competent and Less Fully Human. *Journal of Experimental Social Psychology*, 45(3):598-601.

Heflick, Nathan A., Jamie L. Goldenberg, Douglas P. Cooper, Elisa Puvia. 2011. From Women To Objects: Appearance Focus, Target Gender, and Perceptions of Warmth, Morality, and Competence. *Journal of Experimental Social Psychology*, 47(3): 572-581.

Hehman, Eric, Colleen Carpinella, Kerri Johnson, Jordan B. Leitner, and Jonathan B. Freeman. 2014. Early Processing of Gendered Facial Cues Predicts the Electoral Success of Female Politicans. *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.

Huddy, Leonie and Nayda Terkildsen. 1993. Gender Stereotypes and Perceptions of Male and Female Candidates. *American Journal of Political Science,* 37: 119-147.

Ito, Tiffany A. and Geoffrey R. Urland. 2003. Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals. *Journal of Personality and Social Psychology*, 85(4): 616-626.

Johns, Robert and Mark Shephard. 2007. Gender, Candidate Image and Electoral Preference. *British Journal of Politics & International Relations* 9(2): 434-60.

Kahn, Kim Fridkiin. 1996. *The Political Consquences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns.* New York: Columbia.

Kanazawa, Satoshi and Jody L. Kovar. 2004. Why Beautiful People are More Intelligent. *Intelligence*, 32(3): 227-243.

Kim, H., Adolphs, R., O'Doherty, J. P., & Shimojo, S. 2007. Temporal dissociation of neural processes underlying face preference decisions. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 1823–1825.

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.

Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision Research, 46*, 1901–1910.

Lenz, Gabriel S. and Chappell Lawson. 2011. Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance. *American Journal of Political Science* 55(3): 574-589.

Levine, G. M., Halberstadt, J. B., & Goldstone, R. L. 1996. Reasoning and the weighting of attributes in attitude judgments. *Journal of Personality and Social Psychology*, 70, 230–240.

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, and Caitlin Milazzo. 2014. Pretty Faces, Marginal Races: Predicting Election Outcomes using Trait Assessments of British Parliamentary Candidates. *Electoral Studies*, 34: 177–189.

Mattes, Kyle, Michael Spezio, Hackjin Kim, Alexander Todorov, Ralph Adolph, and R. Michael Alvarez. 2010. Predicting Election Outcomes from Positive and Negative Trait Assessments of Candidate Images. *Political Psychology*, 31(1): 41-58.

Montepare, Joann and Heidi Dobish. 2003. The Contribution of Emotion Perceptions and Their Overgeneralizations to Trait Impressions. *Journal of Nonverbal Behavior*, 27(4): 237-254.

Mouchetant-Rostaing, Yolande and Marie-Helene Giard. 2003. Electrophysiological Correlates of Age and Gender Perception on Human Faces. *Journal of Cognitive Neuroscience,* 15(6): 900-910.

Mouchetant-Rostaing, Yolande, Marie-Helene Giard, Shlomo Bentin, Pierre-Emmanuel Aguera, Jacques Pernier. 2000. Neuropsychological Correlates of Face Gender Processing in Humans. *European Journal of Neuroscience*, 12(1): 303-310.

Naylor, Rebecca Walker. 2007. Nonverbal Cues-Based First Impressions: Impression Formation Through Exposure to Static Images. *Marketing Letters,* 18(3): 165-179.

Olivola, C. Y., Eastwick, P. W., Finkel, E. J., Hortacsu, A., Ariely, D., and Todorov, A. 2009. A picture is worth a thousand inferences: First impressions and mate selection in Internet matchmaking and speed-dating. Unpublished manuscript, Department of Cognitive, Perceptual and Brain Sciences, University College London.

Oliviola, Christopher and Alexander Todorov. 2010a. Elected in 100 Milliseconds: Apperance-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.

Olson, I. R., and C. Marshuetz. 2005. Facial Attractiveness is Appraised in a Glance. *Emotion* 5: 498-502.

Rahn, Wendy, John Aldrich, Eugene Borgida and John Sullivan. 1990. A Social-Cognitive Model of Candidate Appraisal. In *Information and Democratic Processes,* e. John Ferejohn and James Kukilinski. Chicago: The University of Chicago Press.

Riggle, 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.

Rosenberg, Shawn W., Lisa Bohan, Patrick McCafferty, Kevin Harris. 1986. The Image and the Vote: The Effect of Candidate Presentation on Voter Preference. *American Journal of Political Science,* 30(1): 108-127.

Rosenberg, Shawn W., Shulamit Kahn, Thuy Tran. 1991. Creating a Political Image: Shaping Appearance and Manipulating the Vote. *Political Behavior*, 13(4): 345-367.

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.

Rule, Nicholas O. and Nalini Ambady. 2008. The Face of Success: Inferences from Chief Executive Officers' Appearance Predict Company Profits. *Psychological Science*, 19(2): 109-111.

Schubert, James, N. and Margaret Ann Curran. 2001. Stereotyping Effects in Candidate Evaluation: The Interaction of Gender and Attractiveness Bias. Paper prepared for presentation at the Annual Meeting of the Midwestern Political Science Association, Chicago April 19-22.

Sigelman, Lee, Carol K. Sigelman, and Christopher Fowler. 1987. A Bird of a Different Feather? An Experimental Investigation of Physical Attractiveness and the Electability of Female Candidates. *Social Psychology Quarterly,* 50(1): 32-43.

Spezio, Michael L., Antonio Rangel, Ramon Michael Alvarez, John P. O'Doherty, Kyle Mattes, Alexander Todorov, Hackjin Kim, and Ralph Adolphs. 2008. A Neural Basis for the Effect of Candidate Appearance on Election Outcomes. *Social Cognitive and Affective Neuroscience* 3(4): 344-52.

Straube, Thomas, Stephanie Schmidt, Thomas Weiss, Hans-Joachim Mentzel, Wolfgang H.R. Miltner. 2009. Dynamic Activation of the Anterior Cingulate Cortex During Anticipatory Anxiety. *NeuroImage*, 44(3): 975-981.

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, 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, Alexander, Christopher Y. Oliviola, 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., & 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., Mandisotza, A. N., Goren, A., & Hall, C. 2005. Inferences of competence from faces predict election outcomes. *Science*, 308, 1623–1626.

Zebrowitz, L.A. and S. M. McDonald. 1991. The Impact of Litigants' Baby-Facedness and Attractiveness on Adjudications in Small Claims Courts. *Law and Human Behavior*, 15(6): 603-623.

Table 1. Logistic Regression Predicting Election Outcomes

	Same-gender	Mixed-gender	All pairs		
	(1)	(2)	(3)	(4)	(5)
Competence	0.23**	-0.03	0.14*	0.28***	0.28***
	(0.09)	(0.13)	(0.07)	(0.10)	(0.10)
Attractiveness	-0.07	0.02	-0.04	-0.16	-0.16
	(0.10)	(0.12)	(0.08)	(0.12)	(0.12)
Threat	-0.17*	-0.21*	-0.20**	-0.15	-0.15
	(0.09)	(0.12)	(0.08)	(0.10)	(0.10)
Competence x Female-only				-0.51**	-0.51**
				(0.24)	(0.24)
Attractiveness x Female-only				0.74***	0.74***
				(0.27)	(0.27)
Threat x Female-only				-0.12	-0.12
				(0.23)	(0.23)
Competence x Mixed-gender				-0.31*	-0.39**
				(0.17)	(0.18)
Attractiveness x Mixed-gender				0.18	0.24
				(0.17)	(0.17)
Threat x Mixed-gender				-0.06	-0.09
				(0.15)	(0.15)
Female-only				-0.16	-0.16
				(0.19)	(0.19)
Mixed-gender				0.34**	0.10
				(0.16)	(0.19)

Female more competent					0.63***
					(0.12)
Female more attractive					-0.21
					(0.14)
Female more threatening					0.37***
					(0.13)
Constant	0.08	0.46***	0.22***	0.12	0.12
	(0.08)	(0.12)	(0.06)	(0.09)	(0.09)
Pseudo R^2	0.01	0.01	0.01	0.01	0.02
Observations	2397	1255	3652	3652	3652

Notes: Logit coefficients are shown, with robust standard errors in parentheses, clustered by subject.
* p<0.10, ** p<0.05, *** p<0.01

Figure 1: Sample Candidate Image Pair

Figure 2: Marginal differences in likelihood of being an election winner