How Appearance Affects Female vs Male Candidates:

A Potential Mechanism for Women's Underrepresentation in the Candidate Pool

Eden Bjornson

University of California, Los Angeles

Contents

Abstract

Though women's representation in Congress has trended upwards in recent years, US women are still significantly underrepresented in their government. Because women win elections at the same rates as men (Aguiar and Redlin 2014), this analysis focuses on how women are prevented from running for office in the first place. In this paper, I evaluate (1) "How does appearance affect female vs male candidates?" and (2) "Does appearance prevent women from running for office?" I hypothesize appearance is a potential mechanism filtering many women out of the candidate pool. Women without a certain look (what I call a "candidate-worthy appearance") are socialized against running for office from a young age and are continually overlooked as potential candidates throughout their adult lives and careers. I investigated these questions and claims through several phases: (1) an experimental survey to evaluate which traits make male vs female candidates appealing to voters, (2) mock elections to test non-candidate worthy appearing men vs women's chances of winning an election, and (3) a survey of college students about others' appearance-based perceptions of them and whether they've considered running for office. Though I do not successfully pin down what makes up a candidate-worthy appearance, I find that such an appearance exists and is more salient for female candidates than male candidates. Thus, this research strengthens existing literature on gender and appearance in politics and provides a novel explanation for women's underrepresentation in the candidate pool and government.

Keywords: representation; appearance; gender; women; elections; candidate-worthy appearance

1 Introduction

In order to understand the state of American politics, we must understand who gets to office and how they get there. The United States of America's government today is overwhelmingly unrepresentative of its people. One of many groups underrepresented in government is women. Only 28% of lawmakers in the 119th Congress are women (29% in the House of Representatives

and 25% in the Senate). While this is a substantial increase compared to years past – it is a 44% increase from ten years ago (Jackson 2025) – these numbers are still far from reaching gender parity.

However, when women run, they win at equal rates as men. Women are therefore underrepresented because they are not running enough (Aguiar and Redlin 2014). This is why it is important to understand the entire pathway to government. On the surface, equal win rates indicate equal opportunity, but given the persistent underrepresentation of women in office, we must examine how women are systematically discouraged or prevented from running. Furthermore, to fully grasp the situation, we must question whether women who do not choose to run could win if they ran. Nothing happens in a vacuum, so when choosing potential mechanisms to investigate, it is logical to look at ways women are oppressed in society in general. Given the existence of harsh beauty standards for women, this study analyzes appearance as a potential factor preventing women from running for office. Thus, this project poses the following main questions: "How does appearance affect female vs male candidates?" and "Does appearance prevent women from running for office?"

2 Literature Review

Previous research into candidate appearance has uncovered discouraging results about how voters make decisions. Several studies have correctly predicted outcomes of real elections, or shown a clear winner in hypothetical elections, by briefly exposing subjects to images of candidates and asking them to vote for a candidate (Olivola and Todorov 2010). The studies that use real candidates ensure decisions are made purely on appearance by excluding instances where the subjects recognized any candidate; significant results have been found for several different types of US races, including Senate races (Todorov et al 2005) and gubernatorial races (Ballew and Todorov 2007). The senate research exposed subjects to candidate images for one second, and the gubernatorial research exposed subjects for a mere 100 milliseconds or 250 milliseconds in two of its three conditions (Todorov et al 2005; Ballew and Todorov 2007). Any scholar concerned with voters' decision making skills should be alarmed that such dramatically short exposure times yield statistically significant election predictions.

These results still hold up when subjects are given more time to deliberate and when given

more information. In the gubernatorial research discussed above, there was a third condition where subjects were given unlimited response time; though subjects took about twice as long to answer, their accuracy was about the same as the timed conditions (Ballew and Todorov 2007). Manipulating candidates' appearance can also affect electoral results, potentially changing the outcome of elections, even with information on candidate policy positions held constant (Rosenberg, Kahn, and Tran 1991). The implication that physical appearance can be more important to election outcomes than actual policy is frightening. If this truly is the case, voters make illogical decisions that do not reflect their best interests, so the government likely fails to effectively represent the people.

The literature has observed the power of appearance in elections with real races from other countries as well, including Australia (Little et al 2007), France (Antonakis and Dalgas 2009), Ireland (Buckley, Collins, and Reidy 2007), Italy (Castelli et al 2009), New Zealand (Little et al 2007), and the UK (Little et al 2007; Banducci et al 2008). Several of these examples examine races that include photographs of candidates on the real ballot (Banducci et al 2008; Buckley, Collins, and Reidy 2007). These cases particularly emphasize the importance of these findings because in low-information races, these images may be the only information many voters have. Therefore, studies that simply place two images next to each other and ask subjects to pick one may actually represent certain races more accurately than initially envisioned.

Even in high profile races, appearance is likely to still make a difference, especially in the age of television. For example, studies have confirmed watching versus just listening to the infamous 1960 Nixon-Kennedy debate lead to different evaluations of the candidates, both at the time of the debate (Kraus 1996), and in more recent experiments performed over 40 years after the debate (Druckman 2003). One study examining both US senate and gubernatorial races adds to this literature, finding the effect of television is more dramatic for less informed individuals (Lenz and Lawson 2011). Thus, educating the public more about political races may mitigate some of these appearance related concerns.

Current literature on the gender gap in politics largely focuses on which women choose to run. Though it is well established that when women run, they win at the same rates as men (Aguiar and Redlin 2014; Sanbonmatsu 2006), women are still far underrepresented in the US government for many complex reasons. First, the equal win rate does not apply to both parties, as Republican women win at much lower rates than their Democratic counterparts (Thomsen

2019). Women are also recruited by political elites to run for office far less often than men (Fox and Lawless 2010), perhaps because party leaders do not believe women can win (Sanbonmatsu 2006). Women are less likely to enter careers that would traditionally qualify them to run for office (Thomsen and King 2020), and as early as high school and college, girls are already socialized to be less likely to consider a political career than boys (Fox and Lawless 2014). Societal norms around gender also often face women with impossible tasks, complicating a bid for office. For example, although voters equally reward male and female candidates for having families, having a family puts far more strain on women than men (Teele, Kalla, and Rosenbluth 2018). Logically, this phenomenon likely applies to other gender-based double standards, such as the stricter expectations on women's appearances. Because it is so much harder for women to reach Congress than men, the women who do are better at their jobs on average than their male counterparts. Previous research determined female congressmembers secured more funding and sponsored/cosponsored more bills than male congressmembers, even when accounting for confounding variables such as party and district-specific traits (Anzia and Berry 2011). Of course these are not the only potential measures of a congressmember's effectiveness, but they are helpful for research purposes. High bars for qualification and competence may not be the only bars women must reach to assume office. There may be other high bars unrelated to legislator quality such as appearance.

I propose the issues of appearance affecting election outcomes and women's underrepresentation in government are linked. Though research suggests the most highly effective women do make it to office (Anzia and Berry 2011), traits unrelated to candidate qualification, like appearance, may filter out other highly qualified women from running. Outside of politics, research has established that women are deeply affected by beauty standards (Arevalo-Flores 2017; Santoniccolo et al 2023). The political science literature has studied which aspects of female candidates' appearance make them more favorable (Rosenberg, Kahn, and Tran 1991) and established that female candidates are judged differently than male candidates on appearance (Johns and Shepard 2007; Ditonto and Mattes 2018). The idea that how much a candidate's appearance matters may differ by gender has also been implicitly recognized by some studies (Todorov et al 2005). Overall however, the research into how appearance uniquely affects female candidates is limited. Thus, I will fill this gap by bolstering the existing research and extending knowledge in the field by investigating whether appearance prevents women from running.

3 Argument

Since the standards candidates are measured on may not be the same as conventional attractiveness, I will refer to appearances that voters view as more candidate-like as "candidate-worthy appearances". For instance, highly attractive women are often not taken seriously because of their appearance, so attractiveness and candidate-worthiness must not be the same thing. I argue female candidates likely face harsher judgements of their appearance than male candidates, which prevents many women who do not have candidate-worthy appearances from running for office.

The way appearance prevents female candidacy could be related to any of the literature's established barriers, or manifest in another way altogether. For instance, women who do not have a political "look" may be more strongly socialized against running for office than men who do not have the "look". Party elites recruiting future candidates may, consciously or not, weigh women's appearance more strongly than men's appearance, causing them to recruit fewer women. Perhaps women even consciously choose not to run. They may feel that they do not "look the part," to win, or do not want to endure a campaign cycle viciously dissecting their appearance.

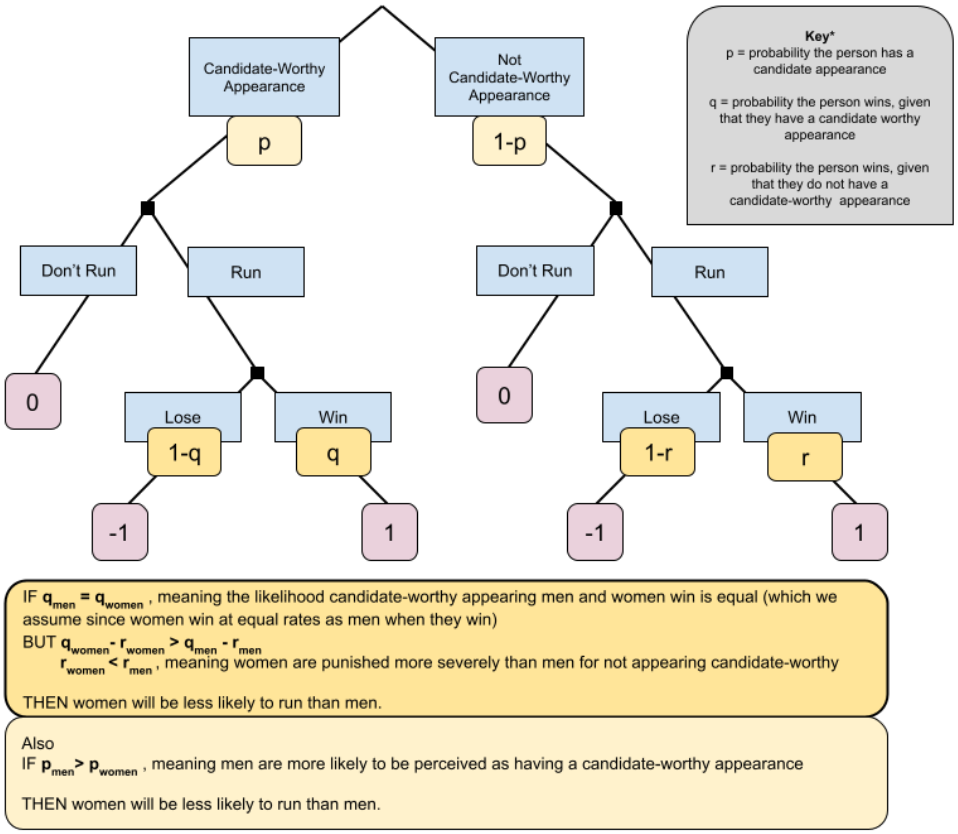
This argument poses a significant contribution to both the literature on the gender gap in government and the importance of appearance in elections. Both literatures are separately quite robust, but there is very little research currently examining their potential link. My argument serves as the bridge between these two literatures. With context from the other, each literature will grow stronger and more thorough. The argument is also relevant to any scholar who cares about representation and how well governments truly represent their people. My argument may apply to other underrepresented groups as well, so it may pave the way for future research into which groups are disproportionately affected by appearance in politics.

The specific causal mechanism of how women are punished more for their appearance than men is more complicated than simply "women with less candidate-worthy appearances do not run", as there are several potential stages preventing these women from running. Visualization 1 presents a potential candidate's choice to run or not as a game theoretic model. There are several nature nodes that the individual cannot change, and only one actual decision: run or not. For the purposes of this diagram, the individual does not get to choose if they have a candidate-worthy appearance or not, and they do not get to ultimately choose whether they

win or lose. While there are aspects of appearance that can be manipulated which may make someone appear more or less candidate-worthy (Rosenburg, Kahn, and Tran 1991), many aspects can not be easily changed, and this diagram serves to break down a very nuanced reality into a more simplified, understandable model. That is also why I chose to present appearance as a binary variable in this diagram. Appearance is certainly more complicated than all or nothing, but this diagram simply serves to illustrate an idea.

There are two places where appearance could affect a decision to run or not: the probability the individual has a candidate-worthy appearance (p) and the probability the person will win if they run given their appearance (q and r). Since the literature has established that appearance does affect election outcomes (Olivola and Todorov 2010), we will assume the probability someone wins if they have a candidate-worthy appearance (q) is higher than if they do not (r). If the probability that a man has a candidate-worthy appearance is higher than the probability that a woman does, then less women are likely to run than men. This would essentially mean that more men already appear candidate-worthy [1], or it is easier for them to look candidate-worthy, so less women would fall into the candidate-worthy appearance category. I also hypothesize that women are punished more than men for not appearing candidate-worthy. Essentially, once a woman is placed in the non-candidate-worthy appearance category, she faces much lower odds of winning than a man in that category, so a non-candidate-worthy appearing woman is less likely to choose to run than a non-candidate-worthy appearing man. In the diagram, this would mean the probability a woman wins if she does not have a candidate-worthy appearance (r_{woman}) is lower than the probability a man wins if he does not appear candidate-worthy (r_{man}).

[1] In my study, when measuring candidate-worthy appearance, men and women are evaluated separately, and rated on their gender's respective scale. When I say perhaps more men already look candidate-worthy, I mean based on the criteria for men, with women having different criteria. In this way, appearance can be separated from gender as a factor.



Visualization 1: Decision Tree when deciding to run for office or not

In my discussion of this diagram thus far, I've framed it as a conscious decision by a prospective candidate to run or not, but this causal model is more flexible than that. For example, political elites may evaluate individuals they consider recruiting to run for office this way. If you replace the labels "Don't Run" and "Run" with "Don't Recruit" and "Recruit" in Visualization 1, the model now applies to recruiters, not prospective candidates. For anyone's decision making purposes (potential candidate or recruiter alike), the probabilities in the diagram effectively represent what the individual believes them to be, not the actual probabilities, since the true probabilities are unknown. Thus, assumptions about appearance and unconscious attitudes are relevant. For example, a recruiter may not consciously realize the reason they are writing off a female candidate is because of her appearance, but really they are unconsciously placing her in the non-candidate-worthy appearance category, and thus mentally assigning her a lower probability of victory. Similarly, a highly qualified woman may never seriously consider running for office if she has always been socialized into believing she does not have a candidate-worthy appearance and that she therefore does not look like someone who would have good chances of

winning. It is logical to believe women would face harsher judgements than men (from recruiters or from socialization in society) since women face much harsher appearance standards in general (Santoniccolo et al 2023).

Some scholars may criticize my argument that it is harder for women to appear candidate-worthy than men, or that, by default, men tend to look more candidate-worthy. They may argue that gender itself affects perceptions of the potential candidates, not appearance. Though individuals are certainly also judged on gender itself, I evaluate men and women's appearances on criteria determined by their respective genders and control for gender as much as possible. Furthermore, if men and women are equally rewarded for appearing candidate-worthy, women may still face more of a burden to achieve that. Although male and female candidates are rewarded equally for having families, women are expected to put in much more effort raising children than men (Teele, Kalla, and Rosenbluth 2018). A similar double standard may apply for appearance. Generally speaking, there are more aspects of appearance women pay attention to than men. Makeup, hair, and clothing styles are all typically more diverse for women than men, so achieving the correct combination may be more difficult for women than men. Therefore, just as there is a higher cost for motherhood than fatherhood, there is likely a higher cost to appear candidate-worthy for women than for men. Thus, it is harder for women to appear candidate-worthy, so less women than men will reach their respective benchmarks. More men than women in the general population can appear candidate-worthy without the measure for appearance effectively only measuring gender.

This argument may also be criticized for being too complicated to test or falsify. Appearance is complex, so it is difficult to empirically evaluate, there are many ways it affects elections, and it is hard to isolate. However, I have integrated several different phases to test my hypotheses in order to ameliorate this issue. I further argue that just because something may be difficult to test does not mean it is unimportant to understand, so scholars should still attempt to answer difficult questions like these.

4 Concept Specification and Operationalization

As previously mentioned, I use the term "candidate-worthy" appearance instead of attractiveness because these are distinct concepts. For the purposes of my studies, individuals with a more

candidate-worthy appearance should be more likely to run (and win) than individuals with a less candidate-worthy appearance.

I measure appearance with a similar method as Rosenberg, Kahn, and Tran (1991). These researchers first collected headshot photographs of women and coded them based on specific traits (ex. eye shape, hairline, age group). Then they gave respondents a packet of 15 images and asked them to rank the images on scales of "political demeanor", "competence", "trustworthiness", and "attractiveness". With these results, they were able to determine which traits made women seem more candidate-worthy (based on the political demeanor scale) and what aspects of appearance correlated with competence, trustworthiness, and attractiveness (traits that could fuel the political demeanor opinion[2]). I conducted a similar test in phase one of my research design with some key differences. First, I conducted the experiment for both men and women (testing men and women separately). Second, rather than asking respondents to rank the images relative to each other on each scale, I asked respondents to rate each candidate on feelings thermometers from zero to 100 to more easily compare results between different groups of images provided to respondents. Third, I used MediaPipe's facial landmark software to calculate the ratios of several facial features and code those ratios as their traits, and I used a Vision Transformer (ViT) age classification model to estimate age group[3]. I accessed these models, ran my images through them, and calculated the proper ratios through a Python script.

Though quantifying an abstract concept like appearance is difficult, I believe my method is reasonably valid. By quantifying the traits through measurements and ratios, I can ensure differences in traits between different images are captured. With this numeric data, I can then easily perform multivariate linear regression to see which traits significantly contribute to the model. Furthermore, I can account for potential correlation between traits through Principal Component Regression. If Principal Component Regression yields a significant model, I can determine which traits are effective by examining their contribution to each principal component used. I also accounted for race, which could be a confounding variable. The male and female photo pools have equal proportions of different races, roughly according to the current US racial

[2]The measures of "competence", "trustworthiness", and "attractiveness" are not key to my main research question, but I plan to include them to answer a subsidiary question of "why do people associate certain traits with a stronger political demeanor?"

[3]MediaPipe is a reputable facial mapping tool developed by Google. It has been used in past research and is roughly 86% accurate (Al-Nuimi and Mohammed 2021; Kumar Rao B et al. 2023). The age estimator tool I used is called the nateraw/vit-age-classifier which I got from HuggingFace.co, which is a platform to publish and use computational and AI tools.

demographic breakdown (Tian et al. 2025). Furthermore, respondents only saw either men or women for the entirety of the survey to prevent respondents from being influenced by the gender of the other images they evaluate.

5 Hypotheses

H_1 : Women are punished more for not having a candidate-worthy appearance than men, so fewer women run for office than men.

H_2 : If women without candidate-worthy appearances did run, they would face lower odds of winning than the candidate-worthy appearing women who do run.

H_3: Women with less candidate-worthy appearances are socialized against running for office.

6 Case Selection, Selection Rationale, and Data Sources/Availability

I collected a large number of stock photos of non-candidates to evaluate what constitutes a candidate-worthy appearance to represent the general population. I collected these photos through Freepik, specifically with a Premium account. I selected real races for my mock elections based on cases where candidate race and gender match each other and the real election was very close, and I chose cases outside of California (to minimize the number of subjects, all of whom attend UCLA, are filtered out of the results for recognizing candidates). My data was collected through my original experimental survey, designed on Qualtrics. This ensures the data best fits my research questions as my experiment is designed around this specific project. Survey respondents are UCLA students whose professors agreed to send my survey out to their class.

7 Research Design

In phase one, I determine if certain traits make up a candidate-worthy appearance. This was done by filtering respondents into either the male photo group or the female photo group, showing them five randomly selected photos from the respective photo pool, and asking them to rate each

image on a scale from zero to 100 for trustworthiness, competence, attractiveness, and political demeanor. The political demeanor question does not use the phrase "political demeanor," rather it asks respondents how comfortable they would be with the individual representing them in the US House of Representatives. There were 400 total photos, 200 in the male photo group and 200 in the female photo group. The races and racialized ethnicities of each stock photo person follow roughly the same breakdown as the US population. Each gender has 120 White photos (60%), 40 Latino/a/Hispanic photos (20%), 26 Black photos (13%), and 14 Asian photos (7%) (Tian et al. 2025). All photos have the background removed, show headshots with smiling faces looking at the camera, and have similar casual office style clothing.

The next part of my survey conducts several mock elections. Each respondent saw four total races: two real US House races from 2024 and two fictional races I fabricated from stock photos. In each race, backgrounds are removed, and both candidates are smiling and wearing similar grey business attire. There are eight total races to analyze, since there will be four with female candidates and four with male candidates. Candidate race/ethnicity is held constant in each race. All races showed two white candidates as there were no real life examples fitting my criteria with two non-white candidates and I wanted to remain consistent between the real and fictional races. I first asked respondents which candidate they would vote for. I then compared the margins of victory for the real races to the stock photo races. I find that since there is a bigger difference in margin of victory for the stock photo women compared to the real women than between the stock and real men, then the less candidate-worthy appearing women are more severely electorally punished than their male counterparts. I also asked respondents to rate each candidate on the same scales of trustworthiness, competence, and attractiveness used in phase one to evaluate what traits correspond to these differences.

The final phase of my survey tests political socialization as a potential mechanism disproportionately preventing non-candidate-worthy appearing women from running compared to their male counterparts. This has a similar design as Fox and Lawless (2014)'s study which identified socialization against political careers from a young age as a reason less women run for office than men. My version however incorporates appearance as well to determine if less candidate-worthy appearing women are more strongly socialized against running than their male peers. I surveyed college students and asked them to rate how they believe others would perceive them on the scales of "political demeanor", "competence", "trustworthiness", and "attractiveness" based on

appearance alone. I then asked if they would ever consider running for office, and whether they think they could win if they ran. Specifically, I asked them about local offices, state offices, the US House of Representatives, the US Senate, and the US Presidency. I predicted the men would more consistently consider running and believe others will rate them highly compared to the women. I further predicted that men would be more likely to believe they could win, regardless of whether they said they would run or not.

If I had found significant traits in phase one, I would have used this criteria to rate real candidates. I would have rated each candidate photo based on the criteria for a candidate-worthy appearance, with a high score indicating a more candidate-worthy appearance and a low score indicating a less candidate-worthy appearance. Then I would have taken the average rating for male candidates and female candidates and evaluate whether the average score for female candidates is higher than the average score for male candidates. Unfortunately, I did not find significant enough results in phase one to warrant this investigation. To my knowledge, this phase's design would have been the most distinct from the existing literature on appearance. This idea could still be employed in future research. I would expect the women would have a higher average score than men. If this were indeed the case, then the women who run would tend to have more candidate-worthy appearances than the men, indicating that women without candidate-worthy appearances are disproportionately filtered out of the candidate pool. This would have supported my hypothesis that women are punished for not having a candidate-worthy appearance, which manifests as less women running for office.

8 Results and Analysis

Data collection for this survey began on October 22, 2025 and ended on December 19, 2025. The survey respondent pool consists of 701 UCLA students who consented to the study, primarily from either Political Science or Statistics courses. This is notably not a sample representative of the US population, but unfortunately this study had limited resources. 453 respondents were female, 226 were male, 5 were non-binary, and 4 were gender-fluid. Since there is not an even split between male and female respondents, I evaluated the data for female vs male respondents separately when appropriate. For the race/ethnicity questions, respondents were allowed to select all categories that applied to them. 249 respondents identified as Asian, 242

as White, 207 as Latino, 52 as Middle Eastern or Northern African (MENA), 46 as Black, 8 as Native American, and 7 as Native Hawaiian or Pacific Islander. All demographic questions were optional, and the race/ethnicity questions allowed multiple selections, so neither the gender nor race/ethnicity information adds up to the exact 701.

I performed some diagnostic tests to ensure that the data was good quality. For any online survey, there is a risk that respondents will simply click random buttons to speed through the survey. Thus, I checked the distribution of how long respondents spent on the survey. The median time was around eight minutes, with the first quartile around five minutes and the third quartile around thirteen and a half minutes. These times are in line with the roughly five to fifteen minutes I expected the survey to take for respondents. The survey tool, Qualtrics, also collected whether respondents finished the survey. About 73.9% of respondents completed the survey, which is a good yield. Since around 16% of respondents spent under a minute on the survey, I then checked how many of those respondents finished the survey, as those who finished in under two minutes likely did not take the survey seriously. Only two respondents who finished the survey did so in under two minutes. This indicates that the data should be of good quality, as those who did not spend long on the survey simply did not finish it. The answers they provided to questions early on in the survey are likely still meaningful. Finally, I confirmed that the CAPTCHA scores collected by Qualtrics overwhelmingly indicated human rather than bot activity. Only two respondents had a CAPTCHA score under 0.5, the typical threshold below which bot activity raises concern. Thus, I can comfortably conclude that my data consists of reasonably thoughtful responses by humans and is therefore suitable to analyze.

8.1 Phase 1

I conducted standard two-sample, Welch t-Tests to determine if men and women were rated differently on each scale (trustworthiness, competence, attractiveness, political demeanor). The results are in Figures 1a-c below. Surprisingly, women were rated statistically significantly higher on average on every scale. Since the sample is skewed female, I checked if these results differed by respondent gender. Though female respondents tend to rate men a bit lower on average compared to male respondents, male respondents still rated women statistically significantly higher on every scale compared to men. The biggest difference between male and female respondents lies in the attractiveness scale: woman respondents rated photos of women far more attractive than

photos of men on average (27.156 points for male photos compared to 54.296 points for female photos), while man respondents demonstrated a far smaller difference in attractiveness ratings by gender (39.280 points for male photos, 45.865 points for female photos).

Figure 1a. Welch t-Tests for Phase 1 Qualitites

All Respondents

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Male Mean	Female Mean
Trustworthiness	-11.026	3,213.237	8.9100e-28	-10.965	-7.654	51.307	60.617
Competence	-10.537	3,202.482	1.5200e-25	-10.209	-7.006	55.283	63.890
Attractiveness	-22.680	3,229.150	8.6400e-106	-22.036	-18.529	31.167	51.449
Political Demeanor	-11.572	3,209.979	2.2800e-30	-12.134	-8.618	48.530	58.906

Figure 1b. Welch t-Tests for Phase 1 Qualitites

Female Respondents Only

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Male Mean	Female Mean
Trustworthiness	-10.497	2,137.660	3.5900e-25	-12.871	-8.819	48.863	59.708
Competence	-10.348	2,136.962	1.6000e-24	-12.128	-8.264	53.895	64.091
Attractiveness	-25.449	2,130.701	5.7200e-125	-29.231	-25.048	27.156	54.296
Political Demeanor	-10.479	2,136.659	4.3400e-25	-13.804	-9.452	47.168	58.796

Figure 1c. Welch t-Tests for Phase 1 Qualitites

Male Respondents Only

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Male Mean	Female Mean
Trustworthiness	-4.085	1,009.081	4.7600e-05	-8.680	-3.047	56.625	62.488
Competence	-3.503	1,000.860	4.8000e-04	-7.859	-2.215	58.415	63.452
Attractiveness	-4.287	1,040.661	1.9800e-05	-9.600	-3.571	39.280	45.865
Political Demeanor	-5.109	1,017.986	3.8700e-07	-10.743	-4.781	51.573	59.335

I then conducted a standard multivariate linear regression to determine if trustworthiness, competence, attractiveness, and photo gender can predict political demeanor. This model met the assumptions required for a linear model, and all qualities contributed statistically significantly to the model. Detailed results are in Figure 13 in the appendix. However, the model has an R-squared value of only 0.3539. Thus, trustworthiness, competence, attractiveness, and photo gender together explain well under half of the variation in political demeanor scores. This demonstrates a surprisingly very weak relationship between how trustworthiness, competence, attractiveness, and gender contribute to comfort with a political candidate. This indicates there are likely other inferred qualities based on appearance relevant to judgments of political

demeanor.

Figure 2: Linear Model - Qualities by Trait

	Dependent variable:			
	Trustworthiness	Competence	Attractiveness	Political Demeanor
	(1)	(2)	(3)	(4)
Estimated Age	−0.3881	−0.4146	−3.4253***	−0.4899
	(0.3413)	(0.3339)	(0.3605)	(0.3663)
Eye Shape	−0.0553	−0.3934	−0.0176	−0.9468
	(0.5878)	(0.5750)	(0.6208)	(0.6307)
Eye Angle	0.1630	0.2131*	−0.0136	0.0657
	(0.1164)	(0.1138)	(0.1229)	(0.1249)
Eye Distance	13.9866***	10.7758**	5.6316	20.0923***
	(5.1129)	(5.0013)	(5.3999)	(5.4862)
Eyebrow Thickness	51.7413***	26.3819***	−7.2765	33.4376***
	(8.9991)	(8.8026)	(9.5041)	(9.6560)
Face Ratio	11.6902	4.7202	23.1909**	0.5475
	(10.0907)	(9.8703)	(10.6570)	(10.8277)
Lip Thickness	0.5199***	0.4749***	0.4595**	0.3021
	(0.1732)	(0.1695)	(0.1830)	(0.1859)
Nose Ratio	−26.9149***	−10.1336*	−5.7893	−10.8017*
	(5.5622)	(5.4407)	(5.8743)	(5.9683)
Gender	8.1367***	7.4596***	18.1074***	8.9492***
	(0.9624)	(0.9414)	(1.0164)	(1.0328)
Constant	10.1587	16.1409	16.5029	18.1162
	(18.2586)	(17.8599)	(19.2832)	(19.5917)
Observations	3,232	3,232	3,232	3,231
R^2	0.0690	0.0451	0.1721	0.0528
Adjusted R^2	0.0664	0.0425	0.1697	0.0502
Residual Std. Error	23.6095	23.0939	24.9344	25.3327
F Statistic	26.5518***	16.9196***	74.3999***	19.9588***
Note:				*p<0.1; **p<0.05; ***p<0.01

When evaluating the physical traits, I first conducted multiple linear regression to determine if any traits correspond with evaluations of trustworthiness, competence, attractiveness, or political demeanor, displayed in Figure 2. All traits, except for estimated age, were numerical

since they were calculated through measurements and ratios of facial features with MediaPipe's facial landmark detection. The model included estimated age, eye shape (how round vs almond shaped the eyes are), eye angle (how drastically upturned or downturned the eyes are), eye distance (distance of the eyes from each other in proportion to eye size), eyebrow thickness, face ratio (face length to face width), lip thickness, nose ratio (nose length to nose width), and photo gender as predictors. Age was categorical, treated as age groups, as it is difficult to distinguish between a one or two year age difference, and determined from a Vision Transformer (ViT) model accessed from HuggingFace.co. I made four separate models, one for trustworthiness, one for competence, one for attractiveness, and one for political demeanor as the respective outcome variables. Though many traits did contribute statistically significantly to many models, none of the models had particularly high explanatory power. The R-squared for trustworthiness was only 0.0690, 0.0451 for competence, 0.1721 for attractiveness, and 0.0528 for political demeanor. I also conducted principal component regression for each scale in order to account for correlation between traits, as displayed in Figure 15 in the appendix. While all individual principal components contributed statistically significantly to most models, these models still had poor R-squared values, each slightly smaller than the R-squared for their respective standard multivariate linear model, meaning they were weak explanatory models. This of course makes sense since Principal Component Analysis does not add any information, it just reduces model dimensionality.

To gauge which variables related to demographic information (of both the photos judged and the survey respondents) may impact the model, I created additional linear models with all of those variables included. These models have the same physical traits included, but include demographic information about both the photo and the respondent as photos of different race/ethnicities or genders may be judged differently, and respondents may judge the photos differently depending on their gender or racial/ethnic background. Complete results are in Figure 16a in the appendix. Since there are many predictors in this model, I confirmed that there are no collinear predictors by calculating the variance inflation factor (VIF) of each predictor. The largest VIF was only 2.745, which is far below the benchmark of 10 that indicates serious collinearity. Thus, we can interpret these coefficients to represent the effect of their respective predictors. For the most part, these models indicate that the race/ethnicity and gender of the photos can statistically significantly contribute to the models, but the demographic informa-

tion for the respondents statistically significantly contributes to the models far less often. This is not a direct indication that different physical traits translate to higher ratings for different racial/ethnic and gender groups of the photos. Rather, this model just illustrates that the gender and racial/ethnic group of the photo itself contributes to the ratings of trustworthiness, competence, attractiveness, and political demeanor. Separate models for different racial/ethnic and gender groups for the photos can determine if different traits are important to different demographic groups. Thus, I conducted additional multiple linear regressions for each gender in the photo pool (male and female) as well as each racial/ethnic group in the photo pool (White, Black, Latino, and Asian).

When examining how these traits contribute to trustworthiness, competence, attractiveness, and political demeanor for each gender, the R-squared values of all models were tiny, even smaller than in their respective full-data multiple linear regression and Principal Component Regression models. This is because most of the explanatory power of the original models came from the gender predictor itself. Once gender is held constant by only including one gender in the model, the predictive power of gender is lost, leaving only the traits themselves. Since the strength of these models overall is very small, it is not useful that the models are technically statistically significant. Thus, when accounting for gender, I still cannot determine what traits contribute to a candidate-worthy appearance to a meaningful degree. Detailed results are shown in Figures 16b and 16c in the appendix.

Figure 3a: White Photos - Qualities by Trait Linear Models Significance

Model	R-Squared	Adjusted R-Squared	P Value
Trustworthiness	0.0576	0.0534	0
Competence	0.0528	0.0486	0
Attractiveness	0.1837	0.1800	0
Political Demeanor	0.0480	0.0438	0

Figure 3b: Black Photos - Qualities by Trait Linear Models Significance

Model	R-Squared	Adjusted R-Squared	P Value
Trustworthiness	0.0425	0.0205	0.0456
Competence	0.0411	0.0192	0.0544
Attractiveness	0.1169	0.0967	0.0000
Political Demeanor	0.0358	0.0138	0.1066

Figure 3c: Latino Photos - Qualities by Trait Linear Models Significance

Model	R-Squared	Adjusted R-Squared	P Value
Trustworthiness	0.0521	0.0377	2e-04
Competence	0.0565	0.0422	1e-04
Attractiveness	0.2361	0.2245	0e+00
Political Demeanor	0.0714	0.0574	0e+00

Figure 3d: Asian Photos - Qualities by Trait Linear Models Significance

Model	R-Squared	Adjusted R-Squared	P Value
Trustworthiness	0.0987	0.0587	0.0108
Competence	0.0978	0.0578	0.0116
Attractiveness	0.2511	0.2179	0.0000
Political Demeanor	0.0620	0.0204	0.1531

When repeating this process for the different races in the photo pools, I still found relatively small R-squared values. Figures 3a-d above summarize the significance and strength of each

race/ethnicity model with the R-squared, Adjusted R-squared, and p-value for each model. Again, although almost all p-values indicate a statistically significant model, the models are too weak to reliably predict trustworthiness, competence, attractiveness, or political demeanor based on physical traits. Interestingly, while still small, the R-squared values for the attractiveness measure notably increased compared to the models with all races/ethnicities pooled. The largest attractiveness R-squared was for Asian photos at 0.2511, next Latino photos at 0.2361, White photos at 0.1837, and Black Photos at 0.1169. These are still generally considered small R-squared values, but the models for the Asian and Latino photos explain roughly a quarter of the variation in attractiveness scores. While not strong enough to reliably use for prediction purposes, the models for Asian and Latino attractiveness specifically appear to indicate which traits contribute to a more attractive appearance for Asians and Latinos. Figure 4 below displays the attractiveness models for each race by physical trait. This indicates that for all races, being younger contributes to a more attractive appearance. For Latinos, eyes placed slightly further apart and thicker eyebrows also contribute to a more attractive rating. For Asians, less dramatically angled eyes and wider (rather than longer) noses led to a more attractive rating.

Figure 4: Linear Model - Attractiveness by Trait, Photo Race Breakdown

	White	Black	Latino	Asian
	(1)	(2)	(3)	(4)
Estimated Age	−2.9779***	−3.2070**	−5.0606***	−3.6725**
	(0.4378)	(1.4108)	(0.9492)	(1.7286)
Eye Shape	−1.2882	4.9736**	0.9906	−0.9863
	(1.0356)	(2.3440)	(1.4875)	(1.4955)
Eye Angle	−0.0192	0.2562	−0.0523	−1.7850***
	(0.1557)	(0.4074)	(0.2681)	(0.6325)
Eye Distance	−3.8913	−21.0709	47.8928***	4.9698
	(6.3622)	(19.1058)	(16.7532)	(35.3985)
Eyebrow Thickness	−59.9935***	28.0169	44.6587**	34.0684
	(12.5497)	(45.8530)	(21.9548)	(58.1453)
Face Ratio	20.9739	−39.4343	43.3689	20.1445
	(13.6310)	(35.9705)	(27.3640)	(44.4512)
Lip Thickness	0.3272	−0.3139	0.0627	−0.6738
	(0.2968)	(0.4835)	(0.4223)	(1.0074)
Nose Ratio	−5.6575	39.4623*	19.7696	62.9223***
	(8.2927)	(22.3672)	(15.1906)	(24.0510)
Photo Gender	19.8481***	14.5788***	16.7409***	26.5623***
	(1.3213)	(3.0430)	(2.4745)	(4.0700)
Constant	51.3965**	43.1815	−91.7655*	106.3007
	(25.1545)	(66.6419)	(48.4203)	(103.9481)
Observations	2,010	403	606	213
R^2	0.1837	0.1169	0.2361	0.2511
Adjusted R^2	0.1800	0.0967	0.2245	0.2179
Residual Std. Error	24.2665	26.5116	24.6179	24.2592
F Statistic	50.0018***	5.7823***	20.4654***	7.5646***
Note:				*p<0.1; **p<0.05; ***p<0.01

Still, the most notable models predicted attractiveness, not political demeanor. Thus, I cannot conclude that the specific traits tested contribute to a candidate-worthy appearance as I had hypothesized. Further, I will not conduct phase four of my research plan, where I had hoped to rate current congressmembers' level of candidate-worthy appearance, since I cannot determine what exactly that candidate-worthy appearance would be.

8.2 Phase 2

Seven mock races were successfully conducted from the survey. Unfortunately, data for one race in the male photo group, based on the real 2024 New York District 19 Congressional race, was corrupted by the survey tool, Qualtrics, and was unable to be recovered. The mock races successfully conducted in the male group were the 2024 Pennsylvania District 8 Congressional race and two fictional races made of men from stock photos. The mock races for the female group were the 2024 Iowa District 1 Congressional Race, the 2024 Virginia District 2 Congressional race, and two fictional races made of women from stock photos. The order candidates were shown to respondents was randomized to prevent bias towards any candidate.

Figure 5. Mock Election Results t-Tests

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Winner Vote Share
Iowa District 1*	8.694	278	3.0900e-16	0.679	0.784	0.731
Virginia District 2*	1.061	287	2.9000e-01	0.473	0.589	0.531
Pennsylvania District 8*	6.679	284	1.2700e-10	0.630	0.739	0.684
Fictional Women - 1	11.018	316	4.1700e-24	0.716	0.810	0.763
Fictional Women - 2	33.886	305	2.0400e-105	0.919	0.970	0.944
Fictional Men - 1	1.243	313	2.1500e-01	0.480	0.591	0.535
Fictional Men -2	12.264	308	1.9900e-28	0.740	0.832	0.786

**Respondents who recognized either candidate in the real races were excluded from the results for that respective race.*

I conducted t-Tests to determine if the winner's vote share was statistically significantly different from 0.5 for each race. The results are displayed in Figure 5 above. All but two mock races (Virginia District 2 and the first fictional men race) had a statistically significant winner. The lower end of all other races' 95% confidence intervals were over 0.630. The most extreme race, the second fictional women's race, had a 95% confidence interval of 0.919 to 0.970 and a winner vote share of 0.944. The average winner vote share for real male races and fictional male races was roughly equal (about 66.05% of the vote for fictional races, and 68.4% of the vote in the real race). However, the average winner vote share for real female races was much smaller than that of the fictional female races. The average winner vote share was 63.1% for the real female races (slightly smaller, though comparable to, the average winner vote share for all male races), but a whopping 85.35% for the fictional female races. This is not just due to the extreme

outcome for the second fictional women race; the first fictional women race had a winner vote share of 76.3%. Thus, both of the fictional women races were well over the average winner vote share in any other category.

Two of the three mock elections produced the same winner as their respective real races. Notably, the results of the IA-01 mock race did not line up with the real race. The real election was a close victory for Mariannete Miller-Meeks, whereas the mock election resulted in a decisive victory for Christina Bohanan. The other real mock races however produced the same winner as in the actual elections. The VA-02 mock race produced a similarly close win for Jennifer Kiggans, and the PA-08 mock race produced a much larger victory for Robert Bresnahan Jr. compared to reality.

Figure 6a. IA-01 Candidate Ratings t-Tests*

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	2.654	279	8.4000e-03	0.715	4.820	2.768
Competence	3.859	279	1.4200e-04	1.853	5.712	3.782
Attractivess	15.480	279	1.9200e-39	15.471	19.979	17.725

Figure 6b. VA-02 Candidate Ratings t-Tests*

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	4.576	289	7.0400e-06	2.810	7.052	4.931
Competence	3.989	289	8.4100e-05	2.100	6.190	4.145
Attractivess	-9.630	289	3.2400e-19	-12.164	-8.036	-10.100

Figure 6c. PA-08 Candidate Ratings t-Tests*

Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	4.514	284	9.3300e-06	3.196	8.138	5.667
Competence	2.063	284	4.0000e-02	0.099	4.237	2.168
Attractivess	10.788	284	5.7600e-23	13.006	18.812	15.909

Figure 6d. Women Fictional Race 1 Candidate Ratings t-Tests						
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	0.174	317	8.6200e-01	-2.243	2.677	0.217
Competence	-0.560	317	5.7600e-01	-3.366	1.875	-0.745
Attractivess	-2.366	317	1.8600e-02	-7.666	-0.705	-4.186

Figure 6e. Women Fictional Race 2 Candidate Ratings t-Tests						
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	1.199	306	2.3100e-01	-1.019	4.198	1.590
Competence	0.248	306	8.0400e-01	-2.143	2.762	0.309
Attractivess	1.432	306	1.5300e-01	-0.786	4.988	2.101

Figure 6f. Men Mock Fictional 1 Candidate Ratings t-Tests						
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	0.183	313	8.5500e-01	-2.083	2.510	0.213
Competence	-0.678	313	4.9800e-01	-3.865	1.884	-0.990
Attractivess	0.688	313	4.9200e-01	-2.873	5.962	1.545

Figure 6g. Men Mock Fictional 2 Candidate Ratings t-Tests						
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Mean Score Difference**
Trustworthiness	-1.389	308	1.6600e-01	-3.997	0.689	-1.654
Competence	1.372	308	1.7100e-01	-0.749	4.199	1.725
Attractivess	-1.721	308	8.6300e-02	-3.212	0.215	-1.498

**Respondents who recognized either candidate in the real races were excluded from the results for that respective race.*

***Mean Score Difference is calculated as mock-race winner score minus mock-race loser score.*

T-Tests were then conducted to determine if candidates were rated statistically significantly differently on each scale in each race. Results are shown in Figures 6a-g. The difference calculated is for the mock race winner minus the mock race loser. The real races showed much more statistically significant differences in trustworthiness, competence, and attractiveness scores of each candidate compared to the fictional races. All three real races showed a statistically significant difference on all three measures, particularly attractiveness scores. For IA-01 and PA-08, the mock-race winner was rated 17.725 and 15.909 points higher for attractiveness on average

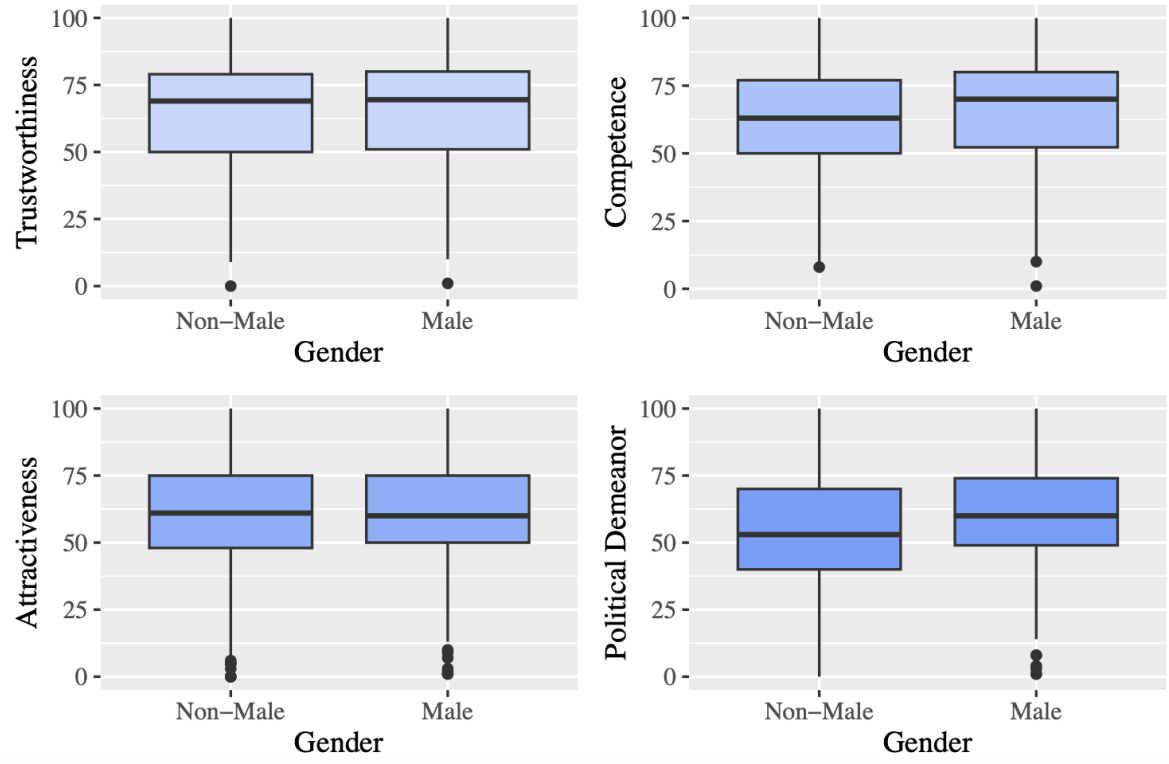
than the loser respectively. These were also the two real races with a statistically significant winner. For VA-02 however, the mock-race winner was rated 10.100 points lower on attractiveness than the loser. In all three real races, the mock-race winner was rated higher on trustworthiness and competence, however not by very much (around two to six points). Oddly enough, although all of the fictional races generally had very clear winners, the only measure that was statistically significant in any race was attractiveness in the first fictional women mock-race, and not by a notable margin. The winner was rated statistically significantly less attractive, but only by 4.186 points on the zero to 100 scale, which does not tell us much. This again suggests that, consciously or not, respondents are likely voting with qualities other than trustworthiness, competence, and attractiveness in mind.

8.3 Phase 3

Though there is strong evidence that men are more socialized to run for office or believe they could win than non-men, it does not appear to be because women (and gender nonconforming people) are socialized to believe people see them as less trustworthy or attractive based on appearance. I ran standard two-mean Welch t-Tests on respondents' ratings of how they believe others would rate them on each scale based on gender, as shown in Figure 7a. For both competence and political demeanor, men rated others' anticipated assumptions statistically significantly higher than women, but only by about 4.465 points for competence and about 5.289 points for political demeanor on the zero to 100 scale.

Figure 7a. How Respondents Believe Others Would Rate Them by Gender t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-Male Mean	Male Mean
Trustworthiness	-0.539	361.329	5.9000e-01	-4.403	2.510	64.538	65.485
Competence	-2.520	402.749	1.2100e-02	-7.948	-0.983	61.839	66.304
Attractivess	0.103	389.161	9.1800e-01	-3.657	4.062	59.976	59.773
Political Demeanor	-2.861	369.449	4.4700e-03	-8.925	-1.653	53.518	58.807

Figure 7b. How Respondents Believe Others Would Rate Them by Gender Boxplots



Based on t-Tests displayed in Figures 8a and 8b, male respondents were statistically significantly more likely to consider running for all offices, and statistically significantly more likely to believe they could win any office.

Figure 8a. Would Respondents Run by Gender t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-Male Mean	Male Mean
Local	-3.095	380.462	2.1100e-03	-0.191	-0.043	0.290	0.407
State	-2.713	381.267	6.9700e-03	-0.175	-0.028	0.284	0.385
House	-4.027	347.334	6.9400e-05	-0.209	-0.072	0.178	0.319
Senate	-3.026	352.245	2.6600e-03	-0.165	-0.035	0.161	0.261
President	-4.385	291.614	1.6200e-05	-0.173	-0.066	0.058	0.177

Figure 8b. Do Respondents Think they Could Win by Gender t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-Male Mean	Male Mean
Local	-3.113	393.521	1.9900e-03	-0.197	-0.044	0.362	0.482
State	-4.236	363.790	2.8900e-05	-0.230	-0.084	0.241	0.398
House	-4.889	321.612	1.6000e-06	-0.227	-0.097	0.126	0.288
Senate	-3.943	310.750	9.9500e-05	-0.170	-0.057	0.081	0.195
President	-3.219	289.569	1.4300e-03	-0.114	-0.027	0.036	0.106

Next, I created logistic regression models to predict how likely a respondent was to report ever considering running for office or believing they could ever win at several different governmental levels (local, state, House of Representatives, Senate, and the Presidency). Statistically significant results in these models indicate why men are more likely to consider running or believe they could win than non-men. Based on these logistic regression models (illustrated in Figures 9a and 9b), higher competence scores do statistically significantly (at a 95% confidence level) increase the odds someone would consider running for president and the odds they believe they could win a presidential election. Thus it is possible that women (and gender nonconforming people) tend to believe people view them as less competent due to their appearance and are then discouraged from considering running for president because they do not believe they could win. However, this effect seems quite small according to the results of my study. The logistic models also illustrate that higher political demeanor scores statistically significantly (at a 95% confidence level) increase the odds someone would run for state level office, the House of Representatives, or the Senate and the odds they believe they could win a local office, state office, the House of Representatives, or the Senate. Therefore it is also possible that women (and gender nonconforming people) believe people would consider them to be worse state or House representatives than men on average, based on appearance, discouraging them from considering running in these offices. Additionally, these models indicate that believing others see the respondents as more attractive makes the respondents more likely to believe they could win a local or state election, but does not make them more likely to consider running. This cannot be directly tied to explaining why less women believe they would run than men however as men and non-men did not rate their perceived attractiveness statistically significantly differently than each other.

Figure 9a: Logistic Models - Running by Qualities and Gender

	Dependent variable:					
	Run at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	−0.0110*	0.0010	−0.0046	−0.0036	−0.0046	−0.0118
	(0.0059)	(0.0058)	(0.0058)	(0.0065)	(0.0067)	(0.0090)
Competence	−0.0045	−0.0083	−0.0013	0.0004	0.0023	0.0209**
	(0.0061)	(0.0061)	(0.0061)	(0.0068)	(0.0071)	(0.0100)
Attractiveness	−0.0022	−0.0050	−0.0050	−0.0069	−0.0048	−0.0058
	(0.0045)	(0.0045)	(0.0045)	(0.0051)	(0.0052)	(0.0072)
Political Demeanor	0.0215***	0.0089	0.0138**	0.0198***	0.0144**	0.0145
	(0.0058)	(0.0056)	(0.0058)	(0.0065)	(0.0067)	(0.0091)
Gender	0.0817	0.1975	0.0802	0.4179**	0.2571	0.9470***
	(0.1821)	(0.1787)	(0.1798)	(0.1917)	(0.1999)	(0.2569)
Constant	0.1981	−0.0614	−0.4406	−1.5281***	−1.5407***	−3.4677***
	(0.3177)	(0.3134)	(0.3165)	(0.3570)	(0.3679)	(0.5490)
Observations	601	601	601	601	601	601
Log Likelihood	−402.2888	−408.5928	−404.9065	−352.3276	−334.8508	−208.6161
Akaike Inf. Crit.	816.5777	829.1855	821.8130	716.6551	681.7016	429.2322

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 9b: Logistic Models - Winning by Qualities and Gender

	Dependent variable:					
	Win at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	−0.0147**	−0.0091	−0.0100	−0.0158**	−0.0142*	−0.0111
	(0.0063)	(0.0059)	(0.0063)	(0.0073)	(0.0082)	(0.0112)
Competence	0.0060	0.0052	0.0066	0.0118	0.0150*	0.0287**
	(0.0065)	(0.0061)	(0.0065)	(0.0078)	(0.0089)	(0.0127)
Attractiveness	0.0109**	0.0103**	0.0115**	−0.0052	0.0008	−0.0053
	(0.0048)	(0.0046)	(0.0049)	(0.0058)	(0.0066)	(0.0089)
Political Demeanor	0.0308***	0.0177***	0.0210***	0.0278***	0.0157*	0.0150
	(0.0061)	(0.0057)	(0.0061)	(0.0075)	(0.0083)	(0.0114)
Gender	0.3046	0.0682	0.3884**	0.6798***	0.6918***	0.8276***
	(0.1981)	(0.1836)	(0.1871)	(0.2064)	(0.2354)	(0.3190)
Constant	−1.3279***	−1.2223***	−2.2652***	−2.5015***	−2.9931***	−4.6554***
	(0.3457)	(0.3292)	(0.3638)	(0.4148)	(0.4848)	(0.7349)
Observations	601	601	601	601	601	601
Log Likelihood	−362.1000	−396.4573	−373.7263	−300.6222	−244.0230	−147.9895
Akaike Inf. Crit.	736.2001	804.9146	759.4526	613.2444	500.0459	307.9790

Note: *p<0.1; **p<0.05; ***p<0.01

When I repeated this analysis based on race, I found some mixed results. I wanted to evaluate whether white people are more likely to believe others perceive them higher on any ratings and whether they are more likely to think they could win or consider running. Since respondents were allowed to select all racial/ethnic groups that described their identity, I defined whiteness in two different ways. First, in Figures 10a through 12b, I define white as anyone who marked the white box, regardless of any additional racial/ethnic identification. Then, in Figures 17a through 19b, I defined white as people who only identified as white, which can be found in the appendix. I chose to evaluate whiteness from different definitions because I was unsure if the presence of a white identity or the absence of a racial/ethnic minority identification would affect the results more.

Figure 10a. How Respondents Believe Others Would Rate Them by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-White Mean	White Mean
Trustworthiness	-1.883	457.779	6.0300e-02	-6.332	0.135	63.763	66.862
Competence	0.278	412.079	7.8100e-01	-3.055	4.061	63.450	62.948
Attractivess	-4.218	493.509	2.9400e-05	-11.265	-4.105	57.234	64.919
Political Demeanor	-1.676	451.927	9.4400e-02	-6.447	0.512	54.176	57.144

Figure 10b. How Respondents Believe Others Would Rate Them by Race Boxplots

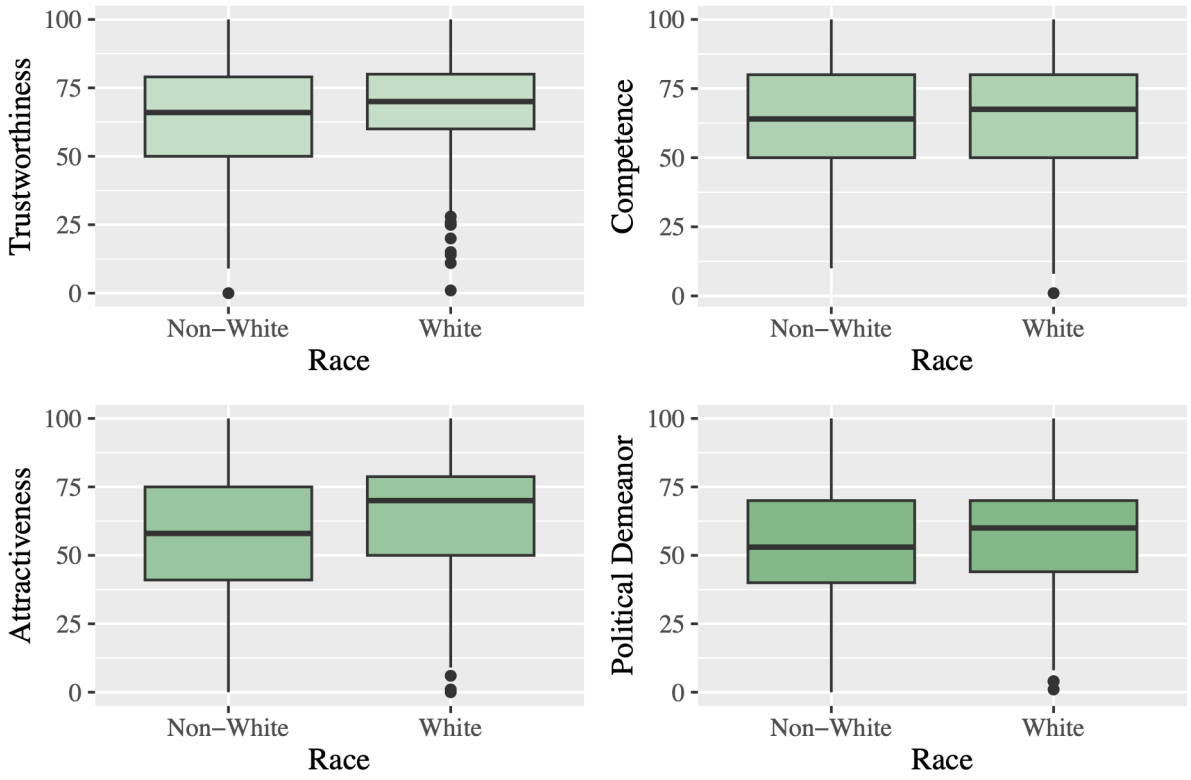


Figure 11a. Would Respondents Run by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-White Mean	White Mean
Local	-0.796	443.926	4.2600e-01	-0.100	0.042	0.314	0.343
State	-1.396	436.572	1.6400e-01	-0.121	0.021	0.297	0.347
House	-0.631	439.870	5.2900e-01	-0.083	0.043	0.211	0.231
Senate	-1.192	425.387	2.3400e-01	-0.098	0.024	0.178	0.215
President	-0.270	441.149	7.8700e-01	-0.050	0.038	0.089	0.095

Figure 11b. Do Respondents Think they Could Win by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	Non-White Mean	White Mean
Local	-4.252	433.697	2.6000e-05	-0.235	-0.087	0.347	0.508
State	-3.448	410.738	6.2300e-04	-0.195	-0.053	0.248	0.372
House	-2.801	389.073	5.3400e-03	-0.147	-0.026	0.145	0.231
Senate	-1.744	394.995	8.1900e-02	-0.096	0.006	0.099	0.145
President	-0.532	422.444	5.9500e-01	-0.045	0.026	0.052	0.062

According to both definitions, only the t-Test for attractiveness showed statistical significance. White respondents believed others would perceive their attractiveness to be higher based on appearance than non-white respondents, by 7.685 points for pooled white respondents and 6.869 points for exclusively white respondents, which is a bit larger than any differences for gender. However, there is much less of a difference between white and non-white respondents potentially running or believing they could win. Neither definition of whiteness showed white people to be statistically significantly more likely to consider potentially running at any level, illustrated in Figures 11a and 18a. As shown in Figure 11b, pooled white respondents are only statistically significantly more likely to believe they could win a local, state, or House race. Exclusively white respondents were only statistically significantly more likely to believe they could win a local or state race than respondents of color, displayed in Figure 18b in the appendix.

Figure 12a: Logistic Models - Running by Qualities and Race

	Dependent variable:					
	Run at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	−0.0108*	0.0011	−0.0046	−0.0042	−0.0051	−0.0133
	(0.0059)	(0.0058)	(0.0058)	(0.0065)	(0.0067)	(0.0090)
Competence	−0.0051	−0.0087	−0.0015	0.0007	0.0028	0.0217**
	(0.0062)	(0.0061)	(0.0061)	(0.0068)	(0.0071)	(0.0098)
Attractiveness	−0.0016	−0.0044	−0.0047	−0.0067	−0.0050	−0.0063
	(0.0046)	(0.0045)	(0.0046)	(0.0051)	(0.0053)	(0.0071)
Political Demeanor	0.0219***	0.0096*	0.0141**	0.0211***	0.0151**	0.0176*
	(0.0058)	(0.0056)	(0.0057)	(0.0065)	(0.0066)	(0.0091)
Race	−0.1882	−0.2328	−0.0989	−0.1686	−0.0184	−0.1542
	(0.1787)	(0.1784)	(0.1788)	(0.1972)	(0.2030)	(0.2754)
Constant	0.2513	0.0257	−0.4046	−1.3963***	−1.4762***	−3.1282***
	(0.3173)	(0.3126)	(0.3155)	(0.3539)	(0.3660)	(0.5325)
Observations	601	601	601	601	601	601
Log Likelihood	−401.8354	−408.3468	−404.8525	−354.3104	−335.6654	−215.2158
Akaike Inf. Crit.	815.6709	828.6936	821.7050	720.6208	683.3308	442.4317

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 12b: Logistic Models - Winning by Qualities and Race

	Dependent variable:					
	Win at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	-0.0161^{**}	-0.0097	-0.0111^{*}	-0.0180^{**}	-0.0159^{*}	-0.0125
	(0.0064)	(0.0060)	(0.0063)	(0.0073)	(0.0082)	(0.0112)
Competence	0.0084	0.0064	0.0082	0.0141^{*}	0.0166^{*}	0.0293^{**}
	(0.0065)	(0.0062)	(0.0065)	(0.0077)	(0.0088)	(0.0125)
Attractiveness	0.0088^{*}	0.0092^{**}	0.0100^{**}	-0.0078	-0.0010	-0.0061
	(0.0048)	(0.0046)	(0.0049)	(0.0058)	(0.0066)	(0.0088)
Political Demeanor	0.0312^{***}	0.0177^{***}	0.0220^{***}	0.0303^{***}	0.0183^{**}	0.0180
	(0.0062)	(0.0057)	(0.0061)	(0.0076)	(0.0083)	(0.0114)
Race	0.4257^{**}	0.2609	0.2169	0.3852^{*}	0.1823	-0.0331
	(0.1962)	(0.1817)	(0.1850)	(0.2104)	(0.2421)	(0.3385)
Constant	-1.3311^{***}	-1.2542^{***}	-2.2061^{***}	-2.3912^{***}	-2.8269^{***}	-4.3706^{***}
	(0.3432)	(0.3281)	(0.3606)	(0.4102)	(0.4781)	(0.7198)
Observations	601	601	601	601	601	601
Log Likelihood	-360.9050	-395.4914	-375.1868	-304.3097	-247.9862	-151.3266
Akaike Inf. Crit.	733.8101	802.9828	762.3736	620.6195	507.9724	314.6532

Note: *p<0.1; **p<0.05; ***p<0.01

In the logistic models in Figures 12a and 12b (nearly identical to 9a and 9b, just swapping the gender predictor for the race predictor, using pooled white respondents), an increase in political demeanor still statistically significantly (at a 95% confidence level) increases the odds a respondent may run at any level, and the odds that they believe they could win, now at every level except president. Attractiveness, the only trait white respondents rated their expected perception statistically significantly higher on, statistically significantly increases the odds someone believes they could win a local or state level office at the 95% confidence level. The logistic models in Figures 19a and 19b, the models where whiteness is defined as respondents who are exclusively white, are nearly indistinguishable for all traits except for race itself. Race itself no longer contributes to any of the models statistically significantly. Thus, it is possible that non-white people believe others perceive them as less attractive, and that decreased attractiveness will make them less likely to believe they could win a local or state office. Interestingly however,

this relationship does not hold for considering a potential run. Further, it appears it is more meaningful to define whiteness as the presence of a white identity rather than the absence of a minority racial/ethnic identity for the purposes of this study. It is more helpful to know if a respondent identifies as white at all than to know if they exclusively identify as white to predict how they believe others will perceive them, whether they may consider running for any office, and whether they believe they could win any race.

9 Discussion

The results of this study are largely consistent with existing research in the fields of appearance and women in politics. For example, most mock elections showed clear winners without any information other than appearance similarly to the results in much of the previous literature (Olivola and Todorov 2010, Todorov et al 2005, Ballew and Todorov 2007, Rosenberg, Kahn, and Tran 1991, Little et al 2007, Antonakis and Dalgas 2009, Buckley, Collins, and Reidy 2007, Castelli et al 2009, Banducci et al 2008), indicating that appearance is in fact important in elections. This study builds on these findings by investigating what a candidate-worthy appearance looks like, directly comparing results of mock races for men vs men and women vs women, and examining a potential mechanism of socialization to explain women's lower candidacy rates compared to men.

Though unable to offer a concrete definition of a candidate-worthy appearance, phase one does present some interesting information about appearance. First of all, trustworthiness, competence, attractiveness, and photo gender can only explain about 35% of the variance in political demeanor scores. This indicates that other perceived qualities likely contribute to evaluations of political demeanor. The literature has largely focused on these qualities to explain judgements of political demeanor based on appearance alone, but my research suggests investigating new qualities may better explain how people judge candidates' appearances. For instance, people may interpret appearances as confident, relatable, or authentic and that judgement may be why they prefer some candidates' appearances over others. I present confidence, relatability, and authenticity as potential qualities as they are often spoken about in relation to politicians. Many voters tend to want "tough" candidates who they believe would stand up for their beliefs; this would correspond to assumptions about confidence. Many voters also judge candidates poorly

because they do not feel as though they would get along with them in a casual setting; this would correspond to assumptions about relatability. Further, voters generally want candidates who are true to themselves and do not put on a facade; this would correspond to assumptions about authenticity.

Further, phase one indicates that the physical traits examined contribute most to judgements of attractiveness. This makes sense as attractiveness is the quality most closely tied to appearance. All of the models, especially the models for attractiveness, had stronger predictability, and had different statistically significant predictors, when only one racial/ethnic group was included. This suggests that different traits are likely more important for different racial/ethnic groups, which makes sense as different racial groups can face different beauty standards. For example, in the model for attractiveness by trait for Asians specifically, the traits align with existing racist beauty standards. Asian photos were considered statistically significantly more attractive with less dramatically angled eyes. Eyes placed at a more noticeable angle are commonly associated with Asian heritage, and unfortunately often deemed unattractive. Eye size, eyelid structure, and other eye-related appearances are heavily scrutinized according to Asian beauty standards, even making plastic surgery to change the appearance of the eye commonplace (Chen et al. 2020). Thus, my results seem to reflect existing biases in judgements of attractiveness, indicating that my methodology was generally sound. Still, the models were too weak to reliably predict trustworthiness, competence, attractiveness, or most importantly political demeanor scores, so I can not determine what exactly a candidate-worthy appearance would be. There are likely different candidate-worthy appearances for different racial/ethnic groups, so any future work on a potential candidate-worthy appearance should take that into account.

Phase two demonstrates appearance clearly matters in elections since the respondents overwhelmingly agreed on a winner in most races, with no information other than appearance. Since the fictional women mock races had a larger margin of victory than the real women races, there is likely something about the appearance of women candidates who run that is different from the general population. This is evidence of some sort of candidate-worthy appearance for women and provides support for my hypothesis that women who do not have a candidate-worthy appearance would face lower odds of winning than women with candidate-worthy appearances; it seems whichever candidate in the fictional races was further from the candidate-worthy appearance held by real candidates was more extremely punished electorally. I attempted to measure

this candidate-worthy appearance through physical traits in phase one, but it appears the traits I measured, or at least the way I measured them, do not reflect this difference in appearance. Still, there appears to be some appearance based X-factor that candidates have that not everyone in the general population has. Otherwise the winner vote share for fictional and real races should have been roughly similar.

Results in phase two make more sense given that there are likely some qualities other than trustworthiness, competence, or attractiveness also inferred based on appearance. This could explain why most of the fictional races did not see statistically significant differences in trustworthiness, competence, or attractiveness scores, despite having winners by large margins. Clearly respondents saw something in candidates in most of the fictional races that they overwhelmingly agreed made one a better choice than the other. Whatever they saw does not appear to be related to judgements about trustworthiness, competence, or attractiveness. The existence of other relevant inferred qualities may also explain why strong attractiveness scores aligned with the winner in the IA-01 and PA-08 cases but not in the VA-02 case. Attractiveness may have been important in IA-01 and PA-08, but a different quality may have been inferred about the candidates in the VA-02 case that cancelled out attractiveness when respondents made their choice. Thus, future research should investigate other possible inferred qualities. Again, these qualities could perhaps include confidence, relatability, or authenticity. However, because the IA-01 and PA-08 cases were the only real elections with statistically significant winners and both had drastically higher attractiveness scores for the winner, attractiveness seems to be the most relevant quality to electoral success tested in this study.

Based on phase three, it appears that men are more socialized to consider running for office or believe they could win than women and gender nonconforming people. Furthermore, there is evidence tying this to inferences about competence and political demeanor, though it is weak. Men believe others would rate them statistically significantly higher on both competence and political demeanor (though not by a dramatic amount). Higher ratings of others' perception of competence are statistically significantly associated with higher odds of potentially running for and believing one could win the presidency. Similarly, higher ratings of political demeanor are statistically significantly associated with higher odds of considering running for state office, the House of Representatives, or the Senate, and believing one could win those offices in addition to local offices. This could mean that women and gender non-comforming people are socialized

to believe others see them as less competent and worse potential political leaders based on appearance, therefore believe they could not win, and choose not to run. Because these results, though statistically significant, may not be practically strong, the perceived qualities tested may not be the only, or most relevant, ones that lead to being seen as more candidate-worthy. Again, it is possible other qualities, such as confidence, relatability, or authenticity, are inferred based on appearance, socializing women against running for office if they lack a confident, relatable, and/or authentic appearance.

When I repeated the analysis for race, I found white and non-white people do believe people view them differently based on appearance, and this may translate to higher confidence of success, but not a higher likelihood of running. This was only statistically significant for attractiveness, not any other qualities tested. Furthermore, white people were only more likely to think they could win for local, state, and the House of Representatives level, and were not more likely to consider running at any level. Additionally, it appears that identifying as white at all is more relevant to respondents' answers in this section than whether a respondent exclusively identified as white.

Notably, some of my analysis in phase three is complicated, as respondents may believe people would perceive them differently on the basis of gender or race itself, not appearance. This seems possible. Perceived competence and political demeanor were the only qualities women (and gender nonconforming people) rated statistically significantly lower, and there are stereotypes about women being less intelligent than men. Therefore, that difference could be because women and gender nonconforming people may believe people will see their gender, and then interpret them to be less intelligent, rather than thinking their appearance itself causes others to judge them as less intelligent. Put differently, women and gender nonconforming people may believe that even without others seeing their appearance, they would automatically assume them to be less competent if they knew they were not a man. Based on perceived political demeanor scores, non-men may also believe people will see their gender, rather than some other aspect of their appearance, and directly judge them to be a less capable representative than a man. Similarly, non-white respondents rated their perceived attractiveness lower than white respondents which may be due to white-skewed, Eurocentric beauty standards. Further research should more cleanly distinguish between gender/race itself and different aspects of appearance when asking these questions, though that will be a difficult task. Gender, race/ethnicity, and appearance are

intertwined in many ways.

10 Limitations and Future Research

It is likely that these results are skewed by the sample of respondents. UCLA tends to have a predominately young, liberal student body, so this may be a biased survey. Future research should be done on other samples, especially more politically diverse samples with wider age ranges, and should collect respondent political affiliations to determine if people are perceived differently by those with different political ideologies. Thus future research should conduct a similar survey on a population different from UCLA students.

Furthermore, this study was financially limited, so I was unable to recruit and photograph people for the purposes of this study and had to rely on stock photos. Though I collected a diverse sample of stock photos for phase one, it is possible that stock photos may not be fully representative of the general population. Even if my finding in phase one, that, on average, women are rated higher on each scale than men, is accurate, it does not account for how much effort was put into that appearance. For instance, the female stock photos may have been rated higher on every quality because there is a higher bar for women's appearance, especially when they are expecting to be photographed. Further research should be conducted on the time and effort female vs male candidates put into their appearance.

The general study design used in phase two should be repeated in future studies to determine if my results were skewed by the specific photos chosen. I attempted to choose reasonable stock photos, and ensured that the clothes of each candidate matched so that each race would be as fair as possible. For the real races, I only chose races where the race and gender of the candidates matched and where the actual election results were very close. Still, despite these efforts, it is possible that results would look very different with different photos. Part of the reason I only ran eight mock races (four for each respondent as they only saw either men or women) was to minimize respondent fatigue as this was one section of a larger study. Future work should focus on this style of mock races, comparing real and fictional races, so that they can run more such races. Furthermore, future work should include matched races with non-white candidates. I focused on the House of Representatives 2024 races for the sake of consistency, and there were not enough close, matched races to pull from different races (especially for women of color), so

I felt it best to focus on the races I chose. To fully grasp the role of appearance in elections however, candidates of all racial/ethnic demographics should be thoroughly studied.

The third phase of my study only investigated socialization as a possible mechanism by which appearance filters women out of the candidate pool. Socialization is one of many such mechanisms explored by the literature. Thus, further research should evaluate the possible link between appearance and barriers to women running for office other than socialization. For example, research into how appearance affects political recruiters' decisions to recruit male vs female candidates would be a valuable contribution to the literature. This would combine the theory behind my research with past research indicating that political recruiters are more likely to recruit men than women (Fox and Lawless 2010).

11 Conclusion

Overall, this study does suggest that appearance is more important for women than men when running for office, and it provides limited evidence for some potential mechanisms for this phenomenon. Women (and gender nonconforming people) may be socialized to believe others perceive them as less competent and worse political leaders than men, making them less likely to believe they could win, thus discouraging them from running for office. If this is the case, it follows that women with a certain candidate-worthy appearance may be less susceptible to this type of socialization, making them more likely to ultimately decide to run for office. Women with this elusive candidate-worthy appearance may have been more likely to be perceived as intelligent, competent leaders growing up, thus encouraging them towards political candidacy.

This candidate-worthy appearance seems more relevant to women than men. Though there were still clear winners in many male races, there was no notable distinction between winner vote share of real and fictional male races. This indicates that it was not particularly important whether the male candidates shown were real candidates or stock photos, suggesting that there may not be such a thing as a candidate-worthy look for men, just a look that makes them more likely to win. This is a subtle distinction but an important one. Women are largely underrepresented in government because they are underrepresented in the candidate pool, not because they are less likely to win when they do run. Thus, this thesis aimed to determine if appearance impacted whether women run for office, not necessarily if appearance impacts

who wins; it is already well established that appearance does affect who wins. Unlike the male races, there was a significant difference in winner vote share between real and fictional female races, with real candidates scoring closer to each other, suggesting that a candidate-worthy appearance exists. There was something about the women' s appearances in the real races that made them harder to choose between than the women in the fictional races. I theorize this is because the real women candidates have more of a candidate-worthy appearance than the stock women. This evidence is limited, especially as I was unable to determine what exactly makes up a candidate-worthy appearance in phase one, but it is notable enough to consider in future research.

Though I've uncovered several interesting findings, my study faces several limitations. Since appearance is such a difficult concept to operationalize, my method of operationalization likely did not capture everything relevant about appearance. Since beauty standards can be different across cultures, my results may apply specifically to the US and countries with similar cultures. Because the US contains a multitude of cultures itself, it may also be difficult to find results that apply universally across the US. This is likely part of the reason I was unable to find which specific traits lead to a candidate-worthy appearance: the candidate-worthy traits may be regional. I also only examined candidates for the US House of Representatives, but appearance may be more relevant in more local, low-information elections. Future research should examine my research questions for other countries and different levels of government. My research design should be applicable to these cases by swapping out candidates from other countries or candidates running for different offices. Therefore, my research design could also be a contribution to the literature. I encourage future research to use my design as a framework for investigation when appropriate.

Similar research should also be applied to other underrepresented groups. Different racial groups for example appear to be judged on their appearance differently, affecting their choices to run. I tried to address race throughout my design but I still focused on gender. Thus, more research dedicated to specifically analyzing how appearance affects the choice to run for different racial/ethnic groups would be valuable.

Further research into gender nonconforming people and the queer community as a whole should also be explored. This research would be logistically difficult to conduct, especially using parts of my design as a template, because so few candidates and representatives openly defy gender norms. Sarah McBride is currently the only openly transgender member of Congress.

Notably, she still identifies with a binary gender, and there are no nonbinary or genderfluid members of congress. It would be nearly impossible to find close races at any level between two openly queer individuals, particularly two queer individuals with the same gender identity and sexuality. However, since my argument revolves around women not running for office in the first place, research into groups even more severely underrepresented in the candidate pool would be all the more important.

12 Works Cited

Aguiar, Gary, and Meredith Redlin. 2014. "Women's Continued Underrepresentation in Elective Office." Great Plains Research24(2). https://www.proquest.com/docview/1619303368/?sourcetype=Schola

Al-Nuimi, Arqam, and Ghassan Mohammed. 2021. "Face Direction Estimation Based on Mediapipe Landmarks." IEEE Conference Publication — IEEE Xplore. https://ieeexplore.ieee.org/abstract/doc$p-yWvNbOdUkAAAAA:ZQOr-07h8Do2UVZ5AlgClroZLYF_1U4MtfZJeygq6DmLhaty-AhEhpyGfdX48nnK4eTPlQNHPDU.$

Antonakis, John, and Olaf Dalgas. 2009. "Predicting Elections: Child's Play!" Science 323(5918): 1183. doi:10.1126/science.1167748.

Anzia, Sarah F., and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" American Journal of Political Science 55(3): 478–93. doi:10.1111/j.1540-5907.2011.00512.x.

Arevalo-Flores, Heather Reena. 2017. "Social Scrutiny: How Viewers' Comments Affect Female TV Journalists' Appearance." The University of Texas Rio Grande Valley. https://scholarworks.utrgv.edu/c

Ballew, Charles C., and Alexander Todorov. 2007. "Predicting Political Elections from Rapid and Unreflective Face Judgments." Proceedings of the National Academy of Sciences 104(46): 17948–53. doi:10.1073/pnas.0705435104.

Banducci, Susan A., Jeffrey A. Karp, Michael Thrasher, and Colin Rallings. 2008. "Ballot Photographs as Cues in Low-Information Elections." Political Psychology 29(6): 903–17. doi:10.1111/j.1467-9221.2008.00672.x.

Buckley, Fiona, Neil Collins, and Theresa Reidy. 2007. "Ballot Paper Photographs and Low-Information Elections in Ireland." Politics 27(3): 174–81. doi:10.1111/j.1467-9256.2007.00297.x.

Castelli, Luigi, Luciana Carraro, Claudia Ghitti, and Massimiliano Pastore. 2009. "The Effects of Perceived Competence and Sociability on Electoral Outcomes." Journal of Experimental Social Psychology 45(5): 1152–55. doi:10.1016/j.jesp.2009.06.018.

Chen, Toby, Kristina Lian, Daniella Lorenzana, Naima Shahzad, and Reinesse Wong. 2020. "Occidentalisation of Beauty Standards: Eurocentrism in Asia." International Socioeconomics Laboratory 1(2). doi:10.5281/zenodo.4325856.

Ditonto, Tessa, and Kyle Mattes. 2018. "Differences in Appearance-Based Trait Inferences for Male and Female Political Candidates." Journal of Women Politics Policy 39(4): 430–50. doi:10.1080/1554477x.2018.1506206.

Druckman, James N. 2003. "The Power of Television Images: The First Kennedy-Nixon Debate Revisited." The Journal of Politics 65(2): 559–71. doi:10.1111/1468-2508.t01-1-00015.

Fox, Richard L., and Jennifer L. Lawless. 2010. "If Only They'd Ask: Gender, Recruitment, and Political Ambition." The Journal of Politics 72(2): 310–26. doi:10.1017/s0022381609990752.

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. doi:10.1017/s0003055414000227.

Jackson, Anna. 2025. "Women Account for 28% of Lawmakers in the 119th Congress – Unchanged from the Last Congress." Pew Research Center. https://www.pewresearch.org/short-reads/2025/02/21/women-account-for-28-of-lawmakers-in-the-119th-congress-unchanged-from-the-last-congress/.

Johns, Robert, and Mark Shephard. 2007. "Gender, Candidate Image and Electoral Preference." The British Journal of Politics and International Relations 9(3): 434–60. doi:10.1111/j.1467-856x.2006.00263.x.

Kraus, Sidney. 1996. "Winners of the First 1960 Televised Presidential Debate between Kennedy and Nixon." Journal of Communication 46(4): 78–96. doi:10.1111/j.1460-2466.1996.tb01507.x.

Kumar Rao B, Narendra, Nagendra Panini Challa, Phalguna Krishna, and Sreenivasa Chakravarthi. 2023. "Facial Landmarks Detection System with OpenCV Mediapipe and Python Using Optical Flow (Active) Approach." IEEE Conference Publication — IEEE Xplore. https://ieeexplore.ieee.org/al
$wiPb{-}eytdq8AAAAA:T_ss{-}BKsjwej7wwxD9BKvJqkD2z{-}eiHAxKhggrefBKu5p_{O}bjfiUTEK4G7Q$
$v8vuzpJo$.

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): 574–89. doi:10.1111/j.1540-5907.2011.00511.x.

Little, Anthony C., Robert P. Burriss, Benedict C. Jones, and S. Craig Roberts. 2007. "Facial Appearance Affects Voting Decisions." Evolution and Human Behavior 28(1): 18–27. doi:10.1016/j.evolhumbehav.2006.09.002.

Olivola, Christopher Y., and Alexander Todorov. 2010. "Elected in 100 Milliseconds: Appearance-Based Trait Inferences and Voting." Journal of Nonverbal Behavior 34(2): 83–110. doi:10.1007/s10919-009-0082-1.

Rosenberg, Shawn W., Shulamit Kahn, and Thuy Tran. 1991. "Creating a Political Image: Shaping Appearance and Manipulating the Vote." Political Behavior 13(4): 345–67.

doi:10.1007/bf00992868.

Sanbonmatsu, Kira. 2006. "Do Parties Know That 'Women Win'? Party Leader Beliefs about Women's Electoral Chances." Politics Gender 2(04). doi:10.1017/s1743923x06060132.

Santoniccolo, Fabrizio, Tommaso Trombetta, Maria Noemi Paradiso, and Luca Rollè. 2023. "Gender and Media Representations: A Review of the Literature on Gender Stereotypes, Objectification and Sexualization." International Journal of Environmental Research and Public Health 20(10): 5770. doi:10.3390/ijerph20105770.

Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." American Political Science Review 112(3): 525–41. doi:10.1017/s0003055418000217.

Thomsen, Danielle M. 2019. "Which Women Win? Partisan Changes in Victory Patterns in US House Elections." Politics Groups and Identities 7(2): 412–28. doi:10.1080/21565503.2019.1584749.

Thomsen, Danielle M., and Aaron S. King. 2020. "Women's Representation and the Gendered Pipeline to Power." American Political Science Review 114(4): 989–1000. doi:10.1017/s0003055420000404.

Tian, Ziyao, Mark Hugo Lopez, Jeffrey Passel, Jens Manuel Krogstad, Nick Zanetti, Sara Atske, and John Carlo Mandapat. 2025. "Counting Race: How the Census Measures Identity and What Americans Think about It." Pew Research Center. https://www.pewresearch.org/race-and-ethnicity/2025/11/03/counting-race-how-the-census-measures-identity-and-what-americans-think-about-it/.

Todorov, Alexander, Anesu Mandisodza, Amir Goren, and Crystal Hall. 2005. "Inferences of Competence from Faces Predict Election Outcomes." Science 308(5728). https://www.proquest.com/docvie

Appendix

Phase 1 Figures

Figure 13: Linear Model - Political Demanor by Qualities and Photo Gender

	Dependent variable:
	Political Demeanor
Trustworthiness	0.2944***
	(0.0244)
Competence	0.3071***
	(0.0249)
Attractiveness	0.0822***
	(0.0173)
Photo Gender	3.3272***
	(0.7920)
Constant	13.8856***
	(1.0370)
Observations	3,231
R^2	0.3539
Adjusted R^2	0.3531
Residual Std. Error	20.9063 (df = 3226)
F Statistic	441.7624*** (df = 4; 3226)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 14: Relationship Between Tested Qualties and Gender

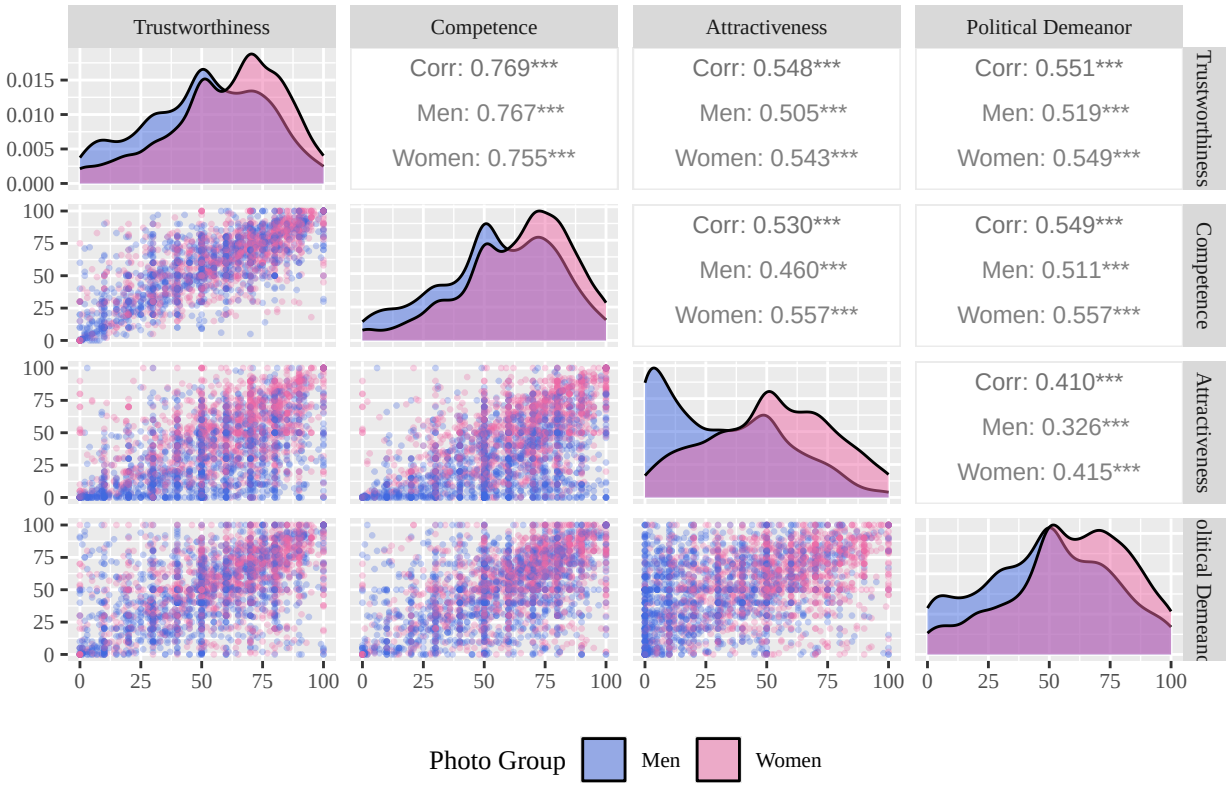


Figure 15: Principal Component Linear Regression for Traits

	Dependent variable:			
	Trustworthiness	Competence	Attractiveness	Political Demeanor
	(1)	(2)	(3)	(4)
PC1	-0.5431^{**}	-1.0695^{***}	-3.8989^{***}	-0.9932^{***}
	(0.2578)	(0.2521)	(0.2786)	(0.2771)
PC2	-5.2158^{***}	-3.7386^{***}	-6.1084^{***}	-4.4564^{***}
	(0.3572)	(0.3492)	(0.3859)	(0.3839)
PC3	-0.2750	-0.6046	-2.3001^{***}	-0.0772
	(0.4088)	(0.3997)	(0.4417)	(0.4395)
Constant	55.9882^{***}	59.6101^{***}	41.3645^{***}	53.7466^{***}
	(0.4162)	(0.4069)	(0.4497)	(0.4474)
Observations	3,232	3,232	3,232	3,231
R^2	0.0633	0.0401	0.1279	0.0437
Adjusted R^2	0.0624	0.0392	0.1271	0.0429
Residual Std. Error	23.6602 (df = 3228)	23.1329 (df = 3228)	25.5668 (df = 3228)	25.4302 (df = 3227)
F Statistic	72.7174^{***} (df = 3; 3228)	44.9702^{***} (df = 3; 3228)	157.8283^{***} (df = 3; 3228)	49.2053^{***} (df = 3; 3227)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 16a: Linear Model - Qualities by Trait, Respondent Demographic Breakdown

	Dependent variable:			
	Trustworthiness	Competence	Attractiveness	Political Demeanor
	(1)	(2)	(3)	(4)
Estimated Age	−0.29 (0.33)	−0.32 (0.33)	−3.43*** (0.36)	−0.44 (0.36)
Eye Shape	−0.30 (0.59)	−0.45 (0.58)	0.26 (0.63)	−1.11* (0.64)
Eye Angle	0.20* (0.11)	0.23** (0.11)	−0.01 (0.12)	0.07 (0.12)
Eye Distance	4.40 (5.12)	2.22 (5.09)	1.94 (5.52)	11.93** (5.58)
Eyebrow Thickness	12.38 (9.30)	−5.09 (9.25)	−26.03*** (10.04)	4.21 (10.15)
Face Ratio	10.07 (9.96)	5.35 (9.91)	21.94** (10.75)	−4.04 (10.87)
Lip Thickness	−0.54*** (0.19)	−0.30 (0.18)	−0.07 (0.20)	−0.33 (0.20)
Nose Ratio	−7.71 (5.79)	2.12 (5.76)	8.43 (6.25)	0.08 (6.32)
Photo Gender	4.83 (3.19)	0.82 (3.17)	18.82*** (3.44)	5.81* (3.48)
Black Photo†	5.90** (2.85)	2.79 (2.83)	10.58*** (3.07)	2.37 (3.11)
Latino Photo†	−3.10 (2.73)	−7.43*** (2.72)	3.18 (2.95)	−8.19*** (2.98)
White Photo†	−14.39*** (2.54)	−14.33*** (2.53)	−1.94 (2.75)	−11.50*** (2.78)
Male Respondents††	5.01*** (0.87)	2.05** (0.87)	1.73* (0.94)	2.00** (0.95)
Other Gender††	−1.71 (3.72)	−1.81 (3.70)	9.74** (4.02)	−3.97 (4.06)
White Respondents	3.15*** (1.15)	0.22 (1.14)	2.49** (1.24)	5.81*** (1.25)
Black Respondents	0.08 (1.72)	1.29 (1.71)	5.39*** (1.86)	0.09 (1.88)
Native Respondents	−2.84 (3.68)	−0.12 (3.66)	−8.48** (3.97)	−1.24 (4.01)
Asian Respondents	1.57 (1.32)	−0.84 (1.31)	2.35* (1.43)	3.43** (1.44)
Latino Respondents	0.14 (1.31)	0.29 (1.30)	−0.51 (1.41)	0.44 (1.43)
MENA Respondents	−0.11 (1.69)	0.53 (1.68)	0.83 (1.82)	2.58 (1.84)
Black Women Photos†††	2.15 (3.96)	2.44 (3.94)	−4.50 (4.27)	2.02 (4.32)
Latina Women Photos†††	2.06 (3.75)	7.37** (3.73)	1.08 (4.05)	6.09 (4.09)
White Women Photos†††	5.78* (3.38)	9.35*** (3.36)	0.04 (3.65)	4.02 (3.69)
Constant	32.93* (18.62)	40.70** (18.52)	15.79 (20.11)	44.44** (20.31)
Observations	3,232	3,232	3,232	3,231
R^2	0.13	0.08	0.19	0.09
Adjusted R^2	0.12	0.07	0.19	0.08
Residual Std. Error	22.87 (df = 3208)	22.75 (df = 3208)	24.69 (df = 3208)	24.95 (df = 3207)
F Statistic	20.84*** (df = 23; 3208)	11.73*** (df = 23; 3208)	33.06*** (df = 23; 3208)	13.00*** (df = 23; 3207)
Note:				*p<0.1; **p<0.05; ***p<0.01

† Compared to Asian Photos

†† Compared to Female Respondents

††† Compared to Asian Men Photos

Figure 16b: Linear Model - Qualities by Trait, Female Photos

	Dependent variable:			
	Trustworthiness	Competence	Attractiveness	Political Demeanor
	(1)	(2)	(3)	(4)
Estimated Age	−0.7359	−0.1905	−2.5338***	−0.2602
	(0.5354)	(0.5178)	(0.5826)	(0.5730)
Eye Shape	−0.1040	0.0491	−0.4982	0.1609
	(0.8757)	(0.8469)	(0.9530)	(0.9371)
Eye Angle	0.2277*	0.2264*	0.2436*	−0.0036
	(0.1347)	(0.1302)	(0.1466)	(0.1441)
Eye Distance	10.3992	−0.8230	−2.8239	4.0762
	(7.1129)	(6.8789)	(7.7404)	(7.6114)
Eyebrow Thickness	21.6533*	5.3448	−11.4398	5.4439
	(12.6513)	(12.2351)	(13.7674)	(13.5381)
Face Ratio	56.6804***	38.8602***	53.3361***	51.7515***
	(15.3344)	(14.8299)	(16.6871)	(16.4110)
Lip Thickness	0.5331**	0.4716**	0.4973*	0.4120
	(0.2485)	(0.2403)	(0.2704)	(0.2659)
Nose Ratio	−33.0948***	−13.8413*	−5.7790	−19.9523**
	(7.8824)	(7.6231)	(8.5778)	(8.4350)
Constant	−17.9742	5.5186	−14.5708	6.1667
	(24.6281)	(23.8179)	(26.8008)	(26.3568)
Observations	1,625	1,625	1,625	1,624
R^2	0.0329	0.0120	0.0370	0.0121
Adjusted R^2	0.0281	0.0071	0.0322	0.0072
Residual Std. Error	22.9217 (df = 1616)	22.1676 (df = 1616)	24.9438 (df = 1616)	24.5279 (df = 1615)
F Statistic	6.8702*** (df = 8; 1616)	2.4432** (df = 8; 1616)	7.7574*** (df = 8; 1616)	2.4790** (df = 8; 1615)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 16c: Linear Model - Qualities by Trait, Male Photos

	Dependent variable:			
	Trustworthiness	Competence	Attractiveness	Political Demeanor
	(1)	(2)	(3)	(4)
Estimated Age	0.0168	−0.3992	−3.7607***	−0.5439
	(0.4520)	(0.4475)	(0.4645)	(0.4852)
Eye Shape	−0.0767	−0.9761	−0.1480	−2.1314**
	(0.8202)	(0.8121)	(0.8429)	(0.8806)
Eye Angle	0.0772	0.2569	−0.5379**	0.3510
	(0.2249)	(0.2226)	(0.2311)	(0.2414)
Eye Distance	18.5051**	22.4750***	13.3057*	35.1033***
	(7.4601)	(7.3857)	(7.6663)	(8.0090)
Eyebrow Thickness	80.2804***	44.5466***	−8.7341	55.5250***
	(13.1376)	(13.0065)	(13.5007)	(14.1042)
Face Ratio	−9.8185	−11.0199	14.9990	−30.0646**
	(13.9937)	(13.8541)	(14.3805)	(15.0234)
Lip Thickness	0.5119**	0.4978**	0.3960	0.2467
	(0.2458)	(0.2433)	(0.2526)	(0.2639)
Nose Ratio	−24.4071***	−10.4307	−11.4269	−6.5553
	(7.9365)	(7.8573)	(8.1559)	(8.5205)
Constant	23.8724	12.1019	72.1177**	3.7763
	(28.6976)	(28.4113)	(29.4909)	(30.8092)
Observations	1,607	1,607	1,607	1,607
R^2	0.0501	0.0240	0.0563	0.0338
Adjusted R^2	0.0453	0.0191	0.0516	0.0289
Residual Std. Error (df = 1598)	24.1512	23.9103	24.8188	25.9283
F Statistic (df = 8; 1598)	10.5283***	4.9086***	11.9159***	6.9809***
Note:				*p<0.1; **p<0.05; ***p<0.01

Phase 3 Figures

Clarifying Note: The following figures, Figures 13a - 15b, evaluate whiteness as exclusively identifying as white, whereas Figures 10a - 12b evaluate whiteness as identifying as white regardless of additional racial/ethnic identification.

Figure 17a. How Respondents Believe Others Would Rate Them by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	POC Mean	Only White Mean
Trustworthiness	-1.346	217.477	1.8000e-01	-6.204	1.168	64.304	66.822
Competence	0.634	203.165	5.2700e-01	-2.776	5.409	63.557	62.240
Attractivess	-3.319	228.745	1.0500e-03	-10.947	-2.791	58.441	65.310
Political Demeanor	-1.008	208.762	3.1500e-01	-6.114	1.977	54.767	56.836

Figure 17b. How Respondents Believe Others Would Rate Them by Race Boxplots

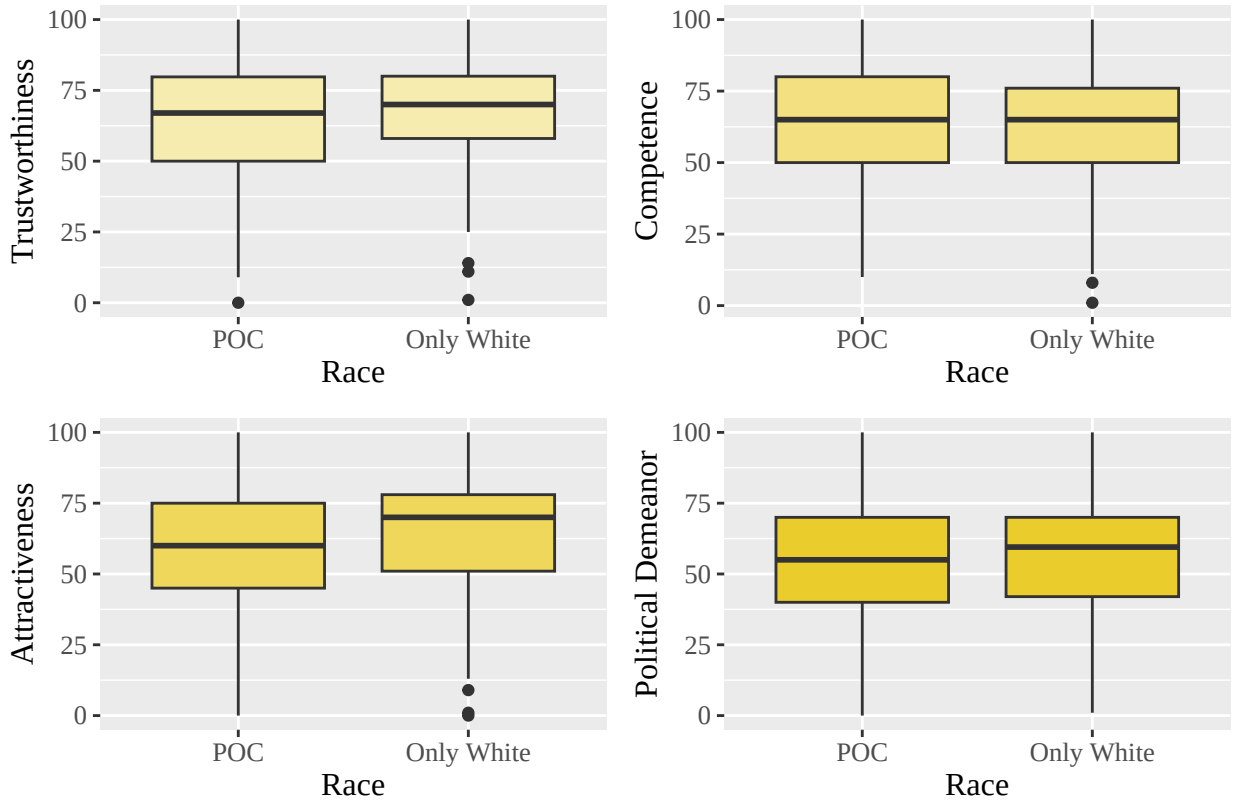


Figure 18a. Would Respondents Run by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	POC Mean	Only White Mean
Local	0.867	225.872	3.8700e-01	-0.046	0.118	0.329	0.293
State	0.146	221.710	8.8400e-01	-0.076	0.089	0.313	0.307
House	-0.519	216.236	6.0400e-01	-0.095	0.055	0.214	0.233
Senate	-0.596	214.776	5.5200e-01	-0.094	0.050	0.185	0.207

Figure 18a. Would Respondents Run by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	POC Mean	Only White Mean
President	0.198	224.881	8.4400e-01	-0.045	0.056	0.092	0.087

Figure 18b. Do Respondents Think they Could Win by Race t-Tests							
Trait	t Statistic	df	p-value	Lower CI	Upper CI	POC Mean	Only White Mean
Local	-2.854	215.865	4.7400e-03	-0.217	-0.040	0.371	0.500
State	-2.152	209.609	3.2600e-02	-0.177	-0.008	0.268	0.360
House	-1.225	208.028	2.2200e-01	-0.115	0.027	0.162	0.207
Senate	-1.321	202.276	1.8800e-01	-0.103	0.020	0.105	0.147
President	-0.632	205.910	5.2800e-01	-0.058	0.030	0.053	0.067

Figure 19a: Logistic Models - Running by Qualities and Race

	Dependent variable:					
	Run at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	-0.0106^{*}	0.0013	-0.0044	-0.0044	-0.0050	-0.0134
	(0.0059)	(0.0058)	(0.0058)	(0.0065)	(0.0067)	(0.0090)
Competence	-0.0057	-0.0092	-0.0020	0.0011	0.0027	0.0216^{**}
	(0.0062)	(0.0061)	(0.0062)	(0.0068)	(0.0071)	(0.0098)
Attractiveness	-0.0010	-0.0040	-0.0042	-0.0072	-0.0050	-0.0064
	(0.0046)	(0.0045)	(0.0046)	(0.0051)	(0.0053)	(0.0070)
Political Demeanor	0.0220^{***}	0.0097^{*}	0.0142^{**}	0.0211^{***}	0.0151^{**}	0.0177^{*}
	(0.0058)	(0.0056)	(0.0057)	(0.0065)	(0.0066)	(0.0091)
Race (Only White)	-0.4490^{**}	-0.4863^{**}	-0.3091	-0.0901	-0.0523	-0.2225
	(0.2059)	(0.2114)	(0.2102)	(0.2281)	(0.2363)	(0.3300)
Constant	0.2674	0.0368	-0.3893	-1.4153^{***}	-1.4735^{***}	-3.1266^{***}
	(0.3175)	(0.3125)	(0.3149)	(0.3530)	(0.3651)	(0.5312)
Observations	601	601	601	601	601	601
Log Likelihood	-400.0094	-406.4854	-403.9085	-354.5999	-335.6449	-215.1395
Akaike Inf. Crit.	812.0188	824.9709	819.8171	721.1998	683.2899	442.2790
Note:						*p<0.1; **p<0.05; ***p<0.01

Figure 19b: Logistic Models - Winning by Qualities and Race

	Dependent variable:					
	Win at All	Local	State	House	Senate	President
	(1)	(2)	(3)	(4)	(5)	(6)
Trustworthiness	-0.0158^{**}	-0.0095	-0.0109^{*}	-0.0174^{**}	-0.0158^{*}	-0.0125
	(0.0063)	(0.0059)	(0.0063)	(0.0073)	(0.0082)	(0.0112)
Competence	0.0077	0.0060	0.0078	0.0132^{*}	0.0165^{*}	0.0296^{**}
	(0.0065)	(0.0062)	(0.0065)	(0.0077)	(0.0088)	(0.0125)
Attractiveness	0.0094^{*}	0.0097^{**}	0.0105^{**}	-0.0064	-0.0008	-0.0065
	(0.0048)	(0.0046)	(0.0049)	(0.0057)	(0.0065)	(0.0088)
Political Demeanor	0.0315^{***}	0.0178^{***}	0.0221^{***}	0.0300^{***}	0.0182^{**}	0.0180
	(0.0062)	(0.0057)	(0.0061)	(0.0075)	(0.0083)	(0.0114)
Race (Only White)	0.3758	0.2042	0.1344	0.1345	0.2062	0.1235
	(0.2291)	(0.2102)	(0.2131)	(0.2439)	(0.2750)	(0.3822)
Constant	-1.2952^{***}	-1.2298^{***}	-2.1824^{***}	-2.3271^{***}	-2.8204^{***}	-4.4034^{***}
	(0.3424)	(0.3274)	(0.3600)	(0.4074)	(0.4777)	(0.7226)
Observations	601	601	601	601	601	601
Log Likelihood	-361.9190	-396.0522	-375.6739	-305.8210	-247.9923	-151.2801
Akaike Inf. Crit.	735.8381	804.1045	763.3477	623.6421	507.9846	314.5602

Note: *p<0.1; **p<0.05; ***p<0.01