

© Copyright 2018

Yiqin Alicia Shen

Gender Double Standard of Aging: How Perception of Age and Related Social
Outcomes are Modulated by the Gender of the Perceived

Yiqin Alicia Shen

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Yuichi Shoda, Chair

Ione Fine

Nancy Kenney

Program Authorized to Offer Degree:

Department of Psychology

University of Washington

Abstract

Gender Double Standard of Aging: How Perception of Age and Related Social Outcomes are Modulated by the Gender of the Perceived

Yiqin Shen

Chair of the Supervisory Committee:
Professor Yuichi Shoda
Department of Psychology

Understanding the influence of group labeling (e.g., gender, race, age) on social processes is one of the most important objectives of social psychology. Research in this tradition often studies the influence of one group label at a time; very little research has examined how multiple group labels (e.g., gender and age) simultaneously influence perception and social outcomes. Experiments in this dissertation combine behavioral methods and big-data analytics to examine (1) How does gender label of a target person influence age perception? (2) Are peoples' intentions to vote for a candidate biased by the candidate's age and gender? (3) How do gender and seniority correlate with representation in high impact journals?

TABLE OF CONTENTS

List of Figures	vi
List of Tables	vii
Chapter 1. Introduction	10
How Does Perceived Age Shape Impression Formation and Behaviors?	11
1.1.1 Age-Related Attitudes, Stereotypes, and Discrimination	11
1.1.2 Is There a Gender Double Standard?	13
How do People Perceive Age? Cues that Influence Age Perception.....	14
1.1.3 Physical and Social Cues Related to Perceived Age	14
1.1.4 Perceptions of Multiple Social Identities: The Case of Age and Gender	17
Individual Differences and the Highly-Repeated Within-Person Design.....	22
Chapter 2. Gender Double Standard of Age Estimation.....	23
Introduction.....	23
Study 1: Gender Moderate the Association Between Facial Features and Perceived Age.....	24
2.1.1 Method	25
2.1.2 Results.....	27
Study 2a: Gendered Names Influence Age Estimation	28
2.1.3 Method	28
2.1.4 Results.....	30
Study 2b: Gendered Pronouns Influence Age Estimation	33
2.1.5 Method	33

2.1.6 Results.....	34
Discussion.....	36
Chapter 3. Gender Double Standard of Aging in Hypothetical Elections	39
Introduction.....	39
Study 1: Gender and Age Interact to Predict Voting Intention.....	40
3.1.1 Results.....	43
Study 2: Gender and Age Interact Predicts Voting When Gender and Age Are not Salient....	46
3.1.2 Method	46
3.1.3 Results.....	47
Study 3: Gender and Age Interact to Predict Voting Intention in Realistic Voting Scenarios.	49
3.1.4 Method	49
3.1.5 Results.....	51
Study 4: The Role of Perceived Attractiveness, Competence, and Warmth	53
3.1.6 Method	53
3.1.7 Results.....	54
General Discussion	58
Chapter 4. Correlates of Gender and Age in the Real World: How Gender and Seniority Relate to Representation in High Impact Journals	60
Introduction.....	60
Method	63
4.1.1 Data Acquisition	63
4.1.2 Gender Determination.....	64

4.1.3	Analysis.....	64
	Result	65
4.1.4	Percentage of Female Authors and Journal Impact Factor	65
4.1.5	Comparisons to Base Rate	66
4.1.6	Change in Representation of Women Over Time.....	66
	Discussion.....	68
Chapter 5.	Summary & Final Remarks.....	72
5.1.1	Summary	72
5.1.2	Final Remarks	73
	Bibliography	74
Chapter 2	Appendix	82
	List of names used in Study 1	82
	Participant Gender Effect in Study 1	82
	Participant Gender Effect in Study 2a and Study 2b	83
	Independent Sample T-Test With Confidence Intervals for Study 2a and Study 2b.....	86
	Multilevel Analysis for Study 2a	87
	Cross-Classified Analysis for Study 2a	89
	Multilevel Analysis for Study 2b.....	91
	Cross-Classified Analysis for Study 2b	91
Chapter 3	Appendix	92
	Study s1: Voting Intention towards Younger vs. Older Candidates with the Same Gender	93

Study s2: Voting Intention towards Female vs. Male Candidates Approximately the Same Age	95
Study s3: Voting Intentions towards 93 Candidates with Profiles	96
Study s4: Ratings of the Initial Set of 365 Stimuli and Their Relations to Voting Intentions	100
The Effect of Participant Gender	103
Study 1	103
Study 2	106
Study 3	107
Study 4	109
The Effect of Participant Age	111
Study 1	111
Study 3	113
Lowess Plots with Bootstrapped 95% Confidence Interval, When Voting Intention is Grand Mean Centered	114
Multilevel Modeling of the Effects of Age on Voting Intentions in Study 3 without Mediators	115
Indirect Effects through the Quadratic Relationship between Perceived Age and Voting Intentions	117
Candidate as Unit of Analysis	119
Pairings of Stimuli Used in Study s2 and Study s3	121
Common Names Used in Study s3	122
Chapter 4 Appendix	123
The proportion of Ungendered and Uncertainly Gendered Authors	123

Results after Thresholding to Authors with High Gender Certainty 124

LIST OF FIGURES

Figure 1.1. Schemas of relationships tested in this dissertation.	11
Figure 2.1. Example morph continuums.	26
Figure 2.2. The relationship between estimated age and morph continuum going from very young (1) to very old (7).	28
Figure 2.3. Example androgynous faces varying in perceived age and apparent race.	30
Figure 2.4. The effect of labeling a face as female versus male, for all dependent variables (dots represents the mean, and horizontal lines represent 95% CI).	32
Figure 2.5. The effect of labeling a face as female versus male, for all dependent variables (dots represents the mean, and horizontal lines represent 95% CI).	36
Figure 3.1. Example stimuli: female and male political candidates from a wide age range.	41
Figure 3.2. The relationship between perceived age and voting intention.	44
Figure 3.3. Predicting voting intention in Study 2 using age estimation made by participants of Study 1 (Lowess smoothing with 95% CI).	47
Figure 3.4. Percentage of participants who voted for the younger candidate over the older candidate by the candidate pairs' gender.	52
Figure 3.5. The relationship between voting intention and perceived age with 95% bootstrapped CI (Study 4).	54
Figure 3.6. Summary of within-subject mediation analyses of the effects of perceived age on voting intention thorough perceived warmth, perceived attractiveness, and perceived competence (Study 4).	58
Figure 4.1. Percentage of female first and last authors between 2005-2017 vs. Journal's 5-year impact factor	65
Figure 4.2. Change in the percentage of female first (Panel A) and last (Panel B) authors over time.	67

LIST OF TABLES

Table 1. Multilevel Modeling Predicting Voting Intention as a Function of Candidate Age and Gender	45
Table 2 Multilevel Modeling Predicting Voting Intention as a Function of Candidate Gender and Age as Estimated by A Separate Group of Participants	48
Table 3 Indirect Effects through Perceived Attractiveness, Perceived Competence, and Perceived Warmth	57
Table 4 Percentage Change for Female First and Last Authors Per Year	68

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisors, Dr. Yuichi Shoda and Dr. Ione Fine, who mentored and cared for me throughout my Ph.D. training, this dissertation would be impossible without you. Thanks also to my friends from the psychology department, especially Kelly Chang and Carol Xu, for sharing your knowledge and your humor with me, helping me pull through the most difficult of times. I also want to thank my parents, Shufen Chen and Jiangtao Shen, for loving me, supporting my choice, and believing in my potential. Last but not least, I would like to thank Dr. Jason Webster, for being my dissertation writing buddy, code debugger, co-author, friend, companion in adventure, and life partner through all those years. Your strength helps to hold everything together.

The research described in this dissertation was made possible by computing resources at Center for Studies in Demography and Ecology (CSDE), and is partly supported by the Bolles fund and Alcor award from Department of Psychology, University of Washington.

DEDICATION

To Susan Sontag, may your words always give me strength.

Chapter 1. INTRODUCTION

Shortly after speculation began that Hillary Clinton would be a candidate in the 2016 presidential election, political pundits began to comment on her age. Some raised the concern that Ms. Clinton might no longer have the energy level to succeed as the president; others even questioned whether she would live to the end of her term. Ms. Clinton's age has a negative impact on people's perception of her suitability as the president. However, several male presidential candidates were significantly older than Ms. Clinton when they ran for president, but their age did not become a point of public scrutiny during their campaigns. *Is this just a special case, or is her gender influencing how people make inferences based on age?* Furthermore, for individuals less well known than Hillary Clinton, knowledge about a person's age is usually not immediately available and has to be estimated from age-related cues. Age-related stereotypes and discrimination faced by the individual are often based on the appearance of age-related cues. *Does gender also influence how people make inferences based on age-related cues?* This dissertation will systematically examine those two questions.

More specifically, Chapter 2 examines whether age perception is moderated by the person's perceived gender. Chapter 3 examines whether the decrease in electability with age occurs at the same rate for male and female candidates, and the mediating role of perceived competence, warmth, and attractiveness in this process. The relationship tested in Chapter 2, and Chapter 3 could be visually summarized in Figure 1.1. The following sections review literature relevant to those two questions respectively.

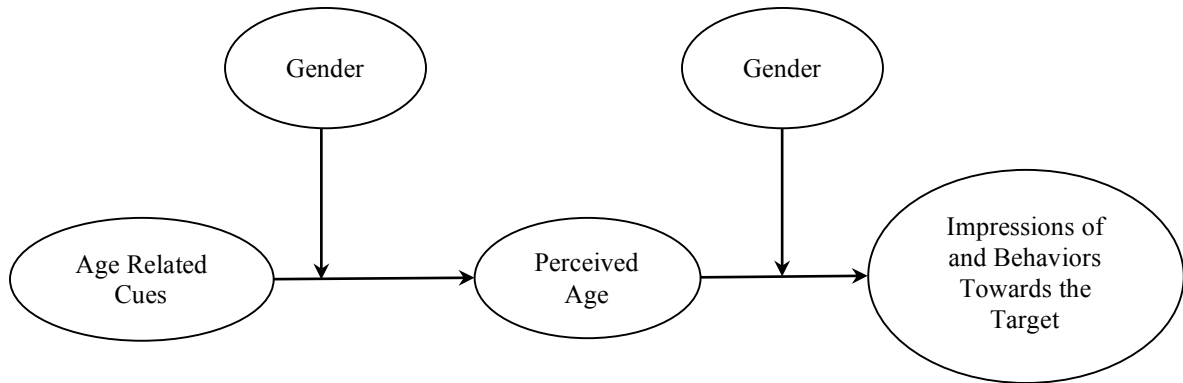


Figure 1.1. Schemas of relationships tested in this dissertation.

HOW DOES PERCEIVED AGE SHAPE IMPRESSION FORMATION AND BEHAVIORS?

1.1.1 *Age-Related Attitudes, Stereotypes, and Discrimination*

In American culture especially, youth is often held in high regard, while old age is often considered undesirable (B. Levy & Langer, 1994). Kite & Johnson (1988) conducted a meta-analysis reviewing attitudes towards older adults as compared to younger adults. They found an overall medium effect favoring younger adults. However, the effect is highly variable across domains. For instance, studies using competence as dependent variable have larger effect sizes as compared to studies using personality traits as dependent variables, which is congruent with common stereotypes of the elderly as being low in competent. Besides explicit preference, people also have an implicit preference for youth over elderly. Both younger and older adults are faster in associating pictures of young adults with positive words, and pictures of older adults and negative words, on average showing a strong implicit preference for young over old (Levy & Banaji, 2002).

While subjects tend to have a more positive attitude towards young adults as compared to older adults, research on stereotypes reveal a more complicated story. Kite, Deaux, & Miele (1991) asked subjects to rate older people on agentic and communal characteristics and found that older individuals to have less agentic traits, and equally likely to have communal traits as compared to younger adults. In other studies involving trait perceptions, elder adults are perceived to be less ambitious and less competent (Singer, 1986), but friendlier (Andreoletti, Maurice, & Whalen, 2001). They are also found to be less competent but warmer than other social groups (“doddering but dear”) (Cuddy, Norton, & Fiske, 2005), which elicits feelings of pity and sympathy (Fiske, Cuddy, Glick, & Xu, 2002). Aside from stereotypes, many (positive and negative) subtypes are associated with old age. Those subtypes include elderly statesman, senior citizen and grandmotherly (Marilynn B Brewer, Dull, & Lui, 1981), perfect grandparents, John Wayne conservative, severely impaired, and despondent (Hummert, 1990), suggesting that stereotypes associated with older adults are inherently complex.

While those mixed stereotypes and subtypes might be beneficial for the elderly in scenarios that require high warmth (Kang & Chasteen, 2009), it could be especially detrimental in settings that require high competence. In audit study, older and younger adults with the same qualifications applied for entry-level positions. The elderly applicants received less favorable responses 41.2% of the time, most of the times without any justification (Bendick, Brown, & Wall, 1999). Under the assumption that older adults are less competent, participants display age accommodating patterns of communication (i.e., “baby talk”), highlighting paralinguistic features and reduction in syntactic complexity (Kemper, Vandepute, Rice, Cheung, & Gubarchuk, 1995; Rubin & Brown, 1975; Thimm, Rademacher, & Kruse, 1998).

Another scenario that requires high competence is politics. Perceived competence is found to be especially predictive of outcomes in political campaigns (Todorov, Mandisodza, Goren, & Hall, 2005). Considering that the elderly are stereotyped to be especially incompetent, one would expect a high preference for young over old in politics. However, empirical results remain inconsistent. Some studies found a preference for younger candidates over old ones. For instance, Sigelman & Sigelman (1982) tested people's voting behavior towards hypothetical candidates that vary in race, gender, and age. Participants were more likely to vote for young White male as compared to middle-aged White male. They are also more likely to vote for middle-aged White males as compared to older White males. However, Todorov and colleagues found that older candidates are not necessarily disadvantaged: the older candidates are perceived as being more competence (Olivola & Todorov, 2010), and are marginally more likely to win (Todorov et al., 2005)¹. With those conflicting results, further research is needed to clarify the influence of candidates' age on voting intentions.

1.1.2 *Is There a Gender Double Standard?*

Majority of past literature on age-related stereotypes and discrimination left gender of the target unspecified. Only a few studies have examined the influence of gender on the perception of older vs. younger adults. Deusch, Zalenski, & Clark (1986) found that photos of older men are rated as more attractive than photos of older women, while middle-aged men and women are rated as being equally attractive. Positive age subtypes (e.g., John Wayne conservative, perfect grandparent) are more likely to be applied to older males as compared to older females, except for old-old targets who are more than 80 years of age. However, it is not known whether the

¹ In Todorov et al. (2005), age did not predict election outcomes in races matched in sex and ethnicity.

gender difference is due to the gendered nature of some of those subtypes (e.g., John Wayne conservative) (Hummert, Garstka, & Shaner, 1997). While most evidence suggests that participants rated aging in females more harshly, O'Connell & Rotter (1979) found that subjects thought aging is more detrimental to males than to females.

Regarding voting behaviors, research has shown that voters use target's gender as well as age to make inferences about their traits (Huddy & Terkildsen, 1993; McDermott, 1997; Sanbonmatsu, 2002; Sigelman & Sigelman, 1982; Todorov et al., 2005). However, it is not known whether gender interacts with age to influence voting behaviors. Todorov et al. (2005) examined (as a control variable) the influence of age on election outcomes, but only in a heavily male-dominant stimulus set. Sigelman & Sigelman (1982) didn't fully cross age and gender, thus making it impossible to examine the interaction effect between gender and age. Though not examining perceived age per se, Poutvaara, Jordahl, & Berggren (2009) found that babyfacedness is positively correlated with electoral success for male candidates but not female candidates, providing some initial evidence that youthful features influence electoral success in females and males differently. More research is clearly needed to examine how gender moderates the relationship between perceived age and voting intentions and outcomes.

HOW DO PEOPLE PERCEIVE AGE? CUES THAT INFLUENCE AGE PERCEPTION

1.1.3 *Physical and Social Cues Related to Perceived Age*

The age of a person is not immediately perceptually available. While age can be estimated from various physical and social cues, the aggregation of such information is not free of bias. Which features are relevant to the perception of age and how are they aggregated to form impressions? This section first discusses the physical cues and social cues relevant to age

perception and how they are aggregated, it then discusses how gender would influence this process drawing from theories on how people perceive targets with multiple social categories.

Physical Cues.

While people can successfully use cues such as voice (Shipp, Qi, Huntley, & Hollien, 1992), movement patterns (Ketcham, Seidler, Gemmert, & Stelmach, 2002), and even body odor (Mitro, Gordon, Olsson, & Lundström, 2012) to estimate age, the visual appearance of facial features exhibited by an individual are by far the most widely utilized cues for age estimation, and consequently the most widely studied (Rhodes, 2009).

The appearance of the skin greatly influences perceived age. Winkles and sagging (Mark et al., 1980), increase in pigment irregularities (Matts, Fink, Grammer, & Burquest, 2007), loss of skin reflectance and homogeneity (Matts et al., 2007), visibility of varicosities (Enlow, 1982), and lower contrast between features and the surrounding skin (Porcheron, Mauger, & Russell, 2013) are all negatively related to perceived age.

Head shape and structure of the face also influence age estimation. The most well studied age-related head shape/structure change is cardioidal strain (Pittenger & Shaw, 1975), which increases most prominently before the age of 20. Characteristics of low strain, which are usually associated with earlier stages of development, include a round, protruding skull casing, nose situated at a lower position in the face, and a smaller chin. Independent of head shape, features such as bigger eyes and shorter nose are also associated with younger age (Berry & McArthur, 1986). Additionally, an increase in facial width to height ratio has also been linked to increased perceived age (Hehman, Leitner, & Freeman, 2014).

While a wealth of research focuses on single features, there is very little theoretical work linking individual physical features to global age estimation. Voelkle, Ebner, Lindenberger, & Riediger (2012) used a linear paramorphic model to simulate the age estimation process. In the linear paramorphic model, estimated age is considered a linearly-weighted combination of all age-related visual features, such that each feature varies in how well it predicts true age (e.g., features such as wrinkles may be especially diagnostic, while features such as grey hair might have greater variability between individuals of the same age). Based on this model, individuals demonstrating more youthful features are more likely to be perceived to be young, while individuals who demonstrate more elderly features are more likely to be categorized as old.

This model represents a data-driven (as opposed to theory-driven) approach towards age estimation. As the authors of this model wrote, the main goal of this approach is not to understand how people estimate age, but how to “simulate behaviors” (Voelkle et al., 2012). Other more complicated algorithms have been developed to optimize the ways features are extracted from images and weighted to predict actual facial age (Geng, Zhou and Smith-Miles, 2007).

Social Cues

While it is often assumed that age-related stereotyping happen after age perception, some argue that age stereotypes can influence age perception itself. Connectionist models, such as the parallel constraint satisfaction model (Kunda & Thagard, 1996) and the dynamic interactive model (Freeman & Ambady, 2010), argue that targets with more stereotypical characteristics increase the activation of the corresponding social category, thus increasing the likelihood of categorization.

The influence of stereotypes on categorization is evident in the domain of race. Individuals are judged to move from one racial category to another based on their social class (Penner & Saperstein, 2008). Targets were more likely to be categorized as Black if they experience downward social migration, such as dropping below the poverty line, were incarcerated or became unemployed, which are circumstances considered more stereotypically Black. Racially ambiguous targets wearing higher status attires have also been experimentally shown to be more likely categorized as White, reflecting the stereotype that Whites are often associated with higher social status occupations (Penner & Saperstein, 2008).

While there is little research on the influence of social cues on age perception, based on those theories and experimental results, it could be predicted that individuals demonstrating more stereotypically elderly stereotypes and behaviors are more likely to be grouped into the category of the “elderly”, while individuals who are congruent with more youthful stereotypes and behaviors are more likely to be categorized as “young”.

1.1.4 *Perceptions of Multiple Social Identities: The Case of Age and Gender*

Before delving into the intersectionality between gender and age, this section first reviews existing theoretical frameworks on this topic to guide our exploration. Intersectionality refers to the modification of psychological meaning attached to a social category depending upon membership in crosscutting social categories (Kang & Chasteen, 2009). As compared to the rich literature on social categorization along a single dimension, the research on intersectionality is still in its infancy. Currently, there is a lack of agreement about whether people even make use of multiple social categories. There are even more unknowns about how they make use of information from multiple social categories.

Category Dominance Hypothesis

The category dominance hypothesis is based on the premise that people, in general, prefer simple, unidimensional forms of social impressions. According to the category dominance hypothesis, multiple salient categories compete for the dominant category, and the category that is most recently or frequently used, most relevant to the perceiver's goal, or most contextually salient become the dominant category (Bodenhausen, 2009). Once the dominance is established, if the task demand does not specifically ask for "dual categorization" (Ross & Murphy, 1996), other social dimensions will be actively suppressed (Hugenberg & Bodenhausen, 2004).

This hypothesis received empirical support in the domain of stereotype activation. Macrae, Bodenhausen, Milne, & Jetten (1994) presented participants with multiply categorizable targets (e.g., Chinese women), and primed targets to either focus on the category "Chinese" or "women." Subjects subsequently showed increased performance in the lexical-decision-making task in the primed category and showed decreased performance in the unprimed category. Similarly, when a Black doctor confederate gave positive feedback to the subjects, subjects tend to inhibit Black stereotypes and perform better on lexical decision tasks related to "doctors." (Sinclair & Kunda, 1999). However, one might argue that subjects must make use of information from the non-dominant category to successfully suppress the non-dominant dimension ("they must somehow know that she is a female to successfully suppress female stereotypes") (Quinn & Macrae, 2005), thus it is unclear whether category dominance happens in earlier or later stages of processing.

Following the logic of the category dominance hypothesis, at the intersectionality of gender and age, when age is made salient through explicitly asking subjects to evaluate the target along the dimension of age, gender information should be suppressed/ignored.

Connectionist Models

While the category dominance hypothesis argues that the non-dominant social category is simply ignored, connectionist models propose that multiple social categories can be simultaneously activated (Freeman & Ambady, 2010). Moreover, the activation of nodes within one social category (e.g. Black, Asian within the social category of race) could both increase and decrease the activation level of nodes in other social categories (e.g. female and male within the social category of gender), depending on whether those two nodes share common stereotypes or phenotypes. Thus rather than settling on a dominant dimension and ignoring non-dominant dimensions, information from multiple dimensions get used dynamically to influence social categorizations.

This view has been supported by research on the intersectionality between race and gender. Participants are more efficient in making judgments about the gender of faces when the race category of the face shares phenotypes or stereotypes with the correct sex category. For instance, as Asian and females share both phenotypes and stereotypes (e.g., docile), gender categorization of Asian female faces is more efficient than for Asian male faces or Black female faces. Similarly, Black and male share phenotypes and stereotypes (e.g., aggressive), gender categorization of Black male faces are more efficient as compared to Black female faces or Asian male faces (Johnson, Freeman, & Pauker, 2012). However, as a relatively newly proposed model, the dynamic interactive model has yet to be tested in the context of gender age interaction.

It remains to be seen whether interactionist predictions from such models would generalize to those two dimensions.

Does Gender Influence Age Estimation Process?

There is preliminary evidence suggesting that people estimate age differently for targets of different genders. When people were asked to attribute age categories (e.g., middle-aged, elderly) to pictures of faces, women were perceived to enter the next life stage earlier than men (Kogan, 1979). However, in a separate series of studies, when participants were asked to estimate age from another set of faces, the age of women were underestimated by 1.8 years, while men's age was overestimated by about 0.6 years (Voelkle et al., 2012). The finding that women's age tend to be underestimated, while also being grouped into the next age category earlier seems to suggest that there is a stricter standard for women when applying age categories to faces.

An alternative way gender could influence age estimation is through associations as described in connectionist models. Freeman and Ambady (2011) pointed out two major routes through which gender could bias age perception: through the top-down process of shared stereotypes and the bottom-up process through shared phenotypes.

Through the top-down route, stereotypes associated with a cross-cutting category influence the likelihood and efficiency of categorization in the other social category. In the case of age and gender, the large number of stereotypes associated with each category makes it difficult to make predictions about covariation. One way to simplify this process is to use known models that simplify stereotype content through mapping them on to two dimensions: competence and warmth (Fiske, Cuddy, Click and Xu, 2002). According to the stereotype

content model, the elderly and women are both associated with the stereotypes of being high on warmth and low on competence. Thus through the top-down route, women might be more likely to be perceived as older due to the shared stereotype content with the elderly.

Through the bottom-up route, phenotypes associated with members of a cross-cutting category influence the efficiency of categorization in other social categories. Johnson, Freeman, and Pauker (2011) examined the influence of race on gender categorization through measuring the similarity of facial features of race and gender using facegen. They found that Asian faces are on average more female-typed, while Black faces are more male-typed. More recently, Carpinella, Chen, Hamilton, & Johnson, (2015) examined the influence of gendered facial cues on race categorization within the same framework. They used both facegen generated images and photographs of real faces. The biasing effect of gender on race was found using facegen stimuli, though subjects were not more efficient in categorizing Asian stimuli as females using real photographs.

In sum, based on previous research, it is hard to form a quantitative measurement of covariation between age-related facial features and gender-related facial features. One way to simplify this process is to also rely on the algorithm provided by facegen and manipulate age as a cluster of features obtained based on linear models constructed from a large number of 3D laser scanned faces. However, since older adults are severely under-represented in (Banz & Vetter, 1999) (i.e., facegen) database, such an approach is unlikely to produce generalizable results. Considering the lack of a satisfactory tool to manipulate multiple facial features, the current research will not directly address gender age feature co-variation.

INDIVIDUAL DIFFERENCES AND THE HIGHLY-REPEATED WITHIN-PERSON DESIGN

While on average individuals may show stronger age-related discrimination towards older women than men, the effect might be stronger for some participants compared to others. One method recently developed to test such individual differences is the Highly-Repeated Within-Person Design (Whitsett & Shoda, 2014). In a Highly-Repeated Within-Person Design, each individual responded to a large number of stimuli that differ in the variable of interest (and other dimensions). The relationship between a situational feature and each individual's response is analyzed separately for each individual. As a result, researchers will be able to identify the amount of individual variation along the dimension of interest. If a certain pattern of behavior is stable within an individual, and significant individual variations are found between individuals, further analysis can be done to identify individual differences dimensions that predict such patterns of responses. We will employ the Highly-Repeated Within-Person design to examine the two main research questions.

Chapter 2. GENDER DOUBLE STANDARD OF AGE ESTIMATION

Shen.Y.A., & Shoda.Y. (2018). The Gender Double Standard of Aging: The Surprising Effect of Gender on Facial Age Estimation. Manuscript in preparation to be submitted to Psychological Science.

INTRODUCTION

The social consequences of being perceived to be of a certain age (e.g., old) has been well documented. But how do people perceive others age to begin with? Most research has assumed that age is largely a simple function of sensory cues such as facial appearance. The features that have been found to be predictive of perceived age include the texture of the skin, looseness of skin along the jawline, length of the nose and ears, the thickness of the eyebrows, size of the eye, and color and quantity of hair (Berry & McArthur, 1986). However, vision research has documented that perception is considerably more complicated than simply encoding retinal images (Bruce & Young, 1986; Gilbert & Li, 2013). Is it possible that perception of a person's age is also more than simply aggregating age-related visual cues? Specifically, could the presumed gender of the target person moderate age perception?

While an emerging line of research has started to reveal how contextual factors can influence the race categorization process (Eberhardt, Dasgupta, & Banaszynski, 2003; Krosch & Amodio, 2014), to our knowledge, no research to date has examined how the age of the perceived is influenced by their perceived gender. Yet, anecdotal evidence suggests that gender of a person may influence how old they are perceived to be. For example, Susan Sontag (Sontag, 1972) suggests in her essay, “double standard of aging,” that there is a greater social expectation for women to maintain a youthful appearance than men. If so, the same person may be perceived

to be older if she is believed to be a woman rather than a man. Similar predictions stem from the dynamic interactionist theory (Cloutier, Freeman, & Ambady, 2014; Freeman & Ambady, 2011), which argues that shared stereotypes between categories cause stronger associations between these categories. Since the category “female” and “old” share stereotypes such as high in warmth and low in competence (Cuddy et al., 2005), it might cause these categories to be more strongly associated with each other.

It has been difficult to test this hypothesis because manipulating perceived gender by the physical features of the perceived can inadvertently change age-related features as well. To overcome this confound, the present studies examined the perceived age of computer-generated faces whose gender-related features are ambiguous, such that they can be perceived as either female or male. Each of these images was presented either as a woman or as a man, and the corresponding perceived age was assessed. Specifically, we tested whether the estimated age of the people shown in these images was believed to be older when the face was presented as a female versus male. Of course, this is only a first step towards testing the more general hypothesis about the effect of gender perception on age perception. However, it allows the use of stimuli that are physically identical and can provide clear evidence that perceived gender can, at least under some circumstances, influence perceived age.

STUDY 1: GENDER MODERATE THE ASSOCIATION BETWEEN FACIAL FEATURES AND PERCEIVED AGE

The purpose of Study 1 is to investigate how changes in age-related facial features correlate with perceived age for men and women. Through morphing young and old faces that were either biologically the same age, or perceived to be the same age, we created an “artificial”

aging process. We hypothesize that, in this “artificial” aging process, the perception of women’s age will increase earlier in the “lifespan” as compared to men.

2.1.1 *Method*

Stimulus Generation

All source images used in this experiment were obtained from the Park Aging Mind stimulus set (Minear & Park, 2004). This stimulus set contained color images of females and males across the lifespan as well as the actual age of the person when the picture was taken.

Pictures of 8 younger females, eight older females, eight younger males, and eight older males were selected from this stimulus set. Half of the female and male faces were equated on their *perceived age* (e.g., as rated by a group of pilot participants, the younger female and male faces were both *perceived to be* 20 years old, and the older female and male faces were both *perceived to be* 70 years old). The other half of the female and male faces were equated on their *actual biological age* (e.g., the younger female and male faces were both 20 years old, and the older male and female faces were both 70 years old).

The eight pairs of younger and older faces were then cropped to include only the section above the upper lip to just above the eyebrows, and between the ends of brow ridge. Each pair was then morphed to create a 7-degree morph continuum from young to old using FantaMorph software (<http://www.fantamorph.com>). Examples of these morph continuums are presented in Figure 2.1.



Figure 2.1. Example morph continuums.

Panel (A) and panel (B) have end faces matched on actual age, panel (C) and panel (D) have end faces matched on perceived age.

Participants

Participants were 134 MTURK workers (59% males; 81% Whites; $M_{\text{age}} = 42.77$, $SD = 13.24$). All workers had IP address within the United States.

Procedure

Participants signed up for a study on “impression formation from faces.” To make sure that the same participant never saw two faces from the same morph continuum, each participant saw eight target faces, with one pseudorandomly selected from each morph continuum. Those eight morphed faces were embedded in 116 none-morphed faces from the Park Aging Mind stimulus set. Participants provided two-digit age estimation for each of the 124 target faces. After all face trials, participants self-reported their gender, age, and race/ethnicity.

2.1.2 Results

We averaged participants' age estimation as a function of the morph continuum (1= youngest, 7 = oldest) separately for pairs that are equated on perceived age and actual age.

For pairs that are matched on actual age, the mean at each morph point and the between-person 95% confidence interval are presented in Figure 2.2 Panel A. While there are no significant differences in perceived age at any point along the morph continuum, there is a significant interaction between face gender and morph point 2 versus morph point 3, $F(1,152) = 7.85, p = .006$. While males are estimated to be older at age point 2, females are estimated to be older at age point 3.

Similar interaction effects were found in pairs that are matched on perceived age (Figure 2.2 Panel B). While there is no significant difference in perceived age between female and male faces at any age point along the morph continuum, there is a significant interaction effect between face gender and age point 2 versus age point 3, $F(1,152) = 5.40, p = .021$. These results suggest that, in the artificial "aging" process created by morphing, women are perceived to start aging earlier than men.

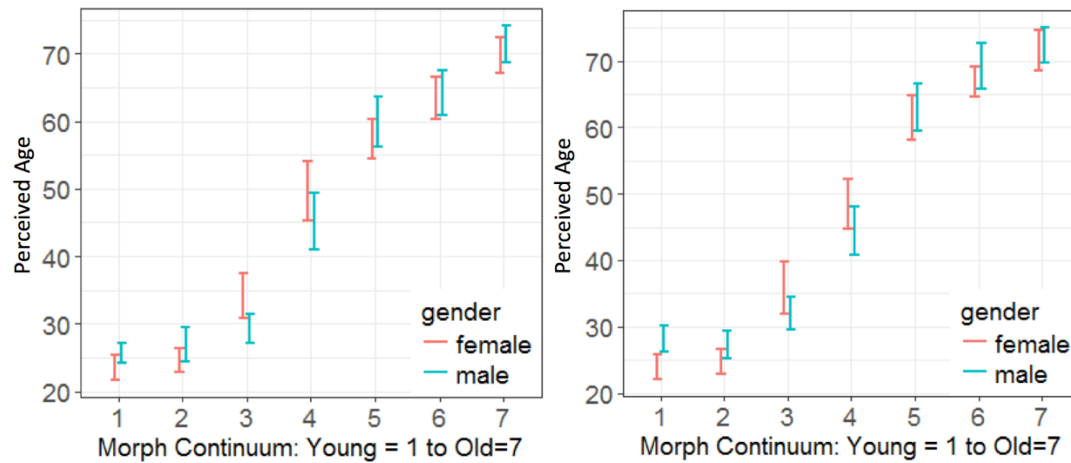


Figure 2.2. The relationship between estimated age and morph continuum going from very young (1) to very old (7). Left panel is morph pairs with end faces matched on actual age, right panel is morph pairs with end faces matched on perceived age. Bars represent 95% CI.

STUDY 2A: GENDERED NAMES INFLUENCE AGE ESTIMATION

Participants were asked to estimate the age of people whose images were referred to with either prototypically female or male names depending on the experimental condition.

2.1.3 Method

Stimuli Generation

Sixteen androgynous faces were generated using the face generation function in Facegen Modeler (www.facegen.com), with the gender slider set halfway between “stereotypically female” and “stereotypically male.” Faces from this stimulus set had features rated as stereotypical of a variety of racial and age backgrounds: four faces were stereotypically White, six were stereotypically Black, and six were stereotypically Asian in appearance according to the Facegen race slider. Faces also represented targets from a wide age range, between age 20 to age 60 on the Facegen age slider.

A total of 16 female and 16 male names were randomly selected from the “top 100 names” from the social security database (for a full set of names and stimuli see appendix).

Participants

195 Amazon Mechanical Turk (MTURK) workers (56% males 43% females) participated in exchange for monetary compensation. Participants’ ages ranged from 18 to 71 years ($M = 38.07$, $SD = 12.37$). All workers had IP address within the United States.

Procedure

The 16 face stimuli were evenly divided into two sets, set A and set B, with an equal number of young and old faces, and faces from each racial group, in each stimulus set. Pairing of face sets with gender was counterbalanced between participants, such that one group of participants ($n = 99$) viewed set A paired with female names and set B paired with male names (“Hi my name is <female name/male name>, nice to meet you”), while the other group of participants ($n=96$) viewed set A paired with male names and set B paired with female names.

With the face and name of each target person on top of the screen, participants were asked to 1) estimate the perceived age on an eight-point scale (“Does <female name/male name> appear young or old to you?”, 0 = “Very Young”, 7 = “Very Old”), 2) provide a two-digit age estimation (“How old do you estimate <female name/male name> to be?”), 3) rate the perceived attractiveness of the target on an eight-point scale (“How attractive do you think <female name/male name> is?”, 0 = “Very Unattractive”, 7 = “Very Attractive”), 4) rate the perceived trustworthiness of the target on a 8 point scale (“How trustworthy do you think <female name/male name> is?”, 0 = “Not Trustworthy At All”, 7 = “Very Trustworthy”). All stimulus presentation and answers collection were conducted using Qualtrics software.



Figure 2.3. Example androgynous faces varying in perceived age and apparent race.

2.1.4 Results

Mean Effect of Pairing a Face with Female versus Male Names

We first conducted two sets of independent t-tests, one for stimulus set A, and the other, for stimulus set B.

For stimulus set A, averaging across all stimuli within the set, when they were paired with female names, participants rated them as significantly older than when the same faces were paired with male names (female names: $M = 3.76$, $SD = 0.56$; male names: $M = 3.52$, $SD = 0.47$; $t(193) = 3.18$, $p = .002$, 95% CI = [0.09, 0.38]). Participants estimated the age of the stimulus person presented as female to be approximately 2 years older than when they were paired with male names (female names: $M = 38.46$, $SD = 5.63$; male names: $M = 36.34$, $SD = 4.76$; $t(193) = 2.84$, $p = .005$, 95% CI = [0.64, 3.59]).

Results from stimulus set B confirmed this result. For stimulus set B, on the eight-point age scale, participants rated faces paired with female names ($M = 3.85$, $SD = 0.48$) to be older than those paired with male names ($M = 3.61$, $SD = 0.53$), $t(193) = 3.31$, $p = .001$, 95% CI = [0.09, 0.38]. Faces paired with female names ($M = 39.19$, $SD = 5.48$) were also perceived to be approximately 2 years older than when paired with male name ($M = 37.03$, $SD = 5.26$), $t(193) = 2.80$, $p = .006$, 95% CI = [0.64, 3.67].

In contrast, pairing androgynous faces with either female or male name has no consistent effect on perceived attractiveness or trustworthiness. The only exception is that, for stimulus set A, participants rated faces paired with female names ($M = 3.50$, $SD = 0.85$) as being more attractive than the same faces paired with male names ($M = 3.22$, $SD = 0.96$), $t(193) = 2.16$, $p = .032$, 95% CI = [0.02, 0.54]. However, this effect failed to replicate with stimulus set B. For both stimuli set A and set B, pairing with the female or male name has no significant influence on perceived trustworthiness.

Stimulus Level Variations

To further examine whether the effect of gender name pairing on perceived age is uniform across all 16 exemplars, we repeated the above mentioned independent sample t-tests separately for each stimulus. The results are summarized in Figure 2.4.

When subjects rated faces on 8-point age estimation ranging from “very young” to “very old,” 15 out of the 16 stimulus faces appeared older when paired with female rather than male names (see Figure 2.4 upper left panel). Under the null hypothesis that pairing a face with female or male names has no effect on perceived age, the probability of observing a pattern of the result as extreme as the current one is 0.0005 (same as tossing a fair coin 16 times and observing 15 or more heads). Similar results were found for 2-digit age estimation, the probabilities of observing a positive effect for 14 or more stimulus faces were 0.004 (see Figure 2.4 upper right panel). Thus, both patterns of results are very unlikely to have occurred due to chance. This suggests that pairing a face with female versus male names influences perceived age in a way that is consistent across a majority of the stimuli.

Conversely, when the same calculations are applied to perceived attractiveness, given the null hypothesis is true, there is a 0.5 probability of obtaining a positive effect for 8 or more stimuli (see Figure 2.4 lower left panel). Similarly, for perceived trustworthiness, there is a 0.45 probability of observing a positive effect for 10 or more stimulus faces (see Figure 2.4 lower right panel).

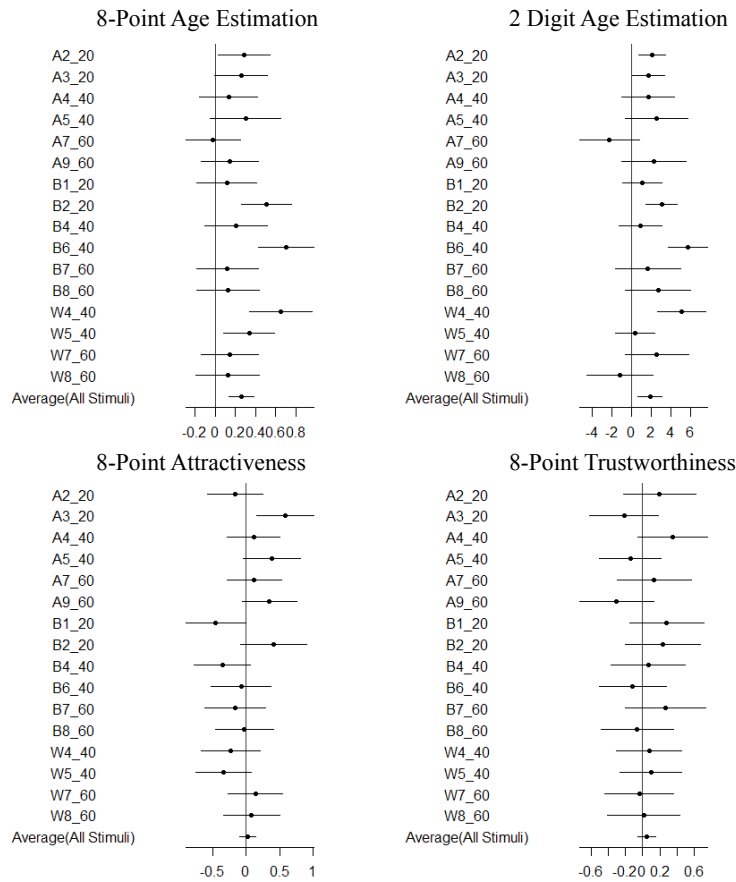


Figure 2.4. The effect of labeling a face as female versus male, for all dependent variables (dots represents the mean, and horizontal lines represent 95% CI). For stimulus labels, “A” represents stereotypical Asian faces, “B” represents stereotypical Black faces, and “W” represents stereotypical White faces (classified according to facegen modeller). The number after the underscore represents the approximate age range in which the faces are sampled from (according to facegen modeller).

STUDY 2B: GENDERED PRONOUNS INFLUENCE AGE ESTIMATION

Study 2a showed that androgynous face paired with a prototypically female name tended to be judged as older than when the same face is paired with a prototypically male name. However, it is possible that the female and male names used in Study 2a conveyed subtle age cues, causing participants to under or overestimate age. Study 2b is a pre-registered conceptual replication of Study 2a, designed to overcome the limitation above by using gender pronouns in place of gender-typical names.

2.1.5 *Method*

Stimuli Selection

The androgynous stimuli used in Study 2b were identical to those in Study 2a, except for the addition of two typical gender faces. The two gender typical faces were generated using Facegen Modella, using the same procedure specified in Study 2a, with age slider set to 40, race slider set to stereotypically Caucasian, and the gender slider set to either typically female or typically male.

Participants

One hundred and ninety-seven Mturk workers were recruited for this experiment (50% males, 50% female). Participants' age range from 20 to 66 years ($M = 36.35$, $SD = 10.49$). According to rules specified in the pre-registration, eight subjects were excluded due to excessively short (shorter than 3 minutes), or excessively long (longer than one hour) completion time. All workers had IP address within the United States.

Procedure

The procedure was identical to that specified in Study 2a except for the following differences. In Study 2b, instead of pairing each face in the two faces sets (Set A and Set B) with either female or male names, they were referred to with either the female gender pronouns (“she” and “her”) or the male gender pronouns (“he” and “him”). Participants first read instructions “You will see photos of nine women one by one. Please let us know how old you think each person is, as well as how attractive and trustworthy she seems to be.”. Unbeknownst to them, the first face they saw was always a gender-typical face, followed by the eight androgynous faces. Whether Set A or Set B appeared first, or whether the female or male gender pronoun we used first, were randomized between participants. Participants rated each face on the same four dimensions (8 point age slider, two digit age estimation, perceived attractiveness, perceived trustworthiness) as in Study 2a, then provided demographic information.

2.1.6 Results

Mean Effect of Pairing a Face with Female versus Male Names

According to the pre-registered analysis plan, we also conducted independent samples *t*-tests comparing the perceived ages of each face set when they are referred to as a woman vs. as a man. For both face set A and face set B, when paired with female gender pronoun, the faces were perceived to be significantly older on the 8-point age slider, $t(187) = 2.08, p = .03, 95\% \text{ CI} = [0.008, 0.314]$; $t(187) = 4.36, p = .001, 95\% \text{ CI} = [0.19, 0.51]$. In terms of two-digit age estimation, only face set B was perceived to be significantly older when paired with the female gender pronouns, $t(187) = 3.845, p = .00001, 95\% \text{ CI} = [1.559, 4.846]$, while for face set A, there is no significant differences, $t(187) = 0.71, p = 0.478, 95\% \text{ CI} = [-0.96, 2.04]$. Pairing

androgynous faces with either female or male gender pronouns have no consistent effect on perceived attractiveness or trustworthiness.

Finally, according to the pre-registered analysis plan, we examined whether the effects of presumed gender depends on the stimulus set/gender order by testing for a three-way interaction between presumed gender and each of the two method factors (i.e., face set A first vs. face set B first; female first vs. male first). There were no significant main effects or interactions involving the two method factors.

As with Study 2a, we also tested the effect of pairing a face with female or male gender pronoun separately for each stimulus. The resulting probabilities of observing the pattern of results as shown in Figure 2.5 upper left panel ($p = 0.0005$) and Figure 2.5 upper right panel ($p = 0.0005$) are significantly smaller than chance, showing that pairing a face has a consistent effect on perceived age, regardless of whether apparent race or approximate age range. In contrast, pairing a face with a female or male gender pronoun has no consistent effect on perceived attractiveness ($p = 0.45$), and has a smaller but significant effect on perceived trustworthiness ($p = 0.021$) in Study 2b, which failed to replicate in Study 2a.

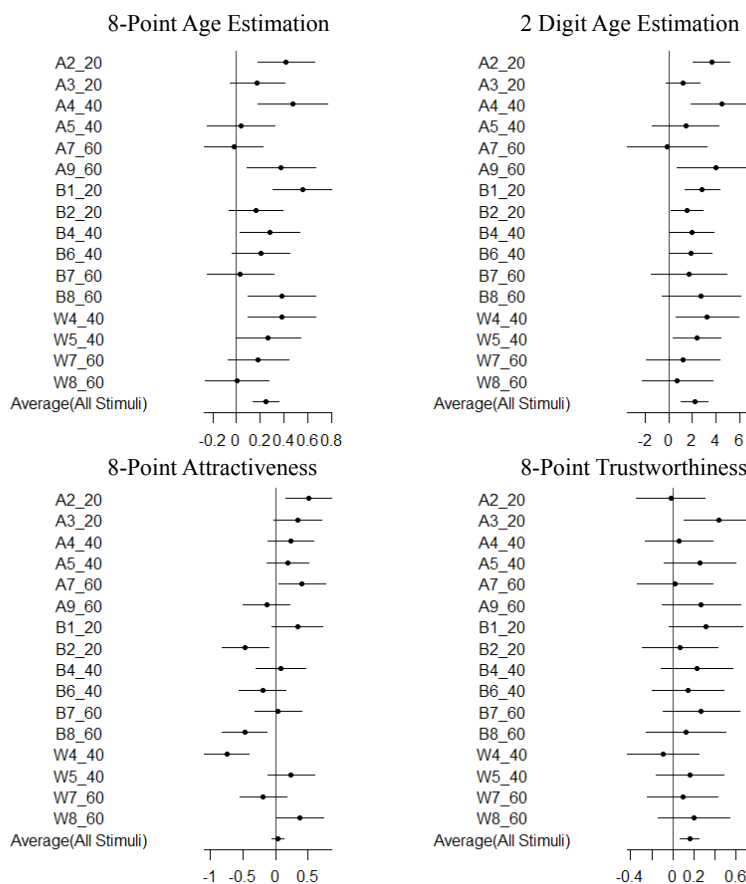


Figure 2.5. The effect of labeling a face as female versus male, for all dependent variables (dots represents the mean, and horizontal lines represent 95% CI).

For stimulus labels, “A” represents stereotypical Asian faces, “B” represents stereotypical Black faces, and “W” represents stereotypical White faces (classified according to facegen modeller). The number after the underscore represents the approximate age range in which the faces are sampled from (according to facegen modeller).

DISCUSSION

Perceiving age from visual appearance is often considered a straightforward, bottom-up process. However, the current research suggests that age perception can be modulated by manipulating the gender label of a person, either through pairing with gender-typical names (Study 1) or gender typical pronouns (Study 2). Faces paired with female names or female gender pronouns were estimated to be older on both age categorization (i.e., young versus old)

and two-digit age-estimation tasks. The same manipulation has no consistent effect on the perceived attractiveness or perceived trustworthiness of the target.

There are many mechanisms potentially responsible for this phenomenon. This effect could purely be based on statistical learning. There is evidence suggesting that, even after controlling for biological age, women on average have more neotenus features (Enlow, 1982; Jones et al., 1995). For instance, women tend to have eyes with greater vertical height (Tanikawa, Zere, & Takada, 2016), they also tend to have a greater facial width to height ratio (Robertson, Kingsley, & Ford, 2017), both are important indicators of neoteny (Jones et al., 1995). Neoteny is negatively correlated with perceived age (Berry & McArthur, 1985). Thus, to achieve accurate age estimation, when the female and male faces show the same age-related facial features, it may be rational for observers to estimate female faces to be older to compensate. The effect could also be driven by social processes. In the age of mass media, male public figures are often allowed to display a wide range of natural aging features, while female figures are often expected to put a lot of effort (e.g., makeup) to look youthful (Sontag, 1972). Accustomed to seeing carefully maintained faces of women, females who show normal signs of aging may be judged to be substantially older. More research is needed to pinpoint the mechanism underlying this phenomenon.

This finding has implications for literature on how perceivers process faces that belong to multiple social categories. According to the category dominance account (Quinn & Macrae, 2005), only the task-relevant dimension (in this case, the age for age perception tasks) were used, while task non-relevant cues (in this case, gender label) are ignored. Evidence supporting this theory has often cited lack of change in processing speed (i.e., fluency) as evidence that the non-relevant category is not activated (for instance, using the Gartner paradigm, Quinn & Macrae

(2005) found that the gender of the target does not influence the processing speed in age categorization task). However, the current research shows that, while gender information might not influence *processing speed*, it influenced the *result* of the age categorization itself. This finding highlights the importance of measuring not only the speed of processing but also the *outcome*, of social categorization when studying the dynamics of social categorization.

This finding also has implications for social outcomes. While being old is perceived negatively in both men and women (Kite & Johnson, 1988; Levy B.R. & Banaji M.R., 2002; M. North & Fiske, 2012), the negativity that comes with aging is often more severe for women (Dockterman, 2016). The current paper shows that a gender double standard exists not only in social evaluation but also in age perception itself – one in which women are evaluated to be older given the same facial appearances. This effect could potentially magnify the already negative influence of aging for women. Documenting this peculiar phenomenon is a first step towards understanding and addressing the societal gender double standard.

Chapter 3. GENDER DOUBLE STANDARD OF AGING IN HYPOTHETICAL ELECTIONS

Shen.Y.A., & Shoda.Y. (2018). Gender Double Standard of Aging on Voting: The Effect of Candidates' Age Depends on their Gender. Manuscript submitted to Psychological Science.

INTRODUCTION

“Nothing more clearly demonstrates the vulnerability of women than the special pain, confusion, and bad faith with which they experience getting older.”

-- Susan Sontag on the Double Standard of Aging

Multiple studies have demonstrated that voters are often influenced by split-second judgments based on candidates' faces (Todorov et al., 2005). However, even though age has been shown to be automatically inferred from faces (M B Brewer & Lui, 1989; M. S. North & Fiske, 2015), to our knowledge the role of candidates' age has been examined in only one published study (Sigelman & Sigelman, 1982). Even that study included only male candidates (Sigelman & Sigelman, 1982). Thus, the goal of the present study was to examine the relationship between candidates' age and their electability, and how this relation potentially differs as a function of the candidates' gender.

From Susan Sontag's classic essay on "the Double Standard of Aging" (Sontag, 1972) to recent reports of gender-based age discrimination in Hollywood (Dockterman, 2016), substantial anecdotal evidence suggests that appearing older has more negative associations for women. Moreover, recent theories of person perception suggest that multiple social categories, such as gender and race, can dynamically interact to influence person perception (Freeman & Ambady,

2011; Kang & Bodenhausen, 2015). For those reasons, we hypothesized that while voters' perception of candidates' age would influence voting intentions, this influence would be more negative for female candidates.

The present research also examined how the role of candidates' age may be accounted for by some of the basic characteristics of person perception, such as warmth and competence (e.g., Fiske, Cuddy, Glick & Xu, 2002; Olivola & Todorov, 2010; Todorov et al., 2005). In addition, we also examined how perceived attractiveness, which has been shown to decrease with age faster for females as compared to males (Chiao, Bowman, & Gill, 2008; McLellan & McKelvie, 1993), influence the relationship between perceived age and voting intention.

STUDY 1: GENDER AND AGE INTERACT TO PREDICT VOTING INTENTION

Study 1 examined how voting intention is related to candidates' apparent age and gender, using a hypothetical election paradigm.

Method

Stimulus Selection and Preparation.

Color images of 365 state legislators were obtained from 14 different states' government websites (CA, NV, ID, WY, NM, TX, ND, IA, OH, MI, FL, VA, PA, and NY). Those states were selected so that roughly half of them voted for Barack Obama and half for Mitt Romney in the 2012 United States presidential election. As described in the APPENDIX, stimuli were sampled to cover a wide age range for both genders, and stimuli were selected *without* regard to their perceived attractiveness, competence, or electability. All images were cropped so that only the head, neck, and shoulders of the person were visible. Backgrounds, often of patriotic themes

(e.g., national flag, state capitol), were digitally removed and replaced with a single-color background. Example stimuli are shown in Figure 3.1.



Figure 3.1. Example stimuli: female and male political candidates from a wide age range.

15 Amazon Mechanical Turk (MTurk) workers rated the perceived age of an initial set of 365 candidates. Based on the average responses, state legislators were grouped into five age bins (30-39 years, 40-49 years, 50-59 years, 60-69 years, 70 years and above). Approximately ten female and ten male legislators were randomly selected from the first four age bins. Due to the small number of legislators in the “70 years old and above” age bin, all in that bin were selected. Approximately half of the selected legislators self-identified as Republican and the other half as Democrat in their profiles. This resulted in a total of 93 legislators in the final stimulus set (51 males, 42 females).

Participants

Participants were 200 MTurk workers, 4 participants were excluded due to failure to complete the experiment, leaving a final sample size of 196 (90 women, 106 men, $M_{age} = 37.66$

years, $SD_{age} = 11.91$ years, 87 self-reported as being White, 29 South Asian, 11 Black or African American, 2 American Indian/Native, 39 East Asian, 5 Native Hawaiian, 10 mixed race, and 13 others).

The sample size was based on previous experiments that used hypothetical election paradigms (e.g., Sussman, Petkova & Todorov, 2013). Data were collected in early 2015. All participants had IP addresses located within the United States. No analyses were conducted before data collection was completed.

Procedure

Study 1 employed a Highly-Repeated Within-Person design (Whitsett & Shoda, 2014), in which each participant responds to a large number of stimuli, and the effect of age on voting intention is assessed separately for each participant. Participants were invited to an experiment on “politics and impression formation.” Using the Millisecond Inquisit Web platform (Draine, 2015), the 93 candidate photos were presented one at a time in random order, and participants were prompted to estimate each of their age in years by entering a two-digit number in a text box below each image. After estimating the age of all candidates, participants were shown each of the 93 candidates again and were asked to indicate their voting intention towards each candidate on a 100-point slider (“In the following task, you will view pictures of the same 93 candidates. Imagine they are all candidates for the next United States of America Senate or House of Representatives. Please indicate how likely you are to vote for each person.”). After rating all the candidates, participants indicated their age, gender, and race/ethnicity.

3.1.1 *Results*

To allow observation of the relationship without imposing assumptions about the shape (e.g., linear, quadratic, etc.) of the relationship, we first used Locally Weighted Scatterplot Smoothing (LOWESS) with 95% bootstrapped confidence intervals to plot participants' voting intentions for each candidate as a function of the perceived age of that candidate (R Core Team, 2017). Because each participant voted for each of the 93 candidates, and because we are interested in the within-participant variation across these candidates, we first mean-centered voting intention within each participant. We then plotted mean-centered voting intention as a function of the perceived age of the candidates, averaged across participants (solid black lines in Figure 3.2). In addition, Figure 3.2 shows the 95% confidence interval at each age point (shaded areas in Figure 3.2, based on bootstrapping with 10000 re-samples). As shown in Figure 3.2, for male candidates the relationship between voting intention and candidates' age appeared to follow an "inverted U" pattern, such that voting intentions increased with age up to the perceived age of approximately 45 years but then decreased beyond it. In contrast, for female candidates, voting intentions decreased almost monotonically with age (results obtained with grand mean centering were consistent with mean centering within the participant, see APPENDIX).

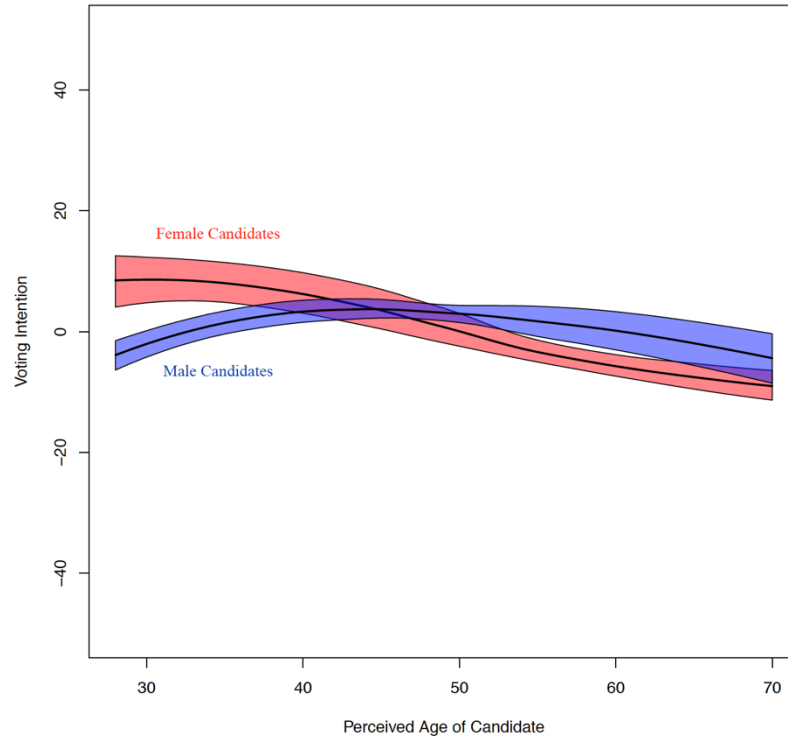


Figure 3.2. The relationship between perceived age and voting intention.

(Lowess smoothing with 95% CI). Y-axis represents voting intention on a 100 point scale mean centered within each participant; X-axis represents individual participants' age estimations. Shaded areas represent 95% bootstrapped confidence intervals.

A 2-level, stimuli-nested-within-subjects random effects model HLM 7 (Bryk, Raudenbush, & Congdon, 2010), was used to formally model the relationship between voting intention and aging. Specifically, the voting intention of participant j for candidate i was modeled as a function of candidate i 's perceived age (Age_{ij}), its quadratic term (Age_{ij}^2), candidate's i 's gender ($Gender_{ij}$), and the interaction of gender and age ($Age_{ij} * Gender_{ij}$, $Age_{ij}^2 * Gender_{ij}$). The model is formally expressed as below, variables with names starting with g are level 1 estimates of global slopes and intercept, variables starting with u are level 2 estimates of individual-to-individual variations, variables starting with r are model residuals:

$$\begin{aligned}
\text{Voting Intention}_{ij} = & g_{00} + (g_{10} + u_{1j}) * \text{Age}_{ij} + (g_{20} + u_{2j}) * \text{Gender}_{ij} + (g_{30} + u_{3j}) * \text{Gender}_{ij} * \text{Age}_{ij} + \\
& + (g_{40} + u_{4j}) * \text{Age}_{ij} * \text{Age}_{ij} + (g_{50} + u_{5j}) * \text{Gender}_{ij} * \text{Age}_{ij} * \text{Age}_{ij} + r_{ij}
\end{aligned} \tag{1}$$

Table 1. Multilevel Modeling Predicting Voting Intention as a Function of Candidate Age and Gender

	Average participant's level-1 slope coefficient	t(df)	p ₁	SD of slopes across participants	Chi-Square	p ₂
Intercept	53.32	67.76(195)	<.001	10.60	1650.90	<.001
Age	-0.34	-9.96(195)	<.001	0.45	828.09	<.001
Gender	1.233	0.971(195)	.333	16.76	838.36	<.001
Age x Gender	0.35	9.62(195)	<.001	0.40	298.40	<.001
Age ²	-0.007	-4.926(195)	<.001	0.01	300.51	<.001
Age ² x Gender	-0.016	-7.844(195)	<.001	0.01	180.00	<.001

Notes. p₁ : Probability of obtaining, in a sample such as the one in the present study, an average participant's slope that is as large or larger than the value obtained, given the null hypothesis that in the population it is 0. p₂: Probability of obtaining an SD that is as large or larger than the value obtained, given the null hypothesis that all participants have the same slope. Age was centered. Gender was coded such that female = -0.5, male = 0.5.

Consistent with the trends visible in Figure 3.2, the main effects of age and its quadratic term were qualified by an interaction with candidate gender.

The SDs for the regression coefficients were significantly greater than what would be expected by chance, indicating substantial individual-to-individual differences in the effects of age. These differences were partially due to participants' gender. Voting intention towards female candidates decreased with age faster when rated by male participants compared to female participants. Details on participant gender effects are presented in the appendix. It also should be noted that even after participant gender was taken into account, there were still statistically

significant individual differences in slopes, suggesting that such individual differences in the effects of perceived candidate age are not all due to participants' gender.

STUDY 2: GENDER AND AGE INTERACT PREDICTS VOTING WHEN GENDER AND AGE ARE NOT SALIENT

In Study 2, to rule out the possibility that the act of estimating candidates' age influenced the results, participants indicated voting intention without estimating age.

3.1.2 *Method*

Participants

Participants were 100 MTurk workers, 2 participants were excluded due to failure to complete the experiment, leaving a final sample size of 98 (49 women, 49 men, $M_{age} = 36.80$ years, $SD_{age} = 11.31$ years, 82 self-reported as being White, 1 South Asian, 5 Black or African American, 1 American Indian/Native, 6 East Asian, 2 mixed race, and 1 others).

Data were collected in late 2016. All participants had IP addresses located within the United States. No analyses were conducted before data collection was completed.

Procedure

Study 2 used the same set of stimuli and followed an almost identical procedure as Study1, except that the participants did not estimate the age of the candidates to rule out the possibility that estimating candidates' age influenced their voting intention. Instead, participants' voting intentions in Study 2 were predicted from the perceived age of each candidate as estimated by participants, and averaged across them, in Study1.

3.1.3 Results

Figure 3.3 displays the relationship between the average perceived age of each candidate as estimated by participants from Study1 (X-axis) and voting intention mean-centered within each participant as estimated by participants in Study 2 (Y-axis). The results suggest that gender and age interaction effect from Study1 was still clearly present when the participants indicated voting intentions without estimating age.

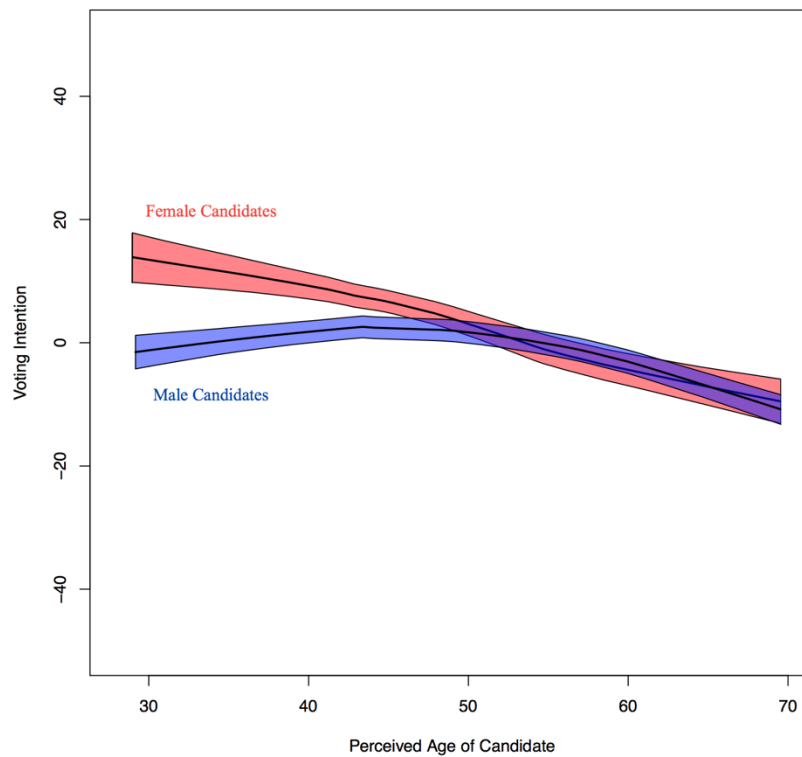


Figure 3.3. Predicting voting intention in Study 2 using age estimation made by participants of Study 1 (Lowess smoothing with 95% CI).

Y-axis represents voting intention measured on a 100 point slider, mean-centered within each participant; X-axis represents the average age for each candidate estimated by participants in Study1. Shaded areas represent 95% bootstrapped confidence interval.

We then used a 2-level, stimuli-nested-within-subjects random effects model (see Study1) to formally test the interaction between candidate age and candidate gender in predicting voting intentions. Results are presented in Table 2. Similar to Study1, the effects of age had a significant quadratic component; both the linear and quadratic components were qualified by an interaction with candidates' gender.

Analysis of the effect of participant gender also replicated findings from Study 1. Voting intention towards female candidates decreased with age faster when rated by male participants compared to female participants (see APPENDIX).

Table 2 Multilevel Modeling Predicting Voting Intention as a Function of Candidate Gender and Age as Estimated by A Separate Group of Participants

	Average participant's level-1 slope coefficient	t(df)	p ₁	SD of slopes across participants	Chi-Square	p ₂
Intercept	54.51	48.56 (97)	<.001	10.68	1282.48	<.001
Age	-.38	-7.58 (97)	<.001	0.47	934.01	<.001
Gender	-.08	-.04 (97)	.961	15.18	695.62	<.001
Age x Gender	.35	6.05 (97)	<.001	0.48	311.43	<.001
Age ²	-.01	-4.80 (97)	<.001	.02	341.75	<.001
Age ² x Gender	-.01	-5.47 (97)	<.001	.01	134.24	.007

Notes. p₁: Statistical significance given the null hypothesis that the average level-1 slope is 0. p₂: Statistical significance given the null hypothesis that the level-1 slopes are the same across participants. Age was centered. Gender was coded such that female = -0.5, male = 0.5.

STUDY 3: GENDER AND AGE INTERACT TO PREDICT VOTING INTENTION IN REALISTIC VOTING SCENARIOS

Study 3 examined the role of candidate gender and age on voting intentions using an expanded stimulus set and a paradigm that closely mimics actual voting. The experimental design and analysis plan were both pre-registered; the pre-registration could be viewed here: https://aspredicted.org/see_one.php?a_id=10332

3.1.4 *Method*

Stimulus Selection and Preparation

Stimuli were sampled from the original set of 365 state legislators that were rated by 15 pilot participants on perceived age (see Study 1). First, we identified younger candidates by selecting those whose mean estimated age was younger than 40; then we selected candidates who were perceived as being clearly older than them by selecting those that were seen as older than the oldest of the younger candidates by at least 14 of the 15 participants. Due to a relatively small number of older females and younger males in the resulting stimulus set, we included all exemplars of these two groups. We then identified younger female candidates and older male candidates who are age-matched to these two groups respectively to minimize any gender difference in perceived age within each age group. The resulting stimulus set (32 younger females and males, 24 older females and males) is reasonably representative of the youngest and the oldest of the state legislators in 14 different states in 2015 (for details on how the 365 stimulus set was sampled see Study 1 Stimulus Selection and Preparation 0).

Participants

According to pre-registration plan, we expect to collect data from 300 MTURK workers. We ended up collecting complete data from 303 participants, among which five were excluded since there is no variation in their responses. This leaves a final sample size of 298 participants (171 women, 126 men, $M_{age} = 37.66$ years, $SD_{age} = 12.74$ years, 241 self-reported as being White, 2 South Asian, 27 Black or African American, 1 American Indian/Native, 16 East Asian, seven mixed race). All participants had IP addresses located within the United States. No analyses were conducted before data collection was completed.

Procedure

On any given trial, participants viewed pairs of candidate profiles presented side by side on the screen on the Millisecond Inquisit Web platform (Inquisit 5 Web, 2015). One candidate was presented on the left, and another candidate on the right. Each profile includes one photo from the candidate, randomly sampled from the corresponding stimulus subset. In order to make the profiles appear as realistic as possible, each profile also includes 1) each candidate's name (randomly generated gender appropriate common name), 2) P.O. box number, 3) phone number, 4) education, 5) occupation, and 6) statement adapted from actual voter's pamphlets. Each pair of candidates had approximately the same education and occupation. Pairs of statements were pilot tested to be about equally likely to receive votes. After reading the two profiles, participants will be asked to vote for one of the two candidates by checking the box below that candidate's profile.

The experiment uses a within-subject design. Each participant saw 16 pairs of candidates presented in pseudorandom sequence: 6 pairs of younger female vs. older female candidates, six

pairs of younger male vs. older male candidates, four same age–different gender pairs featuring middle-aged female vs. middle-aged male candidates. Same age–different gender trials were included to make sure that not all trials are same-gender trials, those trials were excluded from the analysis.

3.1.5 Results

Since the outcome is binary, we conducted multilevel logistic regression to examine whether participants are more likely to vote for the older candidate, and more importantly, whether that is moderated by candidate gender.

The dependent variable is the log odds of the likelihood that the participant voted for *the older candidate* (1 = voted for the older candidate, 0 = voted for the younger candidate). The DV will be regressed on the gender of candidate pairs, with a random intercept and a random slope for each participant.

The model could be written in multilevel modeling annotation as below:

$$\text{Ln}(P_{ij} / (1-P_{ij})) = (g_{00} + u_{0j}) + (g_{10} + u_{1j}) * \text{PairGender}_{ij} \quad (3.1)$$

with *i* standing for candidates, *j* for participants, (PairGender: 1 = male, 0 = female)

The parameters were estimated using the generalized linear mixed model (“glmer”) module in R (Bates, 2005; Kuznetsova, Brockhoff, & Christensen, 2014). Participants were on average more likely to vote for the younger candidate over the older candidate ($b_{0j} = 0.46$, $SE = 0.07$, $p < .001$). The log odds of voting for the older candidate decreased significantly when the candidate is *female relative to male* ($b_{1j} = -0.23$, $SE = 0.08$, $p = 0.004$).

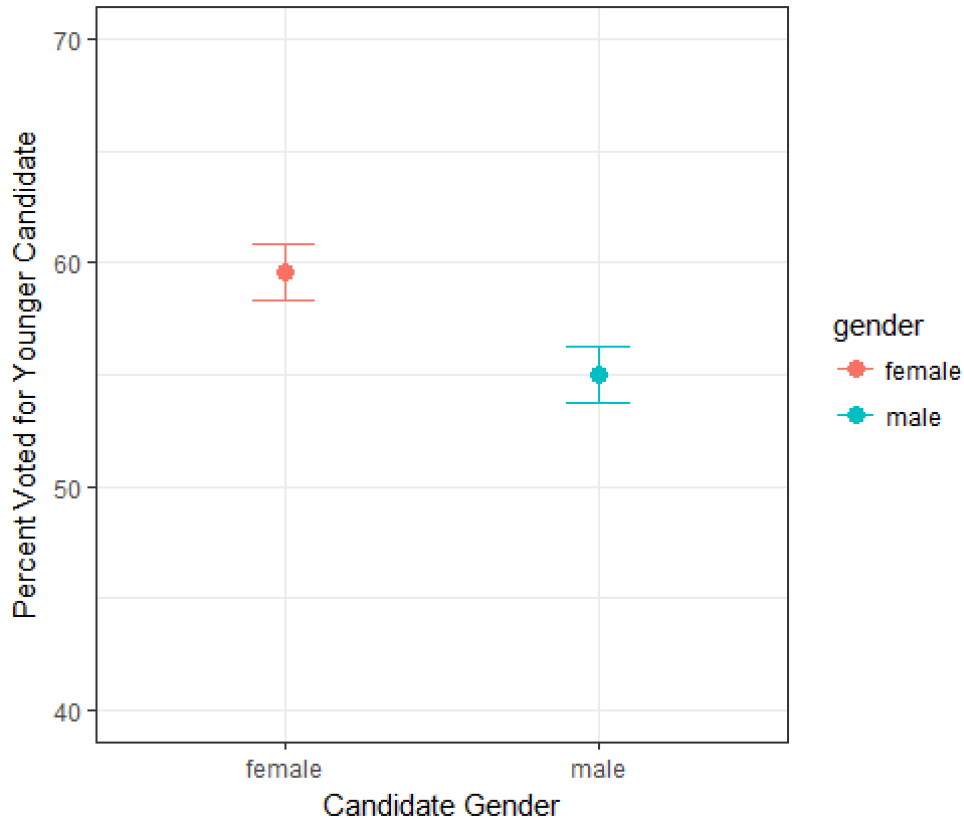


Figure 3.4. Percentage of participants who voted for the younger candidate over the older candidate by the candidate pairs' gender.

100 on Y axis means participants voted for the younger candidate out of the pair 100% of the time. 0 on Y axis means participants voted for the younger candidate out of the pair 0% of the time. Bars represent 1 within-person 95% CI.

There are significant variations in the effect of candidate pair gender between participants ($\chi^2 = 11.99, p = .002$), which means the effect of candidate pair gender on voting preference is not uniform across participants. Through exploratory analysis, we found that this effect was modulated by *participant* gender, in that male *participants* were more likely to exhibit a young preference when choosing between younger and older female candidates. Participant gender and age effects are presented in detail in the appendix.

STUDY 4: THE ROLE OF PERCEIVED ATTRACTIVENESS, COMPETENCE, AND WARMTH

3.1.6 *Method*

Participants

One hundred and eighty undergraduate students at the University of Washington participated in the study in exchange for course credit. Five participants were excluded due to failure to complete the study. Among the remaining 175 participants (49% male, 51% female), $M_{age} = 19.38$ years, $SD_{age} = 2.86$ years; 76 participants self-identified as White, 33 as East Asian, 28 as South Asian, 11 as Black or African American, 10 as mixed race, 5 as Native Hawaiian, 2 as American Indian/Alaska Native, and 10 as other/unspecified. Data were collected in early 2016. No analyses were conducted before data collection was completed.

Procedure

Students came into a lab in groups of five to participate in a study called “Impression Formation and Politics.” They were first asked to estimate the age of the same 93 candidates used in Study 1. After rating all 93 candidates’ age, they were asked to rate the same 93 candidates on perceived competence, perceived warmth, and perceived attractiveness. The sequence in which candidates were presented within each block was randomized. They rated all 93 candidates on one dimension before moving on to rate the same set of candidates on another dimension. Finally, participants indicated their voting intention towards the same 93 candidates on a 100-point slider.

3.1.7 Results

Consistent with the findings from Studies 1-3, for female candidates, voting intentions generally decreased with age, while voting intentions increase then decreased with age for male candidates (see Figure 3.5). For details on the multilevel modeling see APPENDIX.

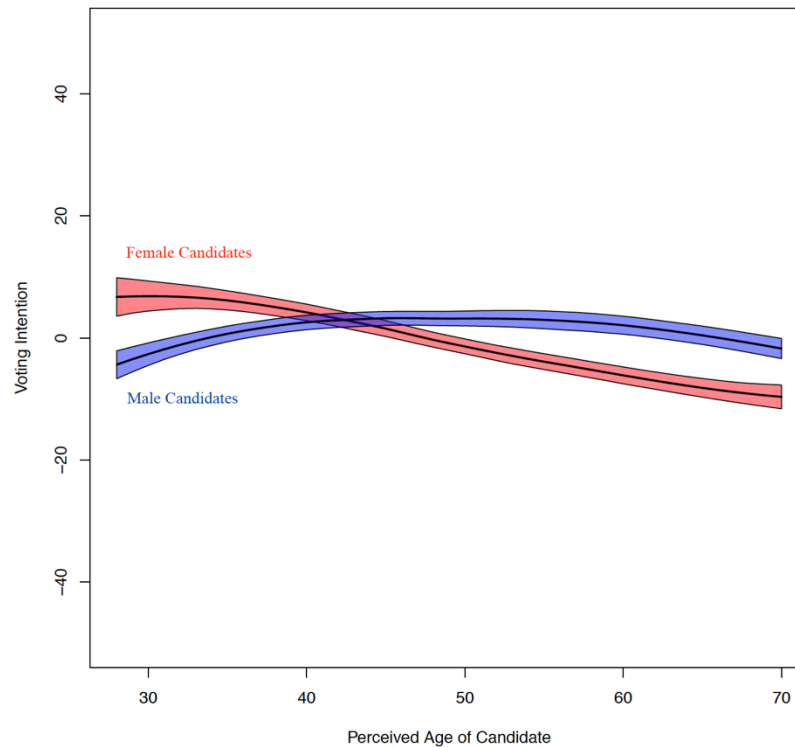


Figure 3.5. The relationship between voting intention and perceived age with 95% bootstrapped CI (Study 4).

Y-axis represents voting intention on a 100-point slider mean-centered within each participant; X-axis represents the ages of the candidates as estimated by participants.

Next, we examined the role of perceived warmth, competence, and attractiveness. In light of the difficulty in interpreting the effects of controlling for these variables from the quadratic effects, because the relationship between age and voting intention was roughly linear when candidates younger or older than about 45 years of age were considered separately, we conducted multiple simultaneous regression analysis separately for those candidates perceived to

be younger or older than 45 years of age. The analyses were conducted within-subject, for each participant.

For candidates younger than 45, controlling for perceived warmth, competence, and attractiveness greatly reduced the variance in voting intention accounted for by perceived age. In what specific ways did these variables account for the relationship between perceived age and voting intention? Are there differences between female and male candidates in the role of these variables in the relationship between perceived age and voting intention? To shed light on these questions, we used the multiple simultaneous regression analysis as described in Preacher & Hayes (2008), except that the analysis was conducted within-subject. Confidence intervals for the indirect effects through perceived attractiveness, competence, and warmth were computed separately for younger or older, female or male candidates (bootstrapped with 5000 samples), and the average indirect effects are reported in Table 3.

The results of these analyses are presented in Figure 3.6, and the section labeled "Difference: Younger Female – Younger Male" of Table 5 of supplementary materials. As shown in Panel A of Figure 3.6, although perceived attractiveness had similar effects on voting intention for male and female candidates (5.43 vs. 6.02, for female vs. male candidates, respectively), age had a much greater negative effect on perceived attractiveness for female compared to male candidates (-1.03 vs. -0.28, for female vs. male candidates), resulting in perceived attractiveness accounting for more of the relationship between perceived age and voting intention for female candidates than male candidates. Similar gender differences in the strength of mediation were observed for perceived warmth and perceived competence.

For candidates perceived to be older than 45, the overall relationship accounted for by perceived attractiveness and competence were stronger for female candidates as compared to

male candidates (see the section labeled "Difference: Older Female – Older Male" in Table 5 in supplementary materials). However, the strength of the mediation effects, as well as the gender difference, were much smaller as compared to those for candidates younger than 45 years of age. These results were consistent with analysis using the “instantaneous indirect effect” method proposed by Hayes & Preacher (2010), shown in supplementary materials.

Table 3 Indirect Effects through Perceived Attractiveness, Perceived Competence, and Perceived Warmth

Mediators	Indirect Effect	Confidence Interval	
		Lower 95% CI	Upper 95% CI
Younger (< 45) Female Candidates			
Attractiveness	-5.60****	-6.74	-4.91
Competence	-0.25 ^{ns}	-1.10	0.57
Warmth	-2.71****	-3.39	-2.22
Younger (< 45) Male Candidates			
Attractiveness	-1.72****	-2.28	-1.35
Competence	1.69****	1.06	2.32
Warmth	0.54****	0.33	0.92
Gender Differences in Indirect Effects: Younger Female – Younger Male			
Attractiveness	-3.87****	-4.93	-3.12
Competence	-1.95****	-2.76	-1.11
Warmth	-3.26****	-4.04	-2.80
Older (>=45) Female Candidates			
Attractiveness	-2.64****	-3.06	-2.24
Competence	-1.40****	-1.68	-1.01
Warmth	-0.43****	-0.63	-0.26
Older (>= 45) Male Candidates			
Attractiveness	-2.01****	-2.38	-1.66
Competence	-0.41**	-0.73	-0.12
Warmth	-0.58****	-0.79	-0.43
Gender Differences in Indirect Effects: Older Female – Older Male			
Attractiveness	-0.63**	-1.09	-0.17
Competence	-0.99****	-1.28	-0.56
Warmth	0.14 ^{ns}	-0.06	0.39

Notes. Mediation analyses were conducted within-subject; the coefficients reported in this table are the means averaged across all participants. We bootstrapped on the participant rather than trial level given the multilevel nature of the data. All predictors were standardized. ns $p > 0.05$, * $p < .05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

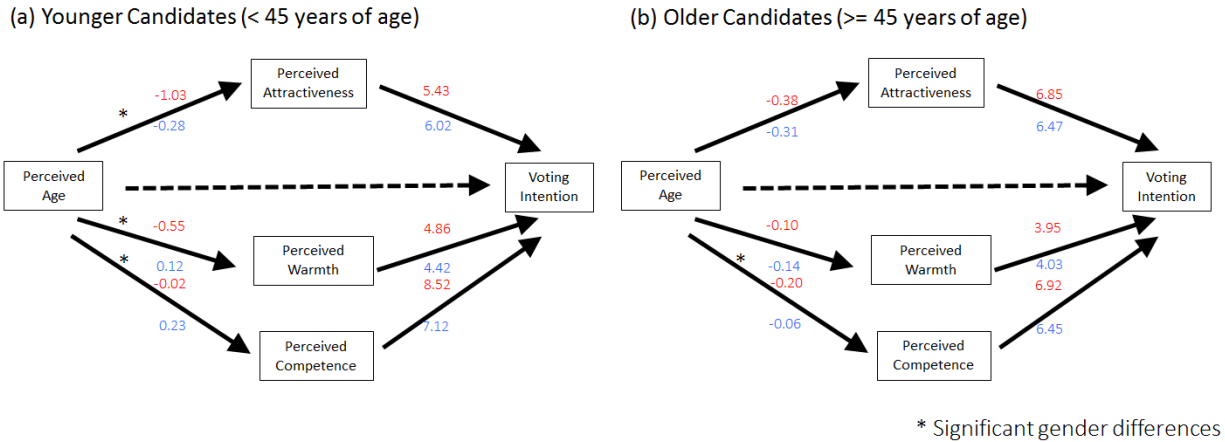


Figure 3.6. Summary of within-subject mediation analyses of the effects of perceived age on voting intention through perceived warmth, perceived attractiveness, and perceived competence (Study 4).

The left panel is for candidates perceived to be younger than 45 years of age. The right panel is for candidates perceived to be older than 45 years of age. Above each path are coefficients for female candidates (averaged across all participants), below each path are coefficients for male candidates. Asterisks indicate paths for which the coefficient for female candidates differs significantly ($p < .05$) from the coefficient for male candidates. The p values were obtained based on 5000 bootstrapped samples. Note that the mean indirect effects reported in Table 3 were not identical to the product of the relevant coefficients in Figure 3.6 because the former was computed without rounding.

GENERAL DISCUSSION

Across participant populations and paradigms, the relations between perceived age and voting intention depended on the candidates' gender. For female candidates, voting intention decreased monotonically with age, while for male candidates, voting intention increased until middle age and then decreased. This effect remained robust even when photos of the candidates were randomly paired with realistic voting materials, and when voting intention is measured using forced choice paradigm. Among candidates seen as younger than 45, perceived attractiveness and warmth, and to some extent, perceived competence, decreased with age faster for female as compared to male candidates. This partially accounts for the gender difference in the effect of perceived age on voting intentions.

We are reasonably confident that the greater age-related decrease in voting intention for female candidates is not limited to the photos of the 93 candidates. This is in part because we also found a significantly stronger negative effect of age for female candidates in a pilot study using photos of all the initial set of 365 candidates (reported in the APPENDIX). However, our stimuli come from individuals who are already elected to office and thus the results may not generalize to non-incumbent candidates. We are not aware of any data or theory indicating that the present results depend on other characteristics of the participants, materials, or context.

The present study provides the first empirical evidence supporting the notion that the “Double Standard of Aging” (Sontag, 1972) may apply to people's voting behavior. Our finding that attractiveness and warmth played a much stronger mediating role for young female candidates suggests that voters may see female and male candidates' ages through different lenses. Documenting and understanding how people perceive male and female candidates differently is a first step towards addressing gender inequality in the electoral democracy.

Chapter 4. CORRELATES OF GENDER AND AGE IN THE REAL WORLD: HOW GENDER AND SENIORITY RELATE TO REPRESENTATION IN HIGH IMPACT JOURNALS

Shen, Y. A., Shoda, Y., & Fine, I. (2018). Leading journals lack female authors. *Nature*. <https://doi.org/10.1038/d41586-018-02833-1>

Shen, Y. A., Webster, J. M., Shoda, Y., & Fine, I. (2018). Persistent Underrepresentation of Women's Science in High Profile Journals. *BioRxiv*, 275362. <https://doi.org/10.1101/275362>

The ultimate question that we would like to address in Chapter 4 is how gender moderates the relationship between quality of research and the likelihood that the research is published in high impact journals, and more importantly, whether this relationship is different for younger and older authors. As a first step, this Chapter uses gender inferred from author first name as a proxy for author gender, and authorship position (whether first or last author) as a proxy for author age. The goal is to examine how the gender of the author is related to their representation in high impact journals, and whether such representation is different for first (typically junior) and last (typically senior) authors.

INTRODUCTION

It has long been known that female representation within STEM fields decreases at every stage of the academic career (National Research Council, 2006; Valian, 1998). Take neuroscience as an example, in the year 2016, over 55% of graduate students were female, however, only 45% of postdoctoral researchers and 32% of faculty were females (McKinley Advisors, 2017).

The reason behind such gender disparity is complex (McKinley Advisors, 2017; National Research Council, 2006; Valian, 1998), one potential cause that has gathered increasing research interest is gender discrepancies within scientific publications. The impact factor of the journal one publishes in is frequently used as a benchmark for grant reviews and tenure evaluations (Rawat, 2014; Wilms, 2013; Zupanc, 2014). Due to the influence journal impact factor continues to have on scientific careers, we examined how it relates to the representation of female authors in those journals.

A series of prominent articles and editorials over the last decade have pointed out that women are underrepresented as authors of commissioned articles in *Nature* (Conley, 2005; “Nature’s sexism,” 2012; H. Shen, 2013). While commissioned opinion pieces and patents are important, one of the most prominent functions of journals such as *Nature* is to disseminate original findings. An initial small-scale analysis to examine gender disparities in research articles across two 3-month periods in 2006 and 2016 in *Nature Neuroscience* found only a 1% increase in the number of female corresponding authors over that time period (“Promoting diversity in neuroscience,” 2018). We sought to broaden this research to examine leading journals which publish neuroscience research, as well as to deepen the research by examining all articles published between 2005 and 2017.

More specifically, we evaluated the proportion of female first and last authors. First authors typically conceived of the project and performed the majority of the experimental work, while the last author is typically the primary investigator who performs the supervisory role for the project. We compared these proportions of females to the portion of female postdoctoral researchers and the proportion of females receiving government research grants, respectively. While gender parity is a laudable goal and one which many researchers, institutions, universities,

and government programs continue to work towards, a humbler goal is to have the number of publications reflect the current composition of the field. Postdoctoral researchers have demonstrated mastery over a particular subject matter and allocate the majority of their time to performing research. Setting this as our baseline for first author positions is conservative since it is not uncommon for Ph.D. candidates, where the proportion of females is higher, to also have first author publications (McKinley Advisors, 2017). Finally, the percent of female receiving prestigious grants such as NIH RO1 is an appropriate, if somewhat conservative, the baseline for the percentage of females in last author positions. Though not rid of bias, being awarded prestigious grants is indicative of the scientific rigor, significance, sophistication, and productivity of the recipient, and thus a somewhat meritocratic reflection of the current composition of primary investigators in the field.

In the current research, we used data mining techniques to examine the proportion of female first and last authors for all research articles published between 2005 and 2017 across a wide range of high-profile journals that publish neuroscience research. In the results describe here, we focus on three findings: First, we show that the proportion of women last authors in high profile research journals is much lower than the proportion of women scientists receiving USA RO1 grants or the European equivalents. Second, we show that, even within this highly selective group of journals, there is a negative relationship between journal impact-factor and proportion of female first and last authors. Finally, we show that the lack of representation of female authors has remained dispiritingly unchanged in most journals over the last 13 years.

METHOD

The full details and code for data acquisition, processing, and analysis are provided in the Github Repo (https://github.com/VisCog/Women_in_high_profile_journals). Here we describe an overview of our approach.

4.1.1 *Data Acquisition*

We downloaded metadata associated with all papers published from 2005 to 2017 from the PubMed's MEDLINE database ("MEDLINE/PubMed Data," 2017). We then subset to focus on research articles in those journals by excluding articles without an abstract.

To focus on high profile journals, we selected 15 journals to include based on the 2016 impact factors from the Thomson Reuters InCite Journal Citation Report (Clarivate Analytics, 2016). Journals which focused on a particular aspect of neuroscience (e.g., EMBO, Stroke) were excluded. This resulted in a list that included both non-specialized multidisciplinary journals (Nature, Science, Proceedings of National Academy of Science (PNAS)), and top non-specialized journals in the field of neuroscience (Nature Review Neuroscience, Nature Neuroscience, Annual Review of Neuroscience, Behavioral and Brain Sciences, Neuron, Trends in Neurosciences, Brain, Cerebral Cortex, Neuropsychology Review, Current Opinion in Neurobiology, Journal of Neuroscience, NeuroImage). We then acquired the subset of the MEDLINE publication metadata based on this list of selected journals.

These steps resulted in a total of 166,979 records for those 15 top journals between the year 2005-2017 which were included for further analysis.

For comparison with our publication data, we also acquired data on the percentage of NIH RO1 grants in the U.S. and the percentage of MRC research grants in the U.K. awarded to

women within this period. This data was obtained from the NIH data book (“MEDLINE/PubMed Data,” 2017) and MRC success rate data (“Medical Research Council 2016/17 Grant and Fellowship application success rates,” 2018), respectively, in aggregated forms.

4.1.2 *Gender Determination*

Due to the large quantities of publication records, manually classifying author gender is infeasible. Instead, we estimated the author’s gender using *genderizeR*, a *genderize.io* interface for R (Wais, 2016). The *genderize.io* database currently contains 216286 distinct first names and gender self-report data from social media platforms across 79 countries and 89 languages. Based on each unique first name, it provides a gender prediction as well as a probability estimation for the prediction. We first analyzed the full set of data, then replicated our analysis with the subset of names for which gender assignment certainty was greater than 0.9.

4.1.3 *Analysis*

To estimate the overall representation of women in each journal, we first calculated the overall percentage of female first and last author for each journal across the entire time range. To estimate the association between author gender ratios and journal profile, we calculated the Spearman’s rank order correlation between the percentage of female first and last author with each journal’s 5-year impact factor (Clarivate Analytics, 2016), with or without the three multidisciplinary journals (i.e., Nature, Science, PNAS).

To see the trends of female representation over time, we also regressed the percentage of female first and last authors in each journal on time (measured in years). The resulting slopes are an indicator of the rate of change of female authorship in each journal.

RESULT

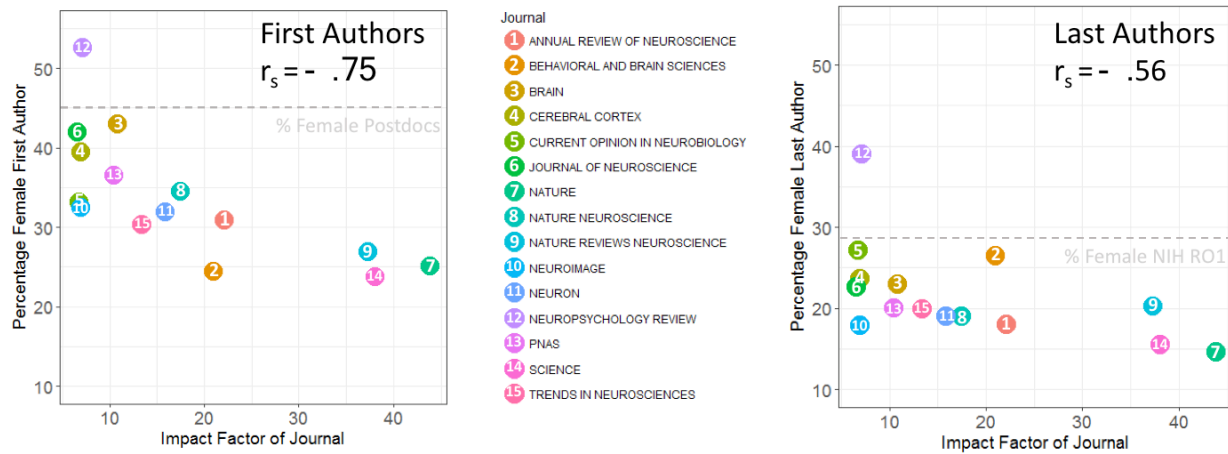


Figure 4.1. Percentage of female first and last authors between 2005-2017 vs. Journal’s 5-year impact factor

4.1.4 Percentage of Female Authors and Journal Impact Factor

Journals with higher impact factors have lower representations of female first and last authors. Through analyzing all articles with abstracts between the year 2005-2017, we found that the percentage of both female first and last authors displayed a strong negative association with journal impact factor (first author $r_s = -0.75$, $p < .01$, last author $r_s = -0.56$, $p < .05$).

It is unlikely that the low representation of women in higher-impact journals is because higher-impact journals are more likely to be multidisciplinary. We excluded all the articles that do not contain the word stems “neuro*”, “neura*”, or “brain*” in their abstracts in multidisciplinary journals (i.e. Nature, Science, and PNAS)², and found that the percentage of female first (Nature: 30.89%, Science: 30.97%, PNAS: 39.46%) and last (Nature: 14.97%, Science: 15.76%, PNAS: 21.90%) authors remain lower than baseline. Even in this subset of

² According to our analysis of the Journal of Neuroscience, 80% of the abstracts contain at least one of the following word stems: “neuro*”, “neura*”, or “brain*”.

articles, a higher impact is still associated with a lower female representation (first author $r_s = -0.48$, $p = .06$, last author $r_s = -0.61$, $p = .01$).

4.1.5 *Comparisons to Base Rate*

The percentage of female authors for high-profile neuroscience journals is lower than expected based on the proportion of women scientists in the field. As shown in Figure 4.1, between 2005-2017, the percentage of female last authors are highest in *Neuropsychology Review* (39.04%) and *Current Opinion in Neurobiology* (27.19%) and were lowest in *Nature* (14.64%) and *Science* (15.53%). This pattern of results is similar for first authors, with *Neuropsychology Review* (52.58%) and *Brain* (43.01%) having the highest percentage of females, and *Nature* (25.22%) having the lowest. Also note that the percentage of female last authors in almost all journals (except for *Neuropsychology Review*) is lower than the percentage of females awarded prestigious grants such as NIH RO1 (~30%, see grey line in Fig. 1 right panel), which is comparable to the proportion of UK Medical Research Council research awards. These grants are an important point of comparison since they are awarded based on the peer-evaluated qualities of significance, impact, research quality, and laboratory productivity.

4.1.6 *Change in Representation of Women Over Time*

As shown in Figure 4.2 and Table 4, the percentage of women first and last authors has increased by less than 1% per year for almost all journals. However, there are variations between journals. Some journals, such as *Brain*, have a steady increase of female last authors of over 1% per year, while other journals, such as *Nature Neuroscience*, have a decrease in the percentage of female last authors per year. Over that period, the percentage of women receiving NIH RO1

awards has remained roughly constant at ~30%, a percentage that reflects the number of women at the Associate/Professor level within STEM fields (see the grey line in Figure 4.2 Panel B).

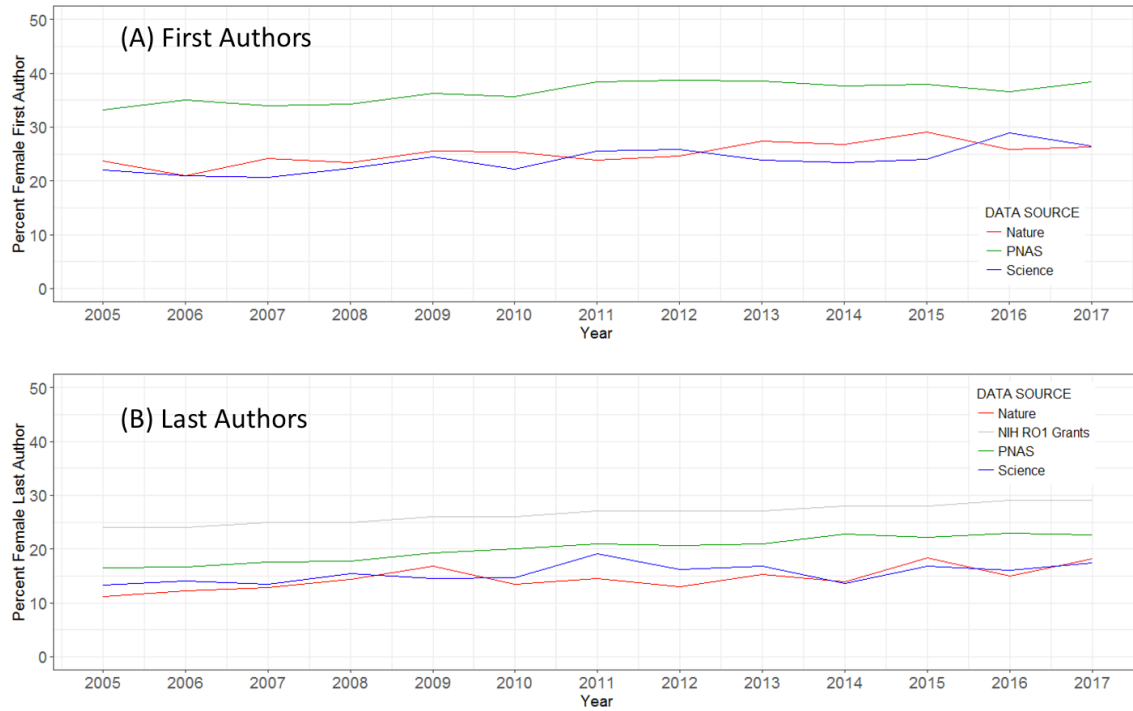


Figure 4.2. Change in the percentage of female first (Panel A) and last (Panel B) authors over time.

Table 4 Percentage Change for Female First and Last Authors Per Year

Journal	% Change in Female First Author Per Year (Slope)	% Change in Female Last Author Per Year (Slope)
Neuropsychology Review	1.85	-0.36
Nature Neuroscience	0.20	-0.11
Nature Reviews Neuroscience	0.97	0.08
Trends in Neurosciences	0.59	0.19
Neuron	0.23	0.22
Annual Review of Neuroscience	0.98	0.22
Science	0.47	0.27
Current Opinion in Neurobiology	0.22	0.31
Nature	0.40	0.40
Behavioral and Brain Sciences	1.80	0.44
Journal of Neuroscience	0.60	0.53
PNAS	0.40	0.58
NeuroImage	0.86	0.77
Cerebral Cortex	0.76	0.78
Brain	1.42	1.03

DISCUSSION

Using data mining techniques, we evaluated the publication records of original research articles for the top 15 journals publishing neuroscience research from 2005-2017. We found that 1) proportion of women authors in high profile research journals is substantially lower than the proportion of women receiving competitive grants, 2) there was a negative relationship between journal impact-factor and proportion of female first and last authors, and 3) the rate of increase in female representation is on average less than one percent per year for first authors and less than half a percent per year for last authors.

While our research demonstrated gender discrepancies, our data do not speak to the underlying causes, which can be difficult to identify (Gannon, Quirk, Guest, & Wallon, 2001). One possibility is that women are submitting less to high profile journals. Editors of *Nature Neuroscience* reported only 21.5% of submissions to their journal were from female authors

(“Promoting diversity in neuroscience,” 2018). Another possibility is that women are less successful in negotiating prestigious authorship positions. While women are more likely to be the person performing experiments (Macaluso, Lariviere, Sugimoto, & Sugimoto, 2016), they are less likely to be in the prestigious lead author positions (West, Jacquet, King, Correll, & Bergstrom, 2013). A third possibility is bias in the publication pipeline. Publishing in high impact journals is often synonymous with academic “genius.” While “genius” is often stereotyped to be a male trait (Bian, Leslie, & Cimpian, 2017), it is possible that women are considered less suitable for those high impact authorship positions. Moreover, experimental evidence suggests that, when reviewers are randomly assigned to evaluate scientific work ostensibly submitted by a female or a male author, they rated the work written by male authors as having higher rigor (Knobloch-Westerwick, Glynn, & Huge, 2013). More research is needed to evaluate the relative importance of those underlying mechanisms.

One potential way to alleviate the negative influence of impact factor on female scientists’ careers is to move away from using impact factors to inform decisions. It is currently forbidden in some countries and institutions to use journal impact-factor for decisions regarding hiring, funding, and advancement (Alberts et al., 2013; “Ending the tyranny of the impact factor,” 2014; Garfield, 2005). The reason behind those decisions is that journal impact factor is a poor proxy for the influence of scientific work -- in almost all cases, journals with high impact factor receive the vast majority of their citations for a very small number of articles (upwards of 80% of the citations come from less than 20% of the articles) (Garfield, 2005; “Time to remodel the journal impact factor,” 2016). The reliance on journal impact-factor only provides a somewhat limited perspective on the performance and influence of a specific journal article (Favaloro, 2008; Habal, 2018).

A number of alternative criteria have been proposed to provide a more meaningful metric of scientific contribution (Banks & Dellavalle, 2008; Oosthuizen & Fenton, 2014). The *h*-index reflects the amount of highly cited work of an individual scientist (Hirsch, 2005), and has been advocated by both Nature (Ball, 2005) and Science (Holden, 2005). The Research Status metric, which seeks to compensate for the correlation between publication rate and the number of citations, has been proposed as an alternative to the *h*-index, which performs more evenly across age and gender (Symonds, Gemmell, Braisher, Gorringer, & Elgar, 2006). However, none of these metrics take into account the quantity or quality of activities such as serving on grant panels, organizing conferences, mentoring, and providing peer review. Ultimately metrics must be devised to provide useful insights into the progress of research and provide incentives which align with the goals of science (Kun, 2015).

Even though far from ideal, the current reality is that publishing in high impact-factor journals remains an important gateway for career advancement. High impact-factor publications have an enormous influence on the likelihood of receiving awards, funding, and positions in highly ranked research institutions. Conversely, the lack of high-profile publications may partially account for the lower rate of recruitment, retention, and promotion for women faculty. The current under-representation of women in high profile journals impacts thousands of talented scientists.

It is now well past time for high-impact journals to begin collaborating with the scientific community to develop and validate evidence-based procedures to remove sources of bias throughout both the editorial and the reviewing process for original scientific articles. We would recommend some obvious first steps. First, all journals should collect gender and minority statistics on submission and acceptance rates for papers and should make these data publically

available. Second, journals should use mandatory double-blind reviewing. Results from other disciplines suggest that double-blind reviewing procedures significantly increase the proportion of female lead research articles (Budden et al., 2008). Finally, reviewers should be provided with more precise guidance about review criteria, and these criteria should implement evidence-based standards based on best practices for the advancement of scientific values (Casadevall & Fang, 2014; Moher et al., 2018).

Chapter 5. SUMMARY & FINAL REMARKS

5.1.1 *Summary*

The research described in this dissertation suggests that gender and age interact to influence perception and social outcomes.

Chapter 2 and showed that the gender of a face could influence age estimation and categorization. Study 1 showed that female faces tend to start aging earlier in the artificial “aging” process created through morphing. Study 2 and Study 3 showed that being labeled as female tended to make a face appear several years older in various age estimation tasks.

Chapter 3 and showed that while candidate age did bias voting intention, the effect was not the same for female and male candidates. Participants' voting intentions for female candidates monotonically decreased as the candidates were perceived to be older, while voting intentions towards male candidates remained the same or only decreased slightly with age (Study 1). The pattern of results holds across different paradigms (Study 2, Study 3), different participant pools (Study 4), and different stimulus pools (Study 3). Study 4 investigated the mediating role of perceived attractiveness, competence, and warmth. Among candidates seen as younger than 45, perceived attractiveness, and to some extent, perceived warmth and competence, decreased with age faster for female as compared to male candidates. This partially accounts for the gender difference in the effect of perceived age on voting intentions.

Chapter 2 and *Chapter 3* employed a Highly-Repeated Within-Person Design, which allowed examining how individual differences, such as *participant* age and gender, influence the magnitude/direction of effects. We found that *participant* gender significantly moderate the

overall effect, in that target age and gender had a stronger influence on outcomes for male participants. *Participant* age showed no consistent effects across experiments.

Chapter 4 establish the link between gender and likelihood of publishing in high impact journals. The result is striking; we found that both junior and senior authors are severely underrepresented as compared to their respective baselines, that the gender discrepancy is larger for higher impact journals, and that these patterns have barely improved over the last 13 years.

5.1.2 *Final Remarks*

This dissertation suggests that perceivers may view aging in females and males through different lenses. The direction of the findings is consistent with Susan Sontag's "double standard of aging" observations (Sontag, 1972), in that aging disproportionately disadvantages women. Such a tendency still exists in contemporary society and reminds us that the battle to combat age discrimination, particularly that towards women, is far from being accomplished.

This dissertation also points to the importance of going beyond traditional practices in psychology of studying the influence of a single social-group label and ignoring cross-cutting social-group labels. For instance, a study on the effect of age that only uses male stimuli could produce results that would not generalize to female stimuli. It is important to qualify the findings if only male stimuli were used, or include both male and female stimuli in the study. This practice will likely make results more replicable, while also provide more insight into how social group labels interact to influence perception and social outcomes.

BIBLIOGRAPHY

- Alberts, B., Ball, P., Banks, M. A., Dellavalle, R., Budden, A. E., Tregenza, T., ... Zupanc, G. K. H. (2013). Beat it, impact factor! Publishing elite turns against controversial metric. *Nature*, *16*(1), 1. <https://doi.org/10.1126/science.1240319>
- Andreoletti, C., Maurice, J. K., & Whalen, H. (2001). Gender, race, and age: Compound stereotypes across the lifespan. In *Poster presented at the Second Annual Meeting of the Society for Personality and Social Psychology*.
- B. Levy, & Langer, E. (1994). Aging free from negative stereotypes: Successful memory in China among the American deaf. *Journal of Personality and Social Psychology*, *66*(6), 989.
- Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, *436*(7053), 900. <https://doi.org/10.1038/436900a>
- Banks, M. A., & Dellavalle, R. (2008). Emerging alternatives to the impact factor. *OCLC Systems & Services: International Digital Library Perspectives*, *24*(3), 167–173.
- Bates, D. (2005). Fitting linear mixed models in R. Using the lme4 package. *R News*, *5*(May), 27–30. <https://doi.org/10.1159/000323281>
- Bendick, M. J., Brown, L. E., & Wall, K. (1999). No foot in the door: An experimental study of employment discrimination against older workers. *Journal of Aging & Social Policy*, *10*(4), 5–23.
- Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, *94*(1–2), 8–15. <https://doi.org/10.1016/j.jpubeco.2009.11.002>
- Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology*, *48*(2), 312–323. <https://doi.org/10.1037/0022-3514.48.2.312>
- Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: the impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*(1), 3.
- Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science (New York, N.Y.)*, *355*(6323), 389–391. <https://doi.org/10.1126/science.aah6524>
- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (pp. 187–194). ACM Press/Addison-Wesley Publishing Co.
- Bodenhausen, G. V. (2009). Diversity in the person, diversity in the group: Challenges of identity complexity for social perception and social interaction. *European Journal of Social Psychology*, *40*(1). <https://doi.org/10.1002/ejsp.647>
- Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, *41*(4), 656.
- Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition*, *7*(3), 262–274. <https://doi.org/10.1521/soco.1989.7.3.262>
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.

- Bryk, A. S., Raudenbush, S. W., & Congdon, R. (2010). HLM 7 for Windows (Computer software). *Chicago: Scientific Software International.*
- Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, *23*(1), 4–6. <https://doi.org/10.1016/j.tree.2007.07.008>
- Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, *41*(12), 1461–1471. <https://doi.org/10.1177/0146167214567153>.
- Casadevall, A., & Fang, F. C. (2014). Causes for the persistence of impact factor mania. *MBio*, *5*(2), e00064-14. <https://doi.org/10.1128/mBio.00064-14>
- Chiao, J. Y., Bowman, N. E., & Gill, H. (2008). The political gender gap: Gender bias in facial inferences that predict voting behavior. *PLoS One*, *3*(10), e3666. <https://doi.org/10.1371/journal.pone.0084677>
- Clarivate Analytics. (2016). 2016 Journal Citation Reports® . Retrieved from <https://incites.thomsonreuters.com/>
- Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the Early Stages of Person Perception: The Asymmetry of Social Categorization by Sex vs. Age. *PLoS ONE*, *9*(1), e84677. <https://doi.org/10.1371/journal.pone.0084677>
- Conley, D. (2005). Women’s efforts are more than a drop in the ocean. *Nature*, *438*(7071), 1078. <https://doi.org/10.1038/4381078c>
- Cuddy, A. J. C., Norton, M. I., & Fiske, S. T. (2005). This old stereotype: The pervasiveness and persistence of the elderly stereotype. *Journal of Social Issues*, *61*(2), 267–285.
- Deusch, F. M., Zalenski, C. M., & Clark, M. E. (1986). Is There a Double Standard of Aging? *Journal of Applied Social Psychology*, *16*(9), 771–785.
- Dockterman. (2016). ACLU: EEOC Is Investigating Sexism Against Female Directors | Time. *Time*. Retrieved from <http://time.com/4326519/aclu-government-sexism-female-directors-hollywood/>
- Draine, S. (2015). Inquisit 4 [computer software]. Millisecond Software. Retrieved from <https://www.millisecond.com/>
- Eberhardt, J. L., Dasgupta, N., & Banaszynski, T. L. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/0146167202250215>
- Ending the tyranny of the impact factor. (2014). *Nature Cell Biology*, *16*(1), 1. <https://doi.org/10.1038/ncb2905>
- Enlow, D. H. (1982). Handbook of facial growth WB Saunders.
- Favaloro, E. J. (2008). Measuring the quality of journals and journal articles: the impact factor tells but a portion of the story. *Seminars in Thrombosis and Hemostasis*, *34*(1), 7–25. <https://doi.org/10.1055/s-2008-1066030>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.
- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247.
- Gannon, F., Quirk, S., Guest, S., & Wallon, G. (2001). Searching for discrimination. Are women

- treated fairly in the EMBO postdoctoral fellowship scheme? *EMBO Reports*, 2(8), 655–657. <https://doi.org/10.1093/embo-reports/kve170>
- Garfield, E. (2005). The Agony and the Ecstasy—the History and the Meaning of the Journal Impact Factor. Report at the Fifth International Congress on Peer Review in Biomedical Publication. Chicago: Fifth International Congress on Peer Review in Biomedical Publication.
- Gilbert, C., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*. Retrieved from <https://www.nature.com/articles/nrn3476>
- Habal, M. B. (2018). The Outdated Thinking That Drove the Dis-impaction of the Journal Impact Factor. LWW.
- Hayes, A. F., & Preacher, K. J. (2010). Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. *Multivariate Behavioral Research*, 45(4), 627–660. <https://doi.org/10.1080/00273171.2010.498290>
- Helman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face–Time Continuum: Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin*, 40(12), 1624–1636. <https://doi.org/10.1177/0146167214552791>
- Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Holden, C. (2005). Random samples: Data point-impact factor. *Science*, 309(8), 1181c. Retrieved from <http://science.sciencemag.org/content/309/5738/r-samples>
- Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, 119–147.
- Hugenberg, K., & Bodenhausen, G. V. (2004). Category membership moderates the inhibition of social identities. *Journal of Experimental Social Psychology*, 40(2), 233–238.
- Hummert, M. L. (1990). Multiple stereotypes of elderly and young adults: a comparison of structure and evaluations. *Psychology and Aging*, 5(2), 182.
- Hummert, M. L., Garstka, T. A., & Shaner, J. L. (1997). Stereotyping of older adults: The role of target facial cues and perceiver characteristics. *Psychology and Aging*, 12(1), 107.
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: how covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, 102(1), 116.
- Jones, D., Brace, C. L., Jankowiak, W., Laland, K. N., Musselman, L. E., Langlois, J. H., ... Symons, D. (1995). Sexual Selection, Physical Attractiveness, and Facial Neoteny: Cross-cultural Evidence and Implications [and Comments and Reply]. *Current Anthropology*, 36(5), 723–748. <https://doi.org/10.1086/204427>
- Kang, S. K., & Bodenhausen, G. V. (2015). Multiple Identities in Social Perception and Interaction: Challenges and Opportunities. *Annual Review of Psychology*, 66(1), 547–574. <https://doi.org/10.1146/annurev-psych-010814-015025>
- Kang, S. K., & Chasteen, A. L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *The International Journal of Aging and Human Development*, 69(3), 201–220.
- Kemper, S., Vandepute, D., Rice, K., Cheung, H., & Gubarchuk, J. (1995). Speech adjustments to aging during a referential communication task. *Journal of Language and Social Psychology*, 14(1–2), 40–59.

- Ketcham, C. J., Seidler, R. D., Gemmert, A. W. Van, & Stelmach, G. E. (2002). Age-related kinematic differences as influenced by task difficulty, target size, and movement amplitude. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 57(1), P54-64.
- Kite, M. E., Deaux, K., & Miele, M. (1991). Stereotypes of young and old: Does age outweigh gender? *Psychology and Aging*, 6(1), 19.
- Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults: a meta-analysis. *Psychology and Aging*, 3(3), 233.
- Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35(5), 603–625.
- Kogan, N. (1979). A study of age categorization. *Journal of Gerontology*, 34(3), 358–367.
- Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, 111(25), 9079–9084. <https://doi.org/10.1073/pnas.1404448111>
- Kun, H. (2015). Evaluation: Moving away from metrics. *Nature*, 520(7549), S18-20. <https://doi.org/10.1038/520S18a>
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, 103(2), 284.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R Package Version*. <https://doi.org/http://CRAN.R-project.org/package=lmerTest>
- Levy B.R., & Banaji M.R. (2002). Implicit ageism. In *Ageism: Stereotyping and Prejudice Against Older Persons*. Retrieved from https://books.google.com/books?hl=en&lr=&id=UvxEoFQ0LYwC&oi=fnd&pg=PA49&dq=Implicit+Ageism&ots=zZUhaKOVWj&sig=ZIIg-jkSEZeHt9IAA_2rIVRMM5Q
- Macaluso, B., Lariviere, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Academic Medicine*, 91(8), 1136–1142. <https://doi.org/10.1097/ACM.0000000000001261>
- Macrae, C. N., Bodenhausen, G. V, Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, 67(5), 808.
- Mark, L. S., Pittenger, J. B., Hines, H., Carello, C., Shaw, R. E., & Todd, J. T. (1980). Wrinkling and head shape as coordinated sources of age-level information. *Perception & Psychophysics*, 27(2), 117–124.
- Matts, P. J., Fink, B., Grammer, K., & Burquest, M. (2007). Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*, 57(6), 977–984.
- McDermott, M. L. (1997). Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science*, 270–283.
- McKinley Advisors. (2017). Society for Neuroscience Survey of Neuroscience Departments & Programs. Retrieved from <https://www.sfn.org/-/media/SfN/Documents/Survey-Reports/NDP-Final-Report.ashx?la=en&hash=41FEFFA45C371F27648DF48DB0A11E46A19171E0>
- McLellan, B., & McKelvie, S. J. (1993). Effects of age and gender on perceived facial

- attractiveness. *Canadian Journal of Behavioural Science*, 25(1), 135–142.
<https://doi.org/10.1037/h0078790>
- Medical Research Council 2016/17 Grant and Fellowship application success rates. (2018). Retrieved from <https://www.mrc.ac.uk/research/funded-research/success-rates/#grant>
- MEDLINE/PubMed Data. (2017). Retrieved from https://www.nlm.nih.gov/databases/download/pubmed_medline.html
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, 36(4), 630–633. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15641408>
- Mitro, S., Gordon, A. R., Olsson, M. J., & Lundström, J. N. (2012). The smell of age: perception and discrimination of body odors of different ages. *PloS One*, 7(5), e38110.
- Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology*, 16(3), e2004089. <https://doi.org/10.1371/journal.pbio.2004089>
- National Research Council. (2006). *To recruit and advance: Women students and faculty in science and engineering*. National Academies Press.
- Nature's sexism. (2012). *Nature*, 491(7425), 495. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23189321>
- North, M., & Fiske, S. (2012). An Inconvenienced Youth? Ageism and Its Potential Intergenerational Roots. *Psychol Bull*, 138(5), 982–997. <https://doi.org/10.1037/a0027843>
- North, M. S., & Fiske, S. T. (2015). Modern attitudes toward older adults in the aging world: A cross-cultural meta-analysis. *Psychological Bulletin*, 141(5), 993–1021. <https://doi.org/10.1037/a0039469>
- O'Connell, A. N., & Rotter, N. G. (1979). The influence of stimulus age and sex on person perception. *Journal of Gerontology*, 34(2), 220–228.
- Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*. <https://doi.org/10.1007/s10919-009-0082-1>
- Oosthuizen, J. C., & Fenton, J. E. (2014). Alternatives to the impact factor. *The Surgeon*, 12(5), 239–243.
- Penner, A. M., & Saperstein, A. (2008). How social status shapes race. *Proceedings of the National Academy of Sciences of the United States of America*, 105(50), 19628–19630. <https://doi.org/10.1073/pnas.0805762105> [doi]
- Pittenger, J. B., & Shaw, R. E. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1(4), 374.
- Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of facial contrast decrease with age and are cues for age perception. *PloS One*, 8(3), e57985.
- Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology*, 45(5), 1132–1135. <https://doi.org/10.1016/j.jesp.2009.06.007>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. In *Behavior Research Methods* (Vol. 40, pp. 879–891). <https://doi.org/10.3758/BRM.40.3.879>
- Promoting diversity in neuroscience. (2018). *Nature Neuroscience*, 21(1), 1. <https://doi.org/10.1038/s41593-017-0052-6>

- Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: the dynamics of person construal. *Journal of Personality and Social Psychology*, 88(3), 467.
- Rawat, S. (2014). How is impact factor impacting our research. *Biomed J*, 37(6), 415–416.
- Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology*, 23(1), 1–12.
- Robertson, J. M., Kingsley, B. E., & Ford, G. C. (2017). Sexually Dimorphic Faciometrics in Humans From Early Adulthood to Late Middle Age: Dynamic, Declining, and Differentiated: <https://doi.org/10.1177/1474704917730640>
<https://doi.org/10.1177/1474704917730640>
- Ross, B., & Murphy, G. (1996). Category-based predictions: influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 736-753. Retrieved from <http://psycnet.apa.org/fulltext/1996-04402-011.html>
- Rubin, K. H., & Brown, I. D. R. (1975). A life-span look at person perception and its relationship to communicative interaction. *Journal of Gerontology*, 30(4), 461–468.
- Sanbonmatsu, K. (2002). Gender stereotypes and vote choice. *American Journal of Political Science*, 20–34.
- Shen, H. (2013). Inequality quantified: Mind the gender gap. *Nature*, 495(7439), 22–24.
<https://doi.org/10.1038/495022a>
- Shen, Y. A., Shoda, Y., & Fine, I. (2018). Leading journals lack female authors. *Nature*.
<https://doi.org/10.1038/d41586-018-02833-1>
- Shen, Y. A., Webster, J. M., Shoda, Y., & Fine, I. (2018). Persistent Underrepresentation of Women’s Science in High Profile Journals. *BioRxiv*, 275362.
<https://doi.org/10.1101/275362>
- Shipp, T., Qi, Y., Huntley, R., & Hollien, H. (1992). Acoustic and temporal correlates of perceived age. *Journal of Voice*, 6(3), 211–216.
- Sigelman, L., & Sigelman, C. K. (1982). Sexism, racism, and ageism in voting behavior: An experimental analysis. *Social Psychology Quarterly*, 263–269.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, 77(5), 885.
- Singer, R. (1986). Agaricales in modern taxonomy.
- Sontag, S. (1972). double standard of aging. *The Saturday Review*, September, 29–38. Retrieved from <http://www.unz.com/print/SaturdayRev-1972sep23-00029/>
- Symonds, M. R. E., Gemmell, N. J., Braisher, T. L., Gorringer, K. L., & Elgar, M. A. (2006). Gender Differences in Publication Output: Towards an Unbiased Metric of Research Performance. *PLoS ONE*, 1(1), e127. <https://doi.org/10.1371/journal.pone.0000127>
- Tanikawa, C., Zere, E., & Takada, K. (2016). Sexual dimorphism in the facial morphology of adult humans: A three-dimensional analysis. *HOMO - Journal of Comparative Human Biology*, 67(1), 23–49. <https://doi.org/10.1016/J.JCHB.2015.10.001>
- Thimm, C., Rademacher, U., & Kruse, L. (1998). Age stereotypes and patronizing messages: Features of age-adapted speech in technical instructions to the elderly.
- Time to remodel the journal impact factor. (2016). *Nature*, 535(7613), 466.
<https://doi.org/10.1038/535466a>
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science (New York, N.Y.)*, 308(5728), 1623–1626.
<https://doi.org/308/5728/1623> [pii]

- Valian, V. (1998). *Why so slow? The advancement of women. Why so slow? The advancement of women*. Cambridge, MA, US: The MIT Press.
- Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging, 27*(2), 265–277. <https://doi.org/10.1037/a0025065>
- Wais, K. (2016). genderizeR: Gender Prediction Based on First Names. R package version 2.0.0. Retrieved from <https://cran.r-project.org/package=genderizeR>
- West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PloS One, 8*(7), e66212. <https://doi.org/10.1371/journal.pone.0066212>
- Whitsett, D. D., & Shoda, Y. (2014). An approach to test for individual differences in the effects of situations without using moderator variables. *Journal of Experimental Social Psychology, 50*, 94–104.
- Wilms, G. (2013). The impact factor. *Neuroradiology, 55*(7), 803–806. <https://doi.org/10.1007/s00234-013-1227-9>
- Zupanc, G. K. H. (2014). Impact beyond the impact factor. *Journal of Comparative Physiology A, 200*(2), 113–116. <https://doi.org/10.1007/s00359-013-0863-1>

CHAPTER 2 APPENDIX

LIST OF NAMES USED IN STUDY 1

List of female names:

Carol, Angela, Laura, Brianna, Rebecca, Elizabeth, Jane, Helen, Bella, Mary, Susan, Sarah, Nancy, Linda, Karen, Joyce

List of male names:

Daniel, James, Ben, Rob, Kevin, Josh, Tony, Ryan, Matt, Dean, Allen, Thomas, Chris, Bill, Henry, Andrew

PARTICIPANT GENDER EFFECT IN STUDY 1

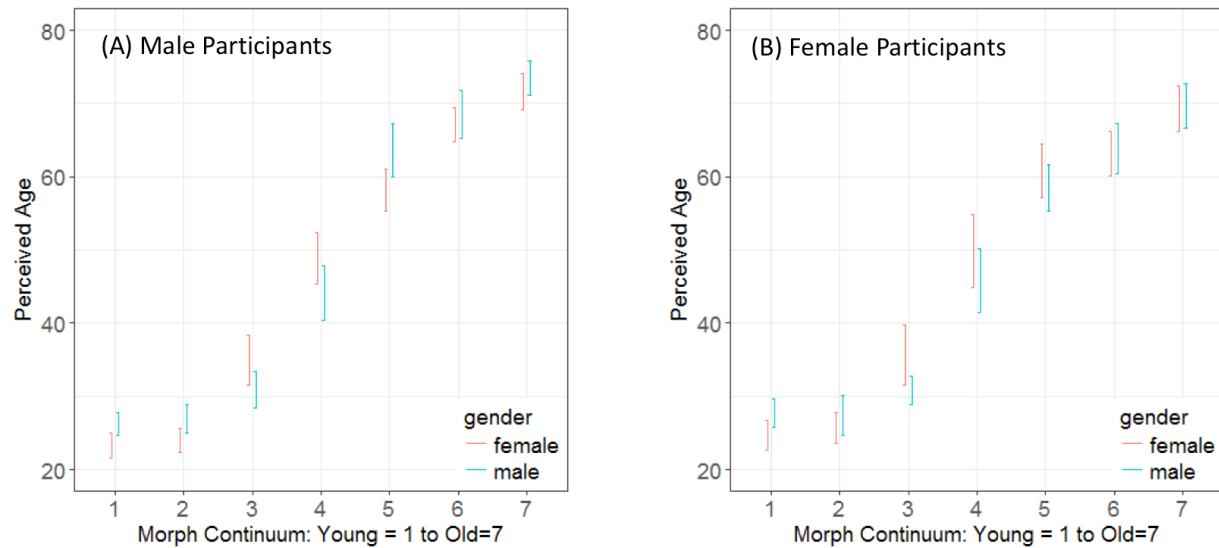


Figure 1. The relationship between morph point and perceived age for female and male participants.

Bars represent 95% CI, collapsing across all four pairs of morph continuums.

Participant gender did not influence the relationship between stimulus gender, morph point, and perceived age, in that morph point 2 versus morph point 3 and stimulus gender interact to predict perceived age equally for female and male participants ($p > .5$).

PARTICIPANT GENDER EFFECT IN STUDY 2A AND STUDY 2B

One individual difference variable that might influence the result is participant gender. Here we investigate how participant gender influences the relationship between gender label and age perception in a within-person analysis. One group of participants ($n = 99$) viewed set A paired with female names and set B paired with male names. For this group of participants, a within-person analysis would compare their mean responses to faces paired with female names to faces paired with male names. Similar methods were used to conduct within person analysis in Study 2a and Study 2b.

We then plotted the mean perceived age for faces paired with female versus male names separately for male and female *participants*.

Study 2a

Confirming results obtained using multilevel analysis, in the within-person analysis, participants rated faces paired with female names to be significantly older on both 8-point age slider or two-digit age estimation tasks (see Figure 1, $p < .05$).

There was no significant mean effect of participant age, nor did participant age interact with gender label of the face to predict perceived age ($p > .15$)

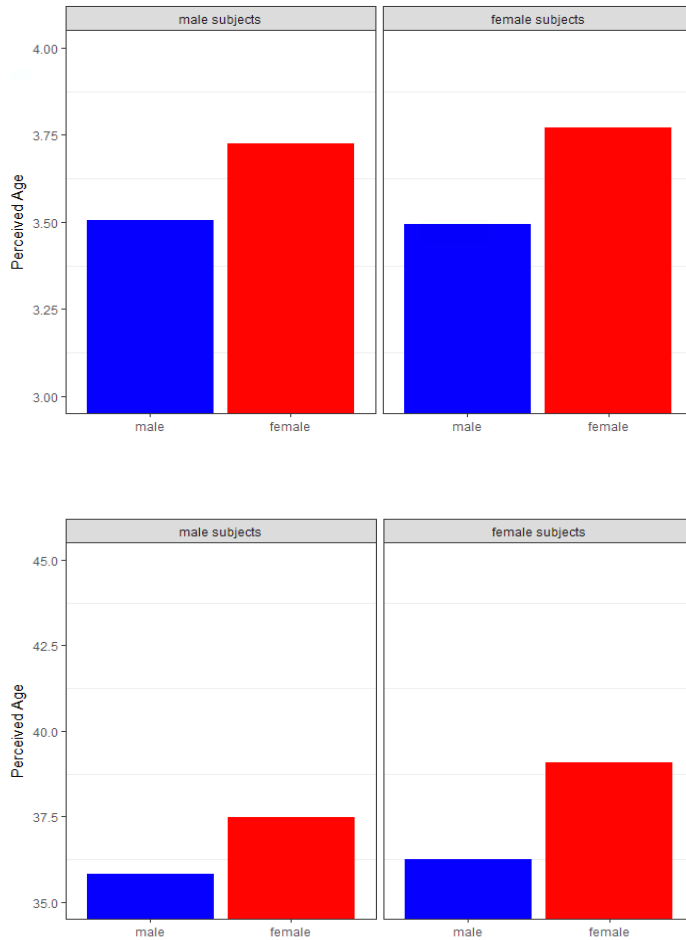


Figure 2. Female and male participants’ mean age estimation on the eight-point slider (upper panel) and two-digit age estimation task (lower panel) in Study 2a.

Study 2b

We also tested whether participant gender interact with stimulus gender to predict perceived age in Study 2b. In Study 2b, while participant gender has a mean effect on 8-point age perception ($\beta = -0.27, p < .001$) and two-digit age estimation ($\beta = -1.68, p = .03$), it did not interact with stimulus gender to predict 8-point age perception ($\beta = 0.04, p = 0.65$), or two-digit age estimation ($\beta = 0.10, p = 0.92$).

Based on those results, we believe that the effect of gender labeling on age perception is not contingent on the gender of the *participant*.

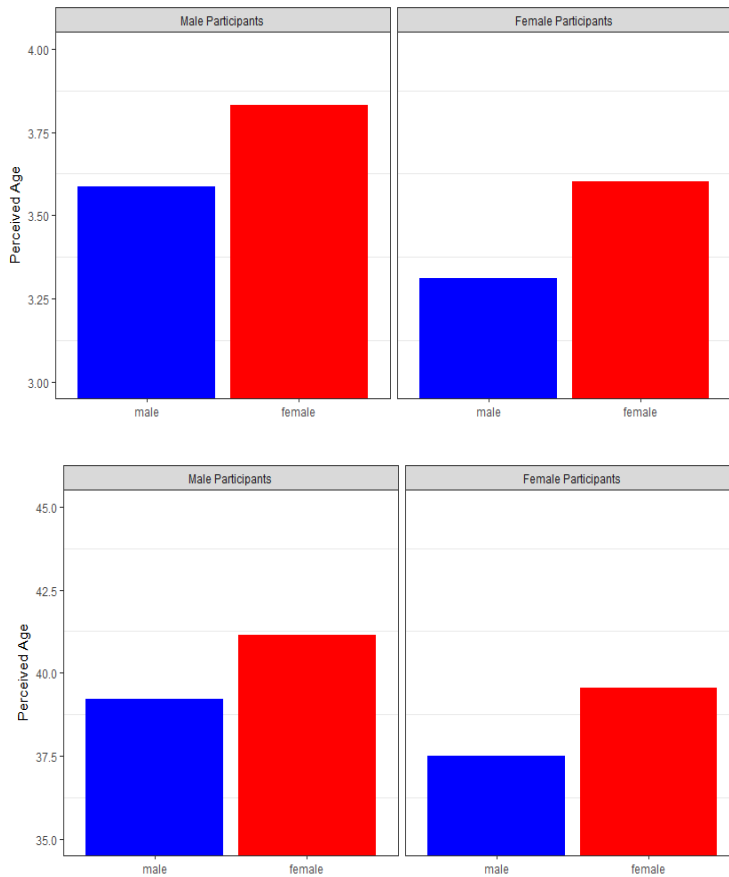


Figure 3. Female and male participants' mean age estimation on eight-point slider (upper panel) and two-digit age estimation task (lower panel) in Study 2a.

INDEPENDENT SAMPLE T-TEST WITH CONFIDENCE INTERVALS FOR STUDY 2A AND

STUDY 2B

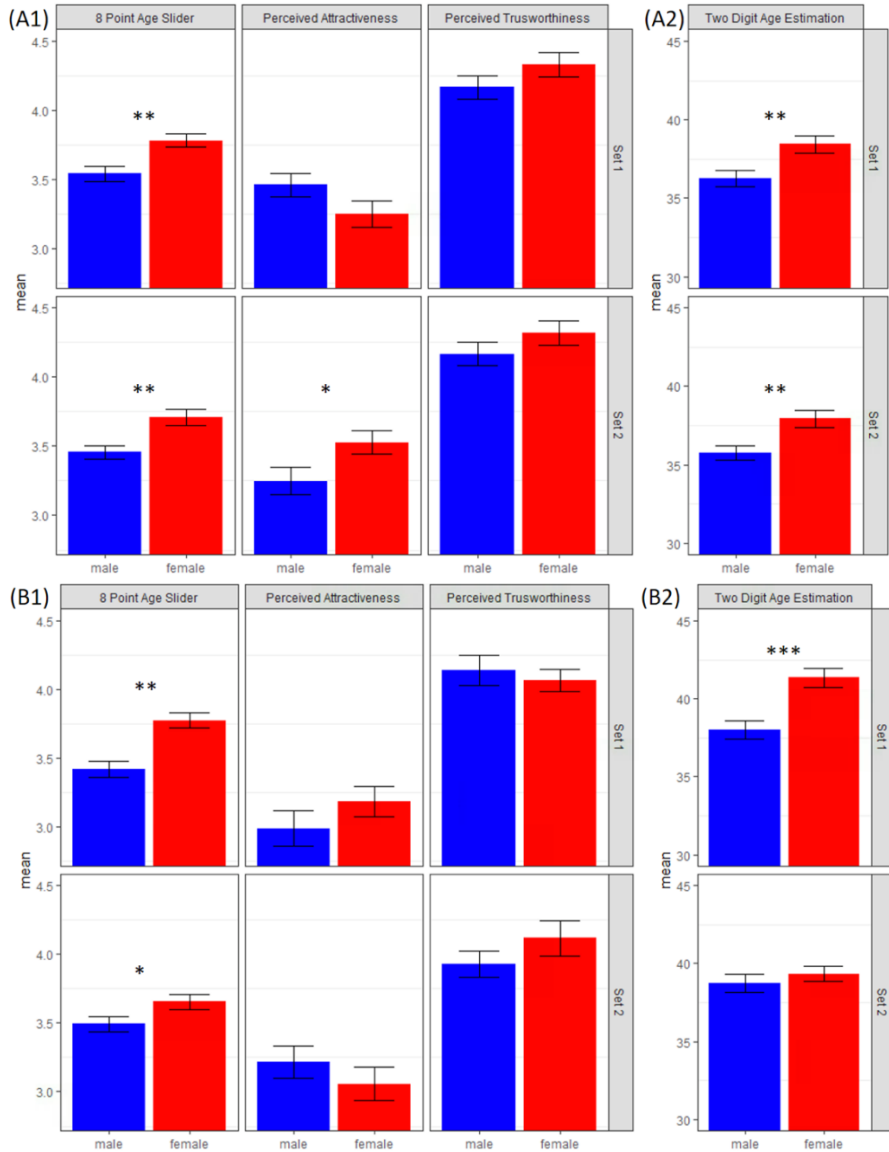


Figure 4. Independent sample t-tests comparing participants' response to the same faces when paired with female or male names (Study 2a), or gender pronouns (Study 2b).

Panel A1: Participants' response on 8-point age slider, perceived attractiveness, and perceived trustworthiness in Study 2a. Panel A2: Participants' response to two-digit age estimation in Study 2a. Panel B1: Participants' response on 8-point age slider, perceived attractiveness, and perceived trustworthiness in Study 2b. Panel B2: Participants' response to two-digit age estimation in Study 2b. (error bars = standard error of the mean, '***' for $p < 0.001$, '**' for $p < 0.01$, '*' for $p < 0.05$, ns is not marked).

MULTILEVEL ANALYSIS FOR STUDY 2A

T-Tests show that, for the average participant, pairing an androgynous face with female names made a face appear older. However, the effect of gender label might be stronger for some participants as compared to others. This section explicitly models the effect of gender label separately for each participant using hierarchical models. Specifically, the dependent variable (e.g., perceived age) as estimated by participant j for stimulus i was modeled as a function of the gender label the stimulus is paired with ($Gender_{ij}$).

The model is formally expressed as below, variables with names starting with g are level 1 estimates of global slopes and intercept, variables starting with u are level 2 estimates of individual-to-individual variations, variables starting with r are model residuals:

$$DV_{ij} = (g_{00} + u_{0j}) + (g_{10} + u_{1j}) * Gender\ Label_{ij} + r_{ij} \quad (1)$$

The model parameters were estimated using the lme4 package in R (Bates, 2005). Results (Table) confirmed observations from Independent Sample T-Tests in the previous section. Participants on average rated faces paired with female names as significantly older on both the 8-point age slider and during two-digit age estimation. Interestingly, the random effects of gender label for those two measures were not significant across participants, indicating that the effect of gender label on perceived age was relatively homogenous across participants. Confirming findings from independent sample T-Tests, gender label did not have a significant influence on perceived attractiveness. Being paired with female names made a face appear significantly more trustworthy. However, considering that the effect of gender label on perceived trustworthiness was not significant in the independent sample T-Test, we refrained from interpreting this effect.

Table 1 Predicting the four DVs as a function of gender label in hierarchical models

	Study 2a		Study 2b	
	Fixed Effects	Random Effects	Fixed Effects	Random Effects
	Average participant's level-1 gender effects	SD of estimates across participants	Average participant's level-1 gender effects	SD of estimates across participants
DV: 8 Point Age Slider				
Intercept	3.74***	0.28	3.71***	0.19
Gender Label	-0.24***	0.01	-0.26***	0.02
DV: Two-Digit Age Estimation				
Intercept	38.18***	2.88	40.33***	2.30
Gender Label	-2.17***	0.34	-1.97**	0.04
DV: Perceived Attractiveness				
Intercept	3.38***	0.81	3.12***	1.04
Gender Label	-0.03	0.25	-0.01	0.44***
DV: Perceived Trustworthiness				
Intercept	4.32***	0.80	4.09***	0.96
Gender Label	-0.15***	0.06	-0.05	0.33***

Note. '***' for $p < 0.001$, '**' for $p < 0.01$, '*' for $p < 0.05$, *ns* is not marked

CROSS-CLASSIFIED ANALYSIS FOR STUDY 2A

In addition to random variations between participants, we also fitted a cross-classified model that also captures the random variations between stimuli. This cross-classified model takes into consideration that 1) presumed gender is nested within participant (each participant always sees a set of faces paired with female name, and another set paired with male names), 2) individual stimuli (i.e., face) are nested within presumed gender (i.e., across responses from all participants, the same stimuli appeared as either female or male). This model is formally expressed as below, in which i represents a response on a given trial, j represents each participant, and k represents each stimulus (i.e., face).

$$DV_{ijk} = (\theta_0 + b_{00j} + c_{00k}) + (\theta_1 + b_{10j} + c_{10k}) * PresumedGender_{ijk} + e_{ijk} \quad (2)$$

The parameter estimates from the above model are presented in “Study 2a” of Table 2. There was a highly significant main effect of presumed gender on the 8-point age slider and two-digit age estimation tasks. There are also significant variations in the effect of gender between different stimuli, indicating that the effect of presumed gender might be different for some stimuli (i.e., faces) than others.

This approach also allowed us to explore individual differences between participants. In the appendix, we report the effect of *participant* gender and age on the effect of presumed age.

Table 2 Predicting the four DVs as a function of gender label in cross-classified random effects model

	Study 2a			Study 2b		
	Fixed Effects	Random Effects		Fixed Effects	Random Effects	
	Average participant's level-1 slope for average stimuli	SD of slopes across participants	SD of slopes across stimuli	Average participant's level-1 slope for average stimuli	SD of slopes across participants	SD of slopes across stimuli
DV: 8 Point Age Slider						
Intercept	3.74***	0.43	1.26*	3.71***	0.41	1.41**
Gender Label	-0.24***	0.01	0.13*	-0.25***	0.05	0.12**
DV: Two-Digit Age Estimation						
Intercept	38.18***	4.73	14.04	40.02***	4.71**	15.04*
Gender Label	-2.17***	0.56	0.64	-1.87***	2.55**	0.92*
DV: Perceived Attractiveness						
Intercept	3.38***	0.83***	0.52***	3.14***	1.08***	0.67***
Gender Label	-0.03	0.36***	0.24***	-0.02	0.56***	0.21***
DV: Perceived Trustworthiness						
Intercept	4.32***	0.80	0.19	4.10***	0.97**	0.16
Gender Label	-0.15***	0.06	0.04	-0.05	0.34**	0.05

Note. '***' for $p < 0.001$, '**' for $p < 0.01$, '*' for $p < 0.05$, *ns* is not marked

MULTILEVEL ANALYSIS FOR STUDY 2B

We also used a hierarchical model to examine whether the effect of gender label on perceived age is homogenous across participants. For details on model specification and model fitting see Study 2a hierarchical model.

Results are reported in Table 2 right-hand side panel. Consistent with results from Study 2a, being paired with the female gender label made a face appear significantly older. The gender slope didn't vary significantly across participants, indicating that the effect of gender label on perceived age was relatively homogenous across participants. Gender label did not have a significant influence on perceived attractiveness or perceived trustworthiness.

CROSS-CLASSIFIED ANALYSIS FOR STUDY 2B

In order to examine if participants tend to rate faces referred to as "she" as being older (as compared to faces referred to as "he"), according to the statistical analysis plan specified in the pre-registration, we modeled participants' perceived age of each face as a function of whether it is presented as female or male, specifying both participant and stimulus as random effects in a cross-classified random effects model (for details see Study 2a).

The parameter estimates from the above model are presented in "Study 2b" of Table 2 Predicting the four DVs as a function of gender label in cross-classified random effects model. There was a highly significant main effect of presumed gender on the 8-point age slider and two-digit age estimation tasks. Presumed gender has no significant effect on either perceived attractiveness or perceived trustworthiness.

CHAPTER 3 APPENDIX

Supplementary Studies:

1. Study s1: Voting Intention towards Younger vs. Older Candidates with the Same Gender
2. Study s2: Voting Intention towards Female vs. Male Candidates Approximately the Same Age
3. Study s3: Voting Intentions towards 93 Candidates with Profiles
4. Study s4: Ratings of the Initial Set of 365 Stimuli and Their Relations to Voting Intentions

Supplementary Analysis:

5. The Effect of *Participant Gender*
6. The Effect of *Participant Age*
7. Lowess Plots with Bootstrapped 95% Confidence Interval, Where Voting Intention is Grand Mean Centered
8. Multilevel Modeling of the Effects of Age on Voting Intentions Without Mediators
9. Instantaneous Mediation Effects in Study 3
10. Candidate as Unit of Analysis

Stimuli/Materials (Other Available Upon Request):

11. Pairings of Stimuli Used in Study s1 and Study s2
12. Common Names Used in Study s3 and Study 4

STUDY S1: VOTING INTENTION TOWARDS YOUNGER VS. OLDER CANDIDATES WITH THE SAME GENDER

Method

Participants

Participants were 104 MTurk workers. Three participants were excluded due to failure to complete the study; two participants were excluded due to extremely small variation in responses (less than 5% of the entire range) across all stimuli. Ninety-nine participants were included in the final analysis (56 female and 43 male, 86 self-reported as White, 4 as Black or African American, 6 as East Asian, 1 as American Indian/Native, 1 as East Asian, and 1 as mixed race, $M_{age} = 38.51$ years, $SD_{age} = 12.03$ years).

Data were collected in late 2015. All participants had IP addresses located within the United States. No analyses were conducted before data collection was completed.

Procedure

The procedure used in Study s1 was similar to that used in Study 2 except that the candidates were presented in pairs, and participants were asked to “Imagine that the two candidates are running for an important position in the government, which one would you vote for?” Participants saw fifty pairs of faces, displayed one pair at a time on the screen. There were two types of pairs. *Same gender – different age* pairs presented two candidates of the same gender that were perceived on average as ten years apart, and *same age – different gender* pairs presented pairs of targets that were from the same perceived age category, but with a different gender. Below each pair of photos, participants were presented with a 100-point slider going

from “definitely vote for A” to “definitely vote for B.” They were asked to move the slider to indicate their relative voting intention between the two candidates. A total of 50 pairs of candidates were presented for each participant.

Results

Consistent with the monotonically decreasing pattern observed in Study 1 and Study 2, participants showed a preference for 30-40 year old female candidates over 40-50 year old female candidates, $t(98) = 4.99, p < .001$, 95% CI of differences = [-12.38, -5.34]; they were also more likely to vote for 40-50 year old female candidates over 50-60 year old female candidates, $t(98) = -6.62, p < .001$, 95% CI of differences = [-13.10, -7.06]; and 50-60 year old female candidates over 60-70 year old female candidates, $t(98) = 4.76, p < .001$, 95% CI of differences = [-9.18, -3.78] (see Figure 1, Panel A).

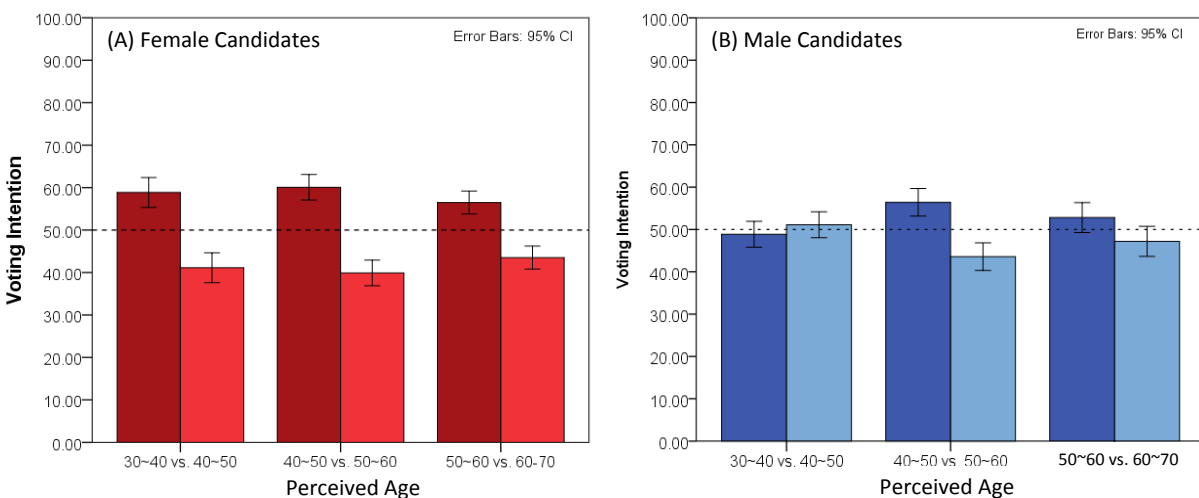


Figure 1. Voting intention towards the candidate in a hypothetical election with two candidates.

Panel A: Voting intention towards female candidates. Panel B: Voting intention towards male candidates. Error bars represent 95% confidence intervals. Darker color represents voting intention towards the younger candidate.

For male candidates, the relationship between voting intention and age was also consistent with the inverted U shape in Study 1 and Study 2. Participants were slightly, but not statistically significantly, less likely to vote for 30-40-year-old male candidates than for 40-50 year male old candidates. They were significantly more likely to vote for 40-50 year old males than 50-60 year old males, $t(98) = 3.94, p < .001, 95\% \text{ CI of differences} = [-9.42, -3.11]$. They were slightly, but not statistically significantly, more likely to vote for 50-60-year-old male candidates compared to 60-70-year-old male candidates (see Figure 1, Panel B). Those results were consistent with the results of Studies 1a and 1b, in that the effect of candidate age on voting intention was dependent on the gender of the candidate.

Also consistent with results from Study 1 and Study 2, candidate age-related decline for female candidates was less steep for female compared to male participants.

STUDY S2: VOTING INTENTION TOWARDS FEMALE VS. MALE CANDIDATES

APPROXIMATELY THE SAME AGE

In the main text, we focused on the age effects and how they differed between female and male candidates. Here, we report results examining the simple effects of gender, while holding age relatively constant. When participants voted for pairs of candidates from the same age bin, participants showed a significant preference for female candidates over male candidates in the 30-40 age category, $t(98) = -4.51, p < .001, 95\% \text{ CI} = [-12.39, -4.82]$. In the 40-50 age category, participants' preferences were not significantly different between female and male candidates, $t(98) = -1.93, p > .05, 95\% \text{ CI} = [-6.43, .08]$. In the 50-60 and 60-70 age categories, however, participants were significantly more likely to vote for male candidates over female candidates ($t(97) = 2.47, p < .05, 95\% \text{ CI} = [1.00, 9.18]$; $t(98) = 3.50, p < .001, 95\% \text{ CI} = [2.54, 9.18]$) (see

Figure 2). These results are consistent with what was observed in the main paper, in that the effect of candidate age on voting intention was dependent on the gender of the candidate.

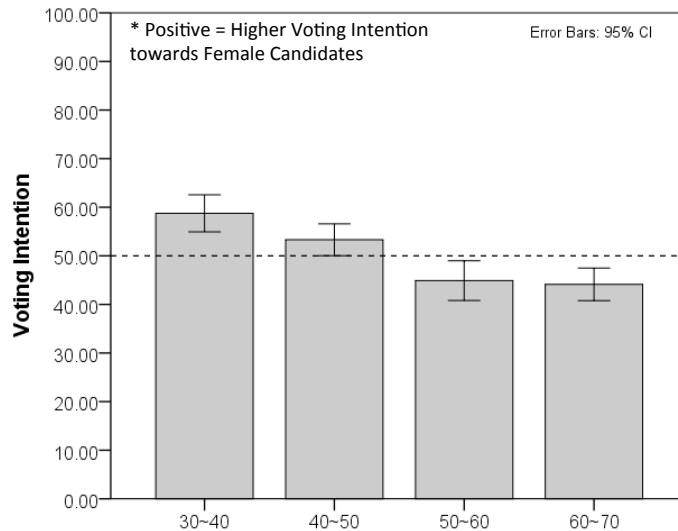


Figure 2. Voting intentions towards female as compared to male candidates in a hypothetical election featuring two candidates of about the same perceived age.

STUDY S3: VOTING INTENTIONS TOWARDS 93 CANDIDATES WITH PROFILES

Method

Participants

140 Amazon Mechanical Turk workers participated in the study. 16 participants were excluded due to failure to complete the study, leaving a final sample size of 124 (73 male, 51 females, $M_{age} = 38.63$ years, $SD_{age} = 11.56$ years, 1 American Indian/Alaska Native, 8 East Asian, 11 Black or African American, 102 White, 2 mixed race).

Data were collected in early 2016. All participants had IP addresses located within the United States. No analyses were conducted before data collection was completed.

Stimulus Selection and Preparation

Policy Statements. We drew from county-level campaign brochures to create an initial list of one-sentence policy statements for the candidates. These policy statements covered a wide range of topics (e.g., public safety, education, environment, infrastructure, housing, government, employment/business). Each statement was edited to focus exclusively on one topic.

To ensure that policy statements are not extremely conservative or liberal, gender-typed, or age-typed, in a pilot study Amazon Mechanical Turk workers (N= 30) rated the policy statements on the following dimensions in randomized orders: “Is this person more likely to be liberal or conservative?” (1= Very Conservative, 5= Very Liberal), “Do you think this person described in the profile is more likely to be a man or a woman?” (1= I am certain this person is a woman, 5= I am certain this person is a man), “Do you think this person described in the profile is more likely to be a younger candidate or an older candidate?” (1= Very Young, 5= Very Old). Policy statements that fall outside of 1 SD above or below the mean on any of the three items were excluded from the final stimulus set. The final stimulus set included 40 policy statements (e.g., “His/Her mission is to keep neighborhoods safe by investing in community programs and increase neighborhood cohesion.”) under seven broad topics.

List of Common Names. Fifty common female and male first names and 50 common last names were selected from social security administration’s database of common names (between the year 1940 and 1980). The first and last names were then randomly paired together to make hypothetical politician names.

Profile Photos. The same 93 profile photos used in previous studies were also used as stimuli in study s3.

Procedure

The study was conducted on Millisecond Inquisit Web platform (Inquisit 5 Web [Computer Software], 2016). Participants were presented with 93 profile photos of the candidates one at a time. Each photo was paired with a randomly selected name from a list of common names, and the average estimated age for each photo obtained in Study 1. In addition, each photo was also paired with two randomly selected non-repetitive policy statements. Participants were asked to indicate their voting intention towards each of the candidates on a 100-point scale “Imagine that the following candidates are running for an important position in the state legislature. Please read the summary carefully and indicate how likely are you to vote for each candidate” (1= *Definitely would not vote*, 100= *Definitely would vote*). Finally, participants reported their gender, age, and race/ethnicity.

Results

Figure 3 shows voting intentions plotted as a function of average perceived age estimated by participants in Study 1. Visual inspection shows that, even when policy statements from the candidates were present, gender differences are present in the effects of perceived age on voting intention similar to what was seen in Studies 1a-b. However, the effect appears to be smaller in magnitude with wider confidence intervals.

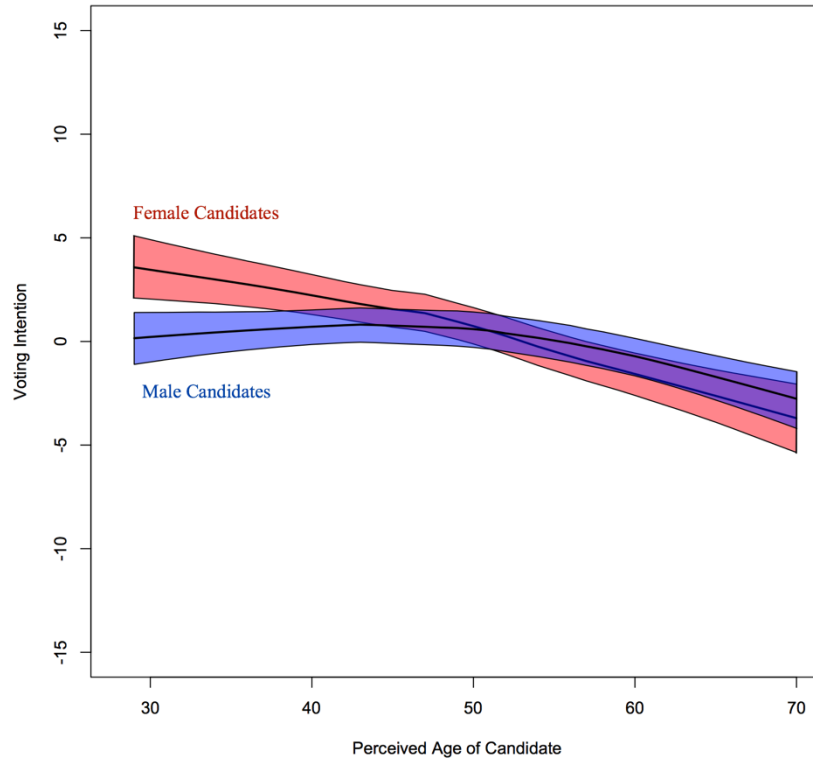


Figure 3. The relationship between perceived age of a candidate and the intention to vote for the candidate.

X-axis represents the age of the candidates as indicated in the candidate profile (based on the age estimated by participants in Study 1). Y-axis represents voting intention on a 100 point scale mean-centered within each participant. Shaded areas represent 95% bootstrapped confidence interval.

Multilevel modeling indicated a significant effect of perceived age and quadratic term of perceived age, the quadratic term of perceived age was qualified by an interaction with gender, in that voting intention decreased with perceived age for female candidates, while it increased slightly and then decreased with perceived age for male candidates.

Table 1
Multilevel Modeling Predicting Voting Intention as a Function of Candidate Gender and Age When Policy Statements from Candidates were Present (Study s3)

	Average participant's level-1 slope coefficient	t(df)	p ₁	SD of slopes across participants	Chi-Square	p ₂
Intercept	55.56	46.71(123)	<.001	13.01	2834.74	<.001
Gender	-0.14	-0.127 (123)	.900	12.12	693.90	<.001
Age	-0.13	-4.546 (123)	<.001	0.31	674.00	<.001
Age ²	-0.004	-2.773(123)	.006	0.01	209.00	<.001
Age ² x Gender	-0.004	-2.456(123)	.015	0.004	93.12	>.050
Age x Gender	0.10	3.484(123)	<.001	0.172	162.40	.010

Notes. p₁ :Probability of a t value that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p₂: Probability of a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the slopes across participants is 0. Age was centered and scaled. Gender was coded such that female = -0.5, male = 0.5.

STUDY S4: RATINGS OF THE INITIAL SET OF 365 STIMULI AND THEIR RELATIONS TO VOTING INTENTIONS

In a pilot study, the ages of the initial set of 365 stimuli were rated by 15 Amazon Mechanical Turk workers, as described in Study 1 “Stimulus Selection and Preparation” section. These 15 Amazon Mechanical Turk workers also indicated their voting intention towards the same set of 365 stimuli. Although the sample size is too small to warrant reliable conclusions, we nonetheless examined the effects of candidate perceived age and gender on voting intention among these 15 MTurk workers and report those preliminary results below.

Fifteen Amazon Mechanical Turk workers (8 women, 7 men, mean age of the participants was 42.73 years, $SD = 9.95$ years, 12 self-reported as being White, 2 as Black or African American, 1 as East Asian), who had IP addresses located within the United States, first rated all 365 stimuli, one at a time, on their perceived age. They were then asked to indicate the likelihood of their voting for each of the 365 candidates. Their voting intention was plotted against the perceived age of each candidate with Lowess smoothing with 95% bootstrapped confidence intervals (see Figure 4). Results showed that, while voting intention decreased with perceived age for both female and male candidates, the decrease was faster for female candidates (see Figure 4).

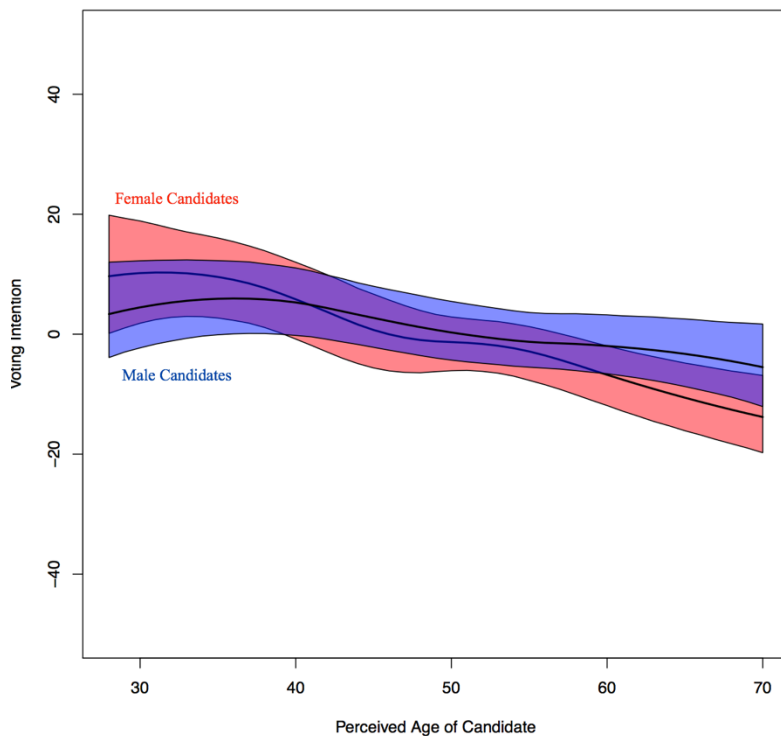


Figure 4. The relationship between the perceived age of the candidates and participants' intention to vote for them (Lowess smoothing with 95% bootstrapped CI) for all 365 stimuli

We then formally modeled the interaction between candidates' perceived age and gender for predicting voting intention. Given that the effects of perceived age on voting intention in Figure 4 appear mostly linear, we did not include the quadratic terms in the multilevel model (see Table 2).

Results revealed a significant effect of perceived age, which was qualified by a significant interaction with the gender of the candidate. These results were consistent with the finding that gender moderated the effect of age on voting intention (Study 1) in a much larger stimulus set, albeit in a much smaller sample of participants, thus demonstrating that the results reported in the main text are not limited to the specific stimulus set we selected. Note that unlike the studies reported in the main text, the quadratic effect of male candidate age was not found. We refrain from interpreting this, however, because the sample size for this pilot study was very small (only 15 participants), compared to the main studies for which the sample sizes ranged from 100 to 200 participants.

Table 2.
Multilevel Modeling Predicting Voting Intention as a Function of Candidates' Age and Gender for All 365 Stimuli (N of participants = 15)

	Average participant's level-1 slope coefficient	t (df)	p1	SD of slopes across participants	Chi-Square	p2
Intercept	45.82	10.76(14)	<.001	16.99	1663.32	<.001
Age [†]	-6.60	-4.62(14)	<.001	5.48	189.67	<.001
Gender ^{††}	1.20	0.21(14)	0.831	22.11	1601.45	<.001
Age x Gender	3.83	2.83(14)	0.013	4.95	94.46	<.001

Notes. p_1 : Probability of obtaining a t value that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p_2 : Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the slopes across participants is 0. [†] Age was centered. ^{††} Gender was coded such that female = 0, male = 1.

THE EFFECT OF PARTICIPANT GENDER

In the multilevel modeling of all of the results reported here and in the main text, for virtually all effects estimated, level-1 coefficients varied significantly between participants more than would be expected by chance. Here, we explore the role of one of individual difference variables: participant gender.

Study 1

We split data from Study 1 based on participant gender and plotted the effect of age on voting intention separately for female and male participants using LOWESS (see figure 5, Panel A and Panel B). While the gender \times age interaction effect was qualitatively the same for female and male participants, male participants' voting intentions towards female candidates declined

more steeply with increasing candidate age, while the decline was less steep for female participants (see figure 5).

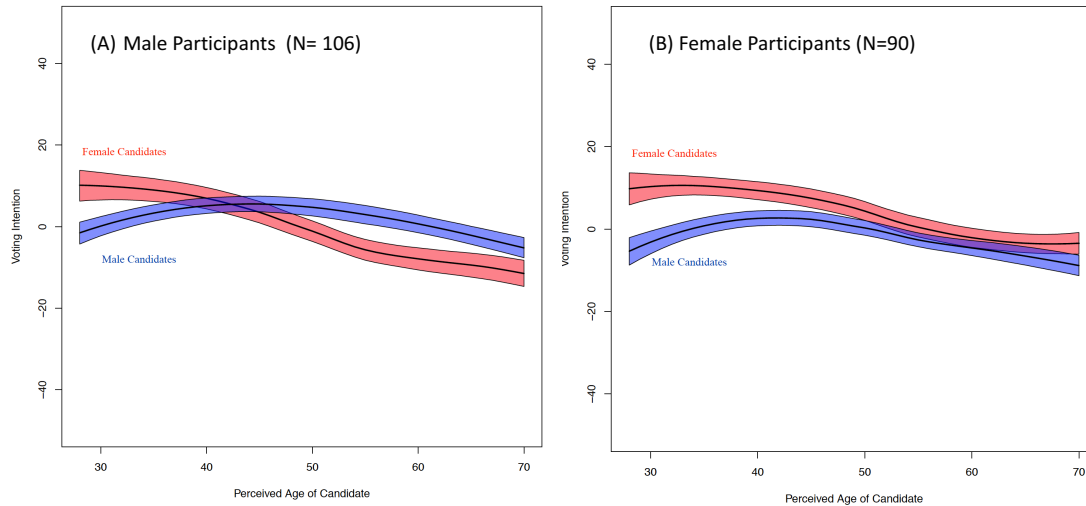


Figure 5. The relationship between likelihood to get votes and perceived age of the candidates in Study 1. Panel A: Male Participants, Lowess curve (95% CI). Panel B: Female Participants, Lowess curve (95% CI).

Multilevel modeling using participant gender as a level 2 predictor (Table 3) indicated that participant gender explained a significant amount of between-participant variation in the effect of age. However, it is worth noting that the SDs of the coefficients across participants remained significant even after participant gender was taken into consideration, indicating that other explanatory factors are needed to fully account for the between-participant variation observed in our data.

Table 3.

Multilevel modeling: predicting voting intention as a function of candidate age and gender as level 1 predictors, and participant gender as level 2 predictor.

	Average participant 's level-1 coefficient	t(df)	p1	Participant gender predicting each coefficient	p2	SD of coefficients controlling for participant gender	p3
Intercept	52.69	48.72 (194)	<.001	7.96	<.001	14.44	<.001
Age [†]	-0.52	-11.52 (194)	<.001	0.15	0.082	0.58	<.001
Gender [†] [†]	1.24	1.009 (194)	0.314	-9.59	<.001	16.12	<.001
Age x Gender	0.35	9.734 (194)	<.001	-0.19	0.008	0.39	<.001
Age*Age	0.0009	0.486 (194)	0.627	-0.003	0.392	0.02	.001
Age*Age * Gender	-0.016	-7.902 (194)	<.001	0.008	0.051	0.01	.021

Notes. p_1 Probability of obtaining an average level-1 coefficient that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p_2 Probability of obtaining an SD that is as large or larger than the value obtained, given the null hypothesis that the difference in Level-1 coefficient between female and male participants is 0. p_3 Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the coefficient across participants is 0. [†] Female is coded as 0 and male coded as 1. ^{††} Age is scaled and mean centered. Voting intention is mean centered.

Study 2

A procedure identical to that used to examine participant gender effect in Study 1 was used to examine participant gender effect in Study 2. The results and parameter estimated are presented below:

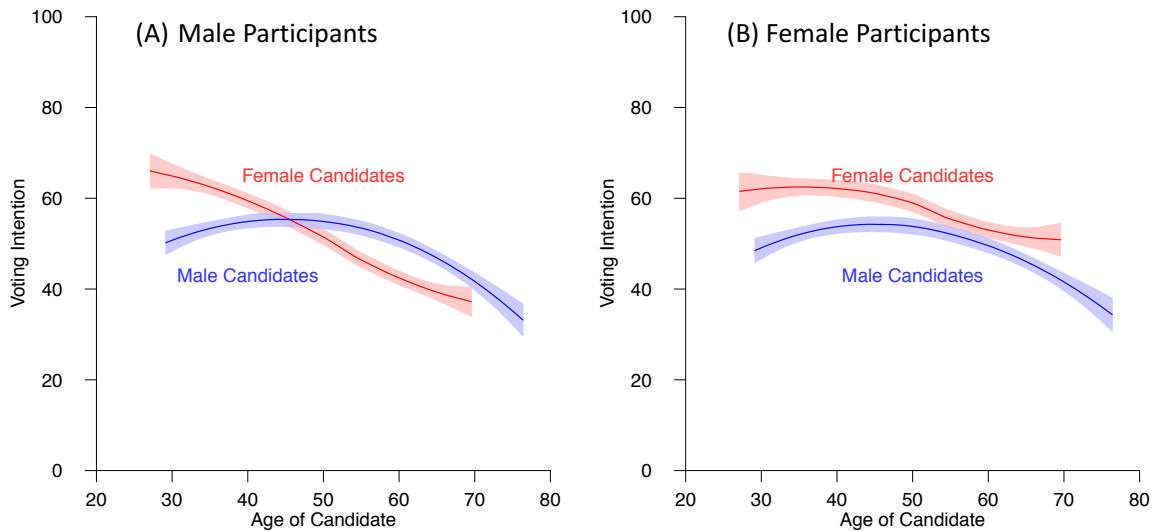


Figure 6. The relationship between likelihood to get votes and perceived age of the candidates in Study 2. Panel A: Male Participants, Lowess curve (95% CI). Panel B: Female Participants, Lowess curve (95% CI).

Results confirmed observations from Study 1, in that male participants' voting intentions towards female candidates declined more steeply with increasing candidate age, while the decline was less steep for female participants.

Table 4.

Multilevel modeling: predicting voting intention as a function of candidate age and gender as level 1 predictors, and participant gender as level 2 predictor.

	Average participant's level-1 coefficient	t(df)	p1	Participant gender predicting each coefficient	p2	SD of coefficients controlling for participant gender	p3
Intercept	54.51	48.73	<0.001	2.88	0.20	10.65	<0.001
Age [†]	-0.39	-7.72	<0.001	0.22	0.04	0.47	<0.001
Gender [†]	-0.08	-0.05	0.96	-8.44	0.01	14.67	<0.001
Age x Gender	0.36	6.36	<0.001	-0.38	0.00	0.45	<0.001
Age*Age	-0.01	-4.78	<0.001	0.00	0.80	0.02	<0.001
Age*Age * Gender	-0.02	-5.47	<0.001	0.01	0.38	0.02	0.01

Notes. p_1 Probability of obtaining an average level-1 coefficient that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p_2 Probability of obtaining an SD that is as large or larger than the value obtained, given the null hypothesis that the difference in Level-1 coefficient between female and male participants is 0. p_3 Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the coefficient across participants is 0. [†] Female is coded as 0 and male coded as 1. ^{††} Age is scaled and mean centered. Voting intention is mean centered.

Study 3

According to pre-registration plan, we also examined whether the effect of candidate pair gender on voting intention is different for female participants as compared to male *participants*, and younger participants compared to older *participants*.

To test the effect of participant gender, we included participant gender as a predictor of the gender slope, and of the overall intercept, in the form of the following model:

Level 1:

$$\text{Ln}(P_{ij} / (1 - P_{ij})) = b_{0j} + b_{1j} \text{PairGender}_{ij}$$

Level 2:

$$b_{0j} = g_{00} + g_{01} \text{ParticipantGender}_j + u_{0j}$$

$$b_{1j} = g_{10} + g_{11} \text{ParticipantGender}_j + u_{1j}$$

The parameters were estimated using lme4 package in R. Participant gender had a significant effect on both the intercept ($g_{01} = 0.58$, $SE = 0.15$, $p < .001$), and the slope of candidate gender ($g_{11} = -0.60$, $SE = 0.16$, $p < .001$). We plotted the mean and within-person standard error of the mean separately for female and male participants in Figure 7. While female participants preferred the younger female candidate and younger male candidates almost equally, that is not the case for male participants; male participants showed a strong young preference for female candidates, driving the gender double standard of aging effect.

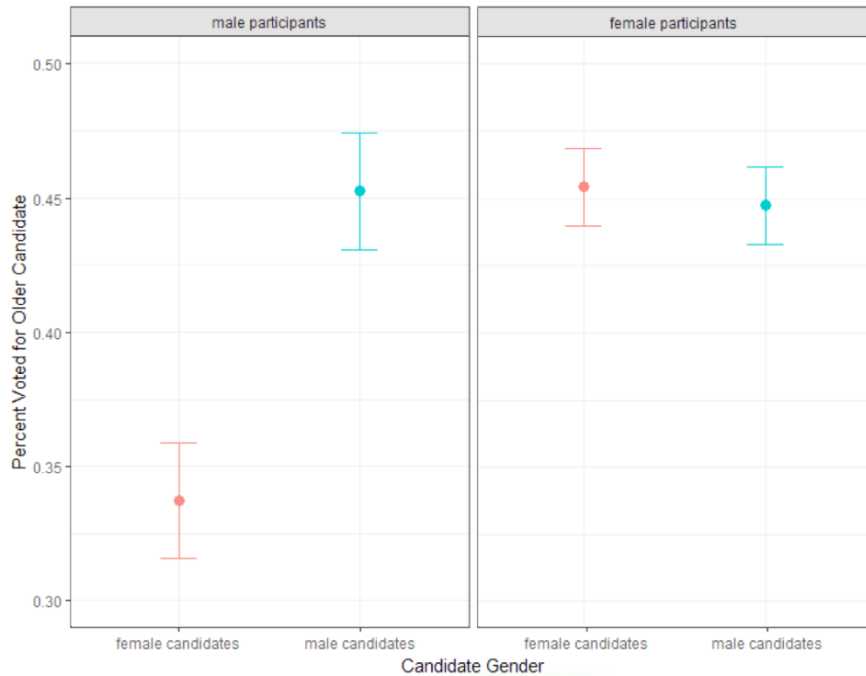


Figure 7. Percentage of participants voting for the older candidate over younger candidate in Study 3, plotted separately for female and male participants (error bars = standard error of the mean).

Study 4

A procedure identical to that used to examine participant gender effect in Study 1 was used to examine participant gender effect in Study 2. The results and parameter estimated are presented below:

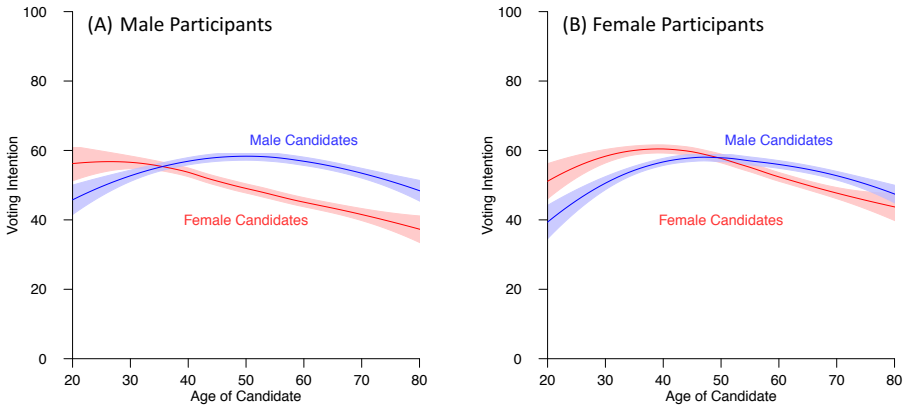


Figure 8. The relationship between likelihood to get votes and perceived age of the candidates in Study 3. Panel A: Male Participants, Lowess curve (95% CI). Panel B: Female Participants, Lowess curve (95% CI).

Results confirmed observations from Study 1, in that male participants' voting intentions towards female candidates declined more steeply with increasing candidate age, while the decline was less steep for female participants.

Table 5. *Multilevel modeling: predicting voting intention as a function of candidate age and gender as level 1 predictors, and participant gender as level 2 predictor.*

	Average participant's level-1 coefficient	t	p_1	Participant gender predicting each coefficient	p_2	SD of coefficients controlling for participant gender	p_3
Intercept	48.38	57.78	<0.001	6.89	<0.001	10.09	<0.001
Age [†]	-6.72	-10.59	<0.001	1.85	0.15	9.35	<0.001
Gender ^{††}	5.17	6.18	<0.001	-8.02	<0.001	7.49	<0.001
Age x Gender	6.56	11.70	<0.001	-2.17	0.05	3.27	<0.001
Age*Age	0.14	0.40	0.69	-2.93	<0.001	2.31	<0.001
Age*Age*Gender	-2.83	-7.77	<0.001	2.50	<0.001	5.74	<0.001

Notes. p_1 Probability of obtaining an average level-1 coefficient that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p_2 Probability of obtaining an SD that is as large or larger than the value obtained, given the null hypothesis that the difference in Level-1 coefficient between female and male participants is 0. p_3 Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the coefficient across participants is 0. [†] Female is coded as 0 and male coded as 1. ^{††} Age is scaled and mean centered.

THE EFFECT OF PARTICIPANT AGE

For all studies reported in the main paper, participant age did *not* significantly interact with candidate age effect or candidate gender effect.

Study 1

Aside from participant gender, another participant factor that might explain the variation in level-1 coefficients is participant age.

First, we split data from Study 1 at the mean of participant age (younger or older than 45 years of age) and plotted the effect of candidate age on voting intention separately for younger and older participants using LOWESS (see figure 9). As can be seen from Figure 9, the gender \times age interaction effect was qualitatively the same for younger and older participants. Multilevel modeling predicting voting intention as a function of candidate age and gender as level 1 predictors, and participant age as level 2 predictor also showed that participant age does not significantly influence the gender \times age interaction effect (see Table 6).

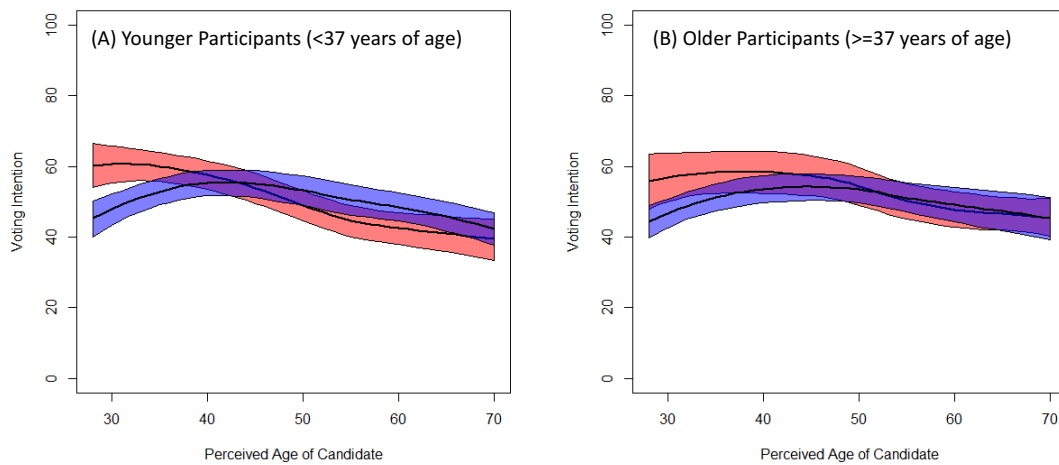


Figure 9. The relationship between likelihood to get votes and perceived age of the candidates in Study 1. Panel A: Younger Participants, Lowess curve (95% CI). Panel B: Older Participants, Lowess curve (95% CI).

Table 6.

Multilevel modeling: predicting voting intention as a function of candidate age and gender as level 1 predictors, and participant age as level 2 predictor.

	Average participant's level-1 coefficient	t(df)	p1	Participant age predicting each coefficient	p2	SD of coefficients controlling for participant age	p3
Intercept	51.57	23.32(194)	<.001	0.35	0.061	15.40	>0.5
Age†	-0.87	-11.37(194)	<.001	0.01	0.034	0.994	>0.5
Gender†	1.21	0.94(194)	0.344	-0.16	0.167	6.09	>0.5
†							
Age x Gender	0.35	9.34(194)	<.001	-0.006	0.052	0.45	>0.5
Age*Age	0.01	4.30(194)	<.001	-0.0006	0.067	0.05	>0.5
Age*Age * Gender	-0.01	-7.09(194)	<.001	0.0002	0.175	0.03	>0.5

Notes. p_1 Probability of obtaining an average level-1 coefficient that is as large or larger than the value obtained, given the null hypothesis that the average level-1 slope is 0. p_2 Probability of obtaining an SD that is as large or larger than the value obtained, given the null hypothesis that the difference in Level-1 coefficient between younger and older participants is 0. p_3 Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the coefficient across participants is 0. † Female is coded as 0 and male coded as 1. †† Age is scaled and mean centered.

Study 3

We tested whether participant age influences the relationship between candidate gender and voting intention, specified in the model below:

Level 1:

$$\text{Ln}(P_{ij} / (1-P_{ij})) = b_{0j} + b_{1j} \text{PairGender}_{ij}$$

Level 2:

$$b_{0j} = g_{00} + g_{01} \text{ParticipantAge}_j + u_{0j}$$

$$b_{1j} = g_{10} + g_{11} \text{ParticipantAge}_j + u_{1j}$$

Participant age had a small effect on the overall intercept ($g_{01} = 0.01$, $SE = 0.005$, $p < .05$), in that older participants are slightly more likely to vote for the older candidate, how candidate gender influence voting intention does not depend on the age of the participant ($g_{11} = -0.01$, $SE = 0.006$, $p > .1$).

LOWESS PLOTS WITH BOOTSTRAPPED 95% CONFIDENCE INTERVAL, WHEN VOTING INTENTION IS GRAND MEAN CENTERED

Figures in the main paper mean centered voting intention within each participant to focus on the within-subject effects of perceived candidate age. Here we replotted the data using grand mean centering across all participants on all trails. Even though these plots reflect participant-to-participant differences in scale use (hence slightly wider confidence interval bands at some age points), the patterns of results described in the main paper is still clearly visible.

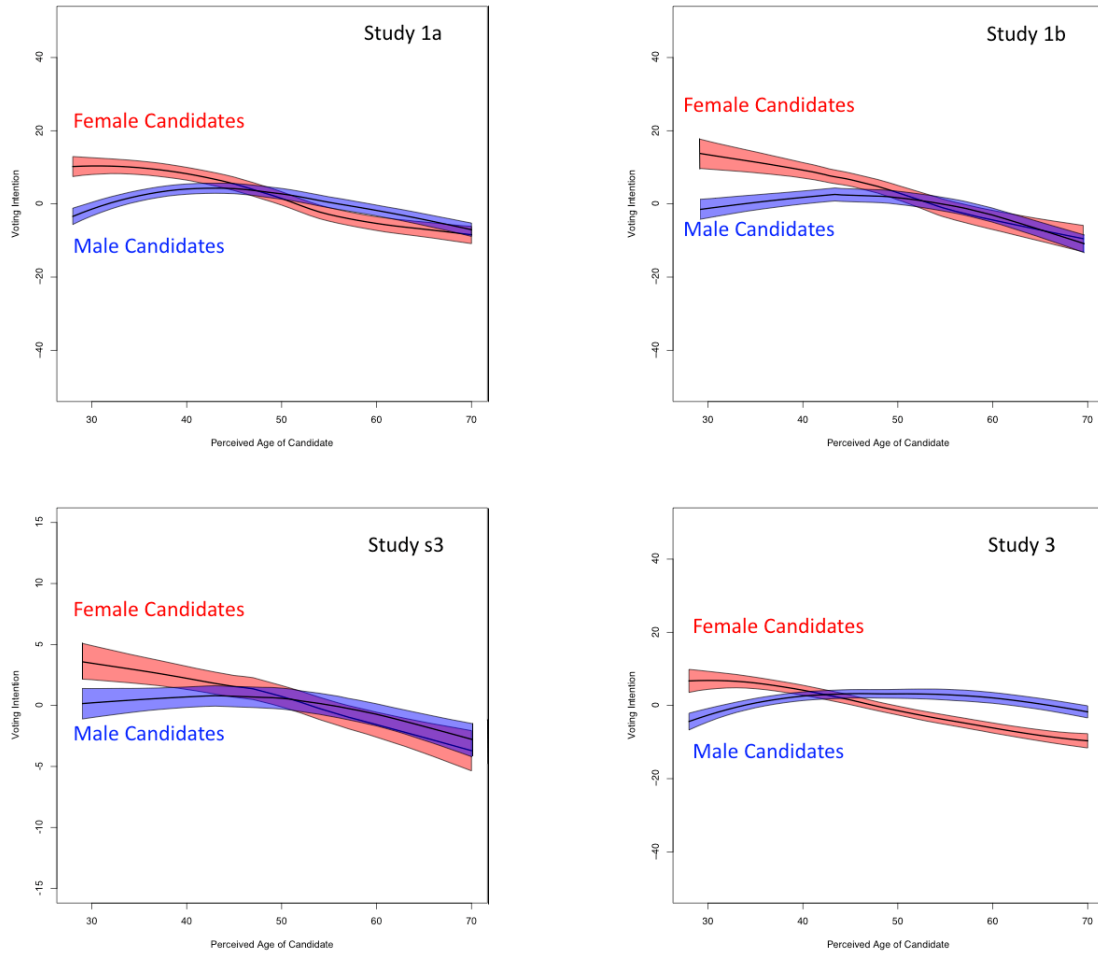


Figure 10. The relationship between perceived age and voting intention (Lowess smoothing with 95% bootstrapped CI, 5000 samples) for Study 1, Study 2, Study s3, and Study 3 when the voting intention is grand mean centered.

Y-axis represents voting intention on a 100 point scale; X-axis represents the age of the candidates as estimated by others (Study 2 and Study s3) or participants themselves (Study 1 and Study 3)

MULTILEVEL MODELING OF THE EFFECTS OF AGE ON VOTING INTENTIONS IN STUDY

3 WITHOUT MEDIATORS

Study 3 focused on testing the role of warmth, competence, and attractiveness in mediating the effect of perceived age on voting intention. We also conducted multilevel analysis

examining the potential quadratic effect of age without mediators, but due to space limitation, we did not report the results of this analysis in the main text. Thus we report them below.

We followed the same analytic strategy illustrated in the Study 1 results section, using multilevel modeling to estimate the relationship between voting intention and candidates' age, gender, a quadratic effect of age, and their interaction terms, allowing all of these effects to vary freely between participants.

Table 7.
Multilevel modeling: predicting voting intention as a function of candidate age and gender with the undergraduate sample.

	Average participant's level-1 coefficient	t(df)	p1	SD of coefficients across participants	Chi-Square	p2
Intercept	-0.06	-1.70(174)	0.091	0.46	1217.56	<.001
Age†	-0.28	-10.86(174)	<.001	0.31	877.09	<.001
Gender††	0.22	5.90(174)	<.001	0.43	726.04	<.001
Age x Gender	-.28	12.01(174)	<.001	0.24	383.74	<.001
Age*Age	0.01	0.78(174)	0.43	0.15	4-5.58	<.001
Age *Age* Gender	-0.13	-8.006(174)	<.001	0.12	271.62	<.001

Notes. p_1 Probability of obtaining a t value that is as large or larger than the value obtained, given the null hypothesis that the average level-1 coefficient is 0. p_2 Probability of obtaining a Chi-Square value that is as large or larger than the value obtained, given the null hypothesis that the SD of the coefficients across participants is 0. † Candidate age was centered within each participant. †† Gender was coded such that female = 0, male = 1.

Results confirmed the findings from Study 1 and Study 2; the effect of candidate age and quadratic effect of candidate age were qualified by an interaction with candidate gender.

INDIRECT EFFECTS THROUGH THE QUADRATIC RELATIONSHIP BETWEEN PERCEIVED AGE AND VOTING INTENTIONS

In light of the difficulty in interpreting mediation analyses in the presence of quadratic effects, in the main text, we reported the results of mediation analyses separately for the younger half, and the older half of the candidates since the relationship between age and voting intention was roughly linear within these age groups. While this approach affords the convenience of interpretation, it requires dichotomizing the data. To overcome this limitation, we re-analyzed the data using methods outlined in Hayes and Preacher (2010).

One major obstacle for analyzing mediation effects of nonlinear relationships is that the mediation effect changes based on the level of the predictor. Hayes and Preacher (2010) overcame this obstacle by calculating instantaneous indirect effect, which is defined as the influence of the independent variable (X: perceived age) on the dependent variable (Y: voting intention) through the mediator (M: perceived attractiveness) at a specific value of X. When X has a quadratic effect on M, but the effect of M on Y is linear (as is the case in Study 2), the instantaneous indirect effect through M when $X = x$ could be expressed as: $(a_1 + 2a_2x) * b$, where a_1 and a_2 are the coefficients for predicting M from X (a_1 for the linear term and a_2 for the quadratic), and b is the linear term predicting Y from M while controlling for X. Note that the indirect effect depends on the level of x. We used bootstrapping (5000 samples) to obtain the distribution of the instantaneous mediation effects at each age point (between 30 to 70 years). A two-tailed 95% confidence interval was obtained by taking the 2.5% quantile and 97.5% quantile of the distribution at each age point.

We conducted this analysis separately for each mediator. To visualize the instantaneous indirect effects, we plotted the point estimate as well as 95% CI for each variable at each age point (Figure 11). A CI that does not overlap with 0 indicates a significant gender difference in the indirect effect at a particular age point. Consistent with previous results, the mediation effect of perceived attractiveness was especially strong for younger female candidates. For older candidates, perceived competence has a slightly stronger instantaneous mediation effect for female candidates as compared to male candidates, in that an increase in perceived age is associated with greater decrease in perceived competence for older female candidates.

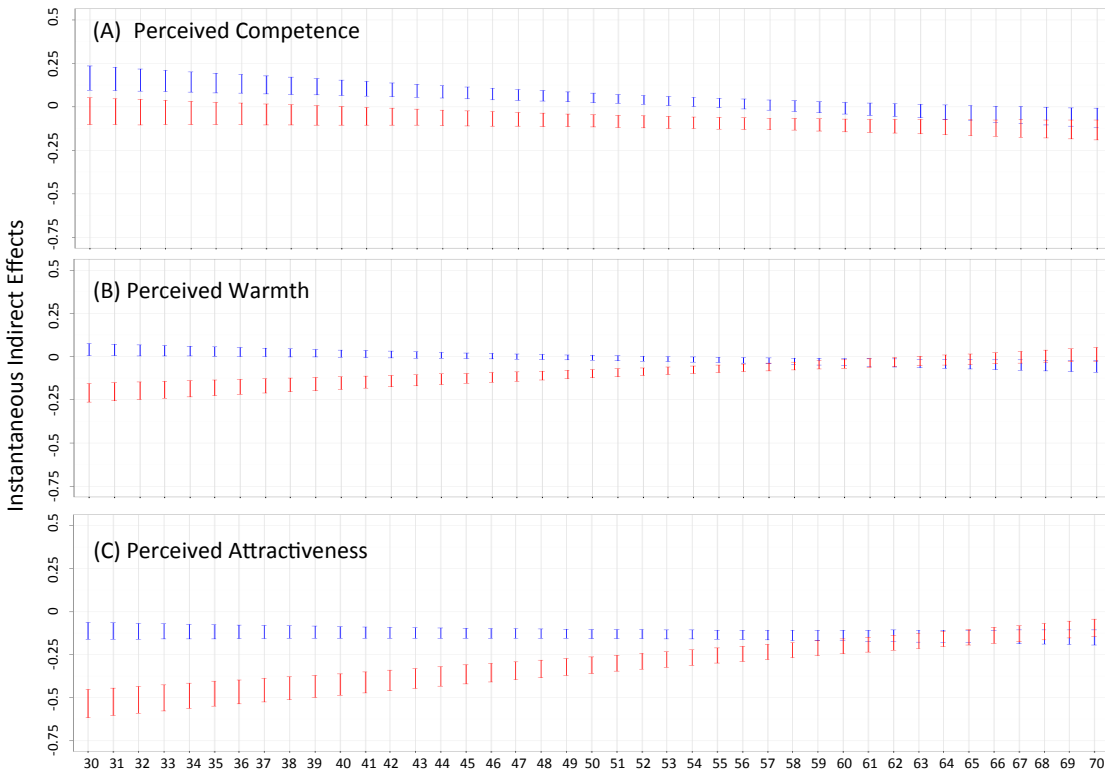


Figure 11. Instantaneous indirect effects through perceived competence (Panel A), perceived warmth (Panel B), and perceived attractiveness (Panel C) at each age point plotted with 95% bootstrapped CI.

Red bars represent CIs for female candidates, blue CIs those for male candidates.

CANDIDATE AS UNIT OF ANALYSIS

Due to the difficulty in interpreting multilevel data analysis, we also conducted a regression analysis on the level of the candidates.

Voting intention and perceived age in Study 1 and Study 2 were averaged across all participants to obtain the mean voting intention and the mean perceived age of each candidate. Perceived attractiveness, competence, and warmth were averaged across all participants in Study 2 to obtain the mean perceived attractiveness, competence, and warmth rating of each candidate.

In the first analysis, we tested whether the influence of age on voting intention towards each candidate is dependent on the candidate's gender. Voting intention towards each candidate was regressed on perceived age (and its quadratic term), gender, and their interaction terms. Results replicated findings from the main text, in that age, as well as its quadratic term, interact with gender to influence voting intention. For female candidates, voting intention decrease with perceived age, while for male candidates, voting intention goes up and then down with perceived age.

In the second analysis, we tested the relative contribution of perceived attractiveness, competence, and warmth in predicting voting intention, also using candidates as the level of analysis. Perceived attractiveness was the strongest predictor of voting intention for older female candidates, after controlling for perceived age, perceived warmth, and perceived competence; while perceived competence was the strongest predictor for voting intention for older male candidates, after controlling for perceived attractiveness, perceived warmth, and perceived age. There were no significant gender differences for candidates younger than 45.

Our finding could partially help to resolve the discrepancy between Todorov et al., 2005, which used predominantly older male candidates and found that perceived competence to be the strongest predictor of electoral success, and Berggren et al., 2010, which used predominantly female candidates in Finnish elections, and found that perceived beauty to be a stronger predictor of relative electoral success. However, more research is needed to reach a reliable conclusion about what kind of social impressions most strongly predicts voting intentions in male and female candidates.

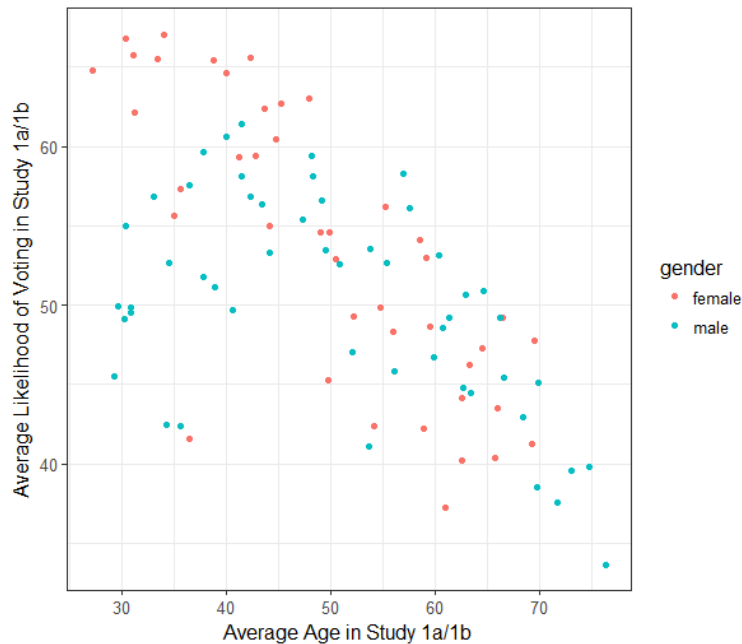


Figure 12. Voting intention towards each candidate as a function of the gender and average perceived age of the candidate.

Table 8.

Linear regression predicting voting intention as a function of candidate age and gender and their interaction terms on candidate (i.e., stimuli) level

	β	SE	t	p
Model: likelihood \sim age + age2 + gender + (age + age2) * gender				
Intercept	53.02	0.78	69.09	<0.001
Age	-5.13	0.54	-9.34	<0.001
Age2	-2.02	0.62	-3.27	0.001
Gender	0.56	1.56	0.36	0.716
Age * Gender	5.05	1.09	4.59	<0.001
Age2 * Gender	-3.18	1.24	-2.57	0.011

Notes. Age was centered and scaled. Gender was coded such that female = -0.5, male = 0.5.

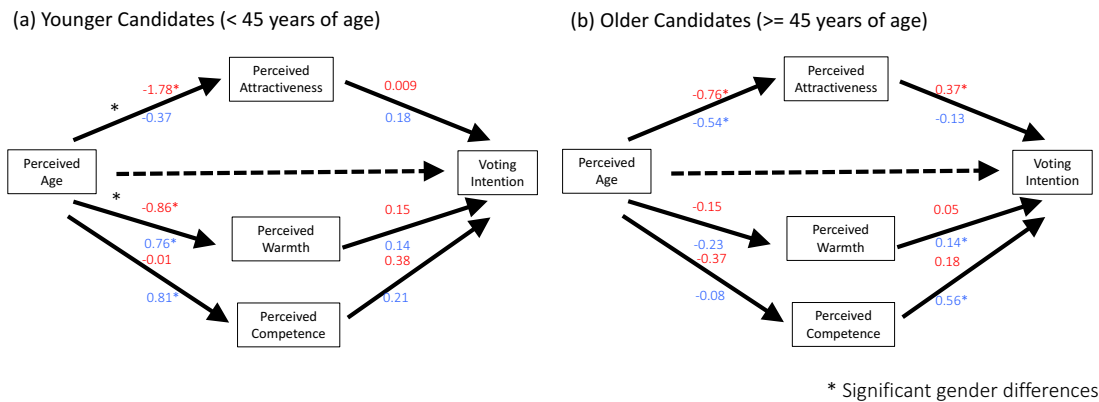


Figure 13. Summary of candidate (i.e., stimulus) level mediation analyses of the effects of perceived age on voting intention through perceived warmth, perceived attractiveness, and perceived competence simultaneously.

Perceived age and voting intention were aggregated from Study 1 and Study 2 of the main paper. Perceived attractiveness, warmth, and competence came were aggregated from Study 4 in the main paper. The left panel is for candidates perceived to be younger than 45 years of age. The right panel is for candidates perceived to be older than 45 years of age. Above each path are coefficients for female candidates, below each path are coefficients for male candidates. Asterisks indicate paths for which the coefficient for female candidates differs significantly ($p < .05$) from the coefficient for male candidates. The p values were obtained based on 5000 bootstrapped samples.

PAIRINGS OF STIMULI USED IN STUDY S2 AND STUDY S3

Pairs of candidates were compared against each other in a hypothetical election according to the scheme illustrated in table 9. This design ensures that each target appears on the right and left-hand side of the screen equally across participants. There was a total of 80 hypothetical

elections. Due to time limitations, each participant voted in 50 randomly selected hypothetical elections.

Table 9.

Number of Specific Combinations of Stimuli in Study s1

Same Gender Different Age Pairs	30-40 (right)	40-50 (right)	50-60 (right)	60-70 (right)
30-40(left)	--	8*		
40-50(left)	8*	--	8*	
50-60(left)		8*	--	8*
60-70(left)			8*	--
Opposite Gender Same Age Pairs	30-40 (male)	40-50 (male)	50-60 (male)	60-70 (male)
30-40(female)	8†			
40-50(female)		8†		
50-60(female)			8†	
60-70(female)				8†

Notes. * Half of the pairs were between females, half between males.

† On half of the trials, female appeared on the left side of the screen.

COMMON NAMES USED IN STUDY S3

All names used in Study 2 and Study s3 are constructed based on common first and last names from the social security administration (<https://www.ssa.gov>) database.

List of male politician names:

Alan Smith, Albert Johnson, Andrew Williams, Anthony Brown, Arthur Jones, Barry Miller, Bill Davis, Brandon Garcia, Carl Martinez, Charles Anderson, Chris Taylor, Craig Thomas, Daniel Hernandez, Danny Moore, Darrell Martin, David Jackson, Dennis Thompson, Donald White, Douglas Lopez, Eddie Lee, Edward Gonzalez, Eric Harris, Frank Clark, Frederick Lewis, Gary Robinson, George Walker, Gerald Perez, Gregory Hall, Harold Young, Harry Allen, Howard Sanchez, Jack Wright, James King, Jason Scott, Jay Green, Jeffery Baker, Jerry Adams, Jimmy Nelson, Joe Hill, John Ramirez, Johnny Campbell, Jonathan Mitchell, Joseph Roberts, Justin Phillips, Keith Evans, Kenneth Turner, Kevin Torres, Lawrence Collins, Lee Edwards, Leonard Stewart, Louis Flores

List of female politician names:

Alice Morris, Amber Murphy, Angela Rivera, Anita Cook, Leslie Morris, Ashley Morgan, Barbara Peterson, Betty Cooper, Beverly Reed, Brenda Bell, Brittany Gomez, Carol Kelly,

Carolyn Howard, Catherine Ward, Cathy Cox, Cheryl Diaz, Christina Richardson, Christine Wood, Cindy Watson, Connie Brooks, Crystal Bennett, Cynthia Gray, Danielle Reyes, Debbie Hughes, Deborah Price, Denise Myers, Diana Long, Donna Sanders, Doris Ross, Dorothy Morales, Elaine Powell, Elizabeth Sullivan, Ellen Russell, Erin Ortiz, Frances Jenkins, Gail Gutierrez, Gloria Perry, Heather Butler, Helen Barnes, Jacqueline Fisher, Linda Foster, Laura Web

CHAPTER 4 APPENDIX

THE PROPORTION OF UNGENDERED AND UNCERTAINLY GENDERED AUTHORS

Table 1. Number and Percentage of Ungendered and Authors by Journal

Journal	Total First Authors	Total Last Author	Gendered First Author	Gendered Last Author	Gendered First Author With 90% Certainty	Gendered Last Author With 90% Certainty	% First Author Gendered	% Last Author Gendered
Annual Review of Neuroscience	288	238	270	224	248	206	93.75	94.12
Behavioral and Brain Sciences	1536	920	1485	894	1396	839	96.68	97.17
Brain	4355	3789	4069	3591	3513	3108	93.43	94.77
Cerebral Cortex	4061	4012	3694	3725	3088	3222	90.96	92.85
Current Opinion in Neurobiology	1666	1355	1542	1258	1388	1126	92.56	92.84
Journal of Neuroscience	19533	19125	17173	17593	15061	16217	87.92	91.99
Nature Neuroscience	3522	3163	3140	2911	2773	2685	89.15	92.03
Nature Reviews Neuroscience	1576	721	1518	691	1443	646	96.32	95.84
Nature	28191	14910	25962	13830	22622	11602	92.09	92.76
NeuroImage	11530	11073	10441	10154	8313	8433	90.56	91.70
Neuron	5675	5113	5088	4746	4483	4344	89.66	92.82
Neuropsychology Review	355	279	342	270	311	239	96.34	96.77
PNAS	49954	46937	43333	42866	37011	38384	86.75	91.33
Science	28015	15302	25984	14332	22230	11786	92.75	93.66
Trends in Neurosciences	1034	895	970	838	893	781	93.81	93.63

RESULTS AFTER THRESHOLDING TO AUTHORS WITH HIGH GENDER CERTAINTY

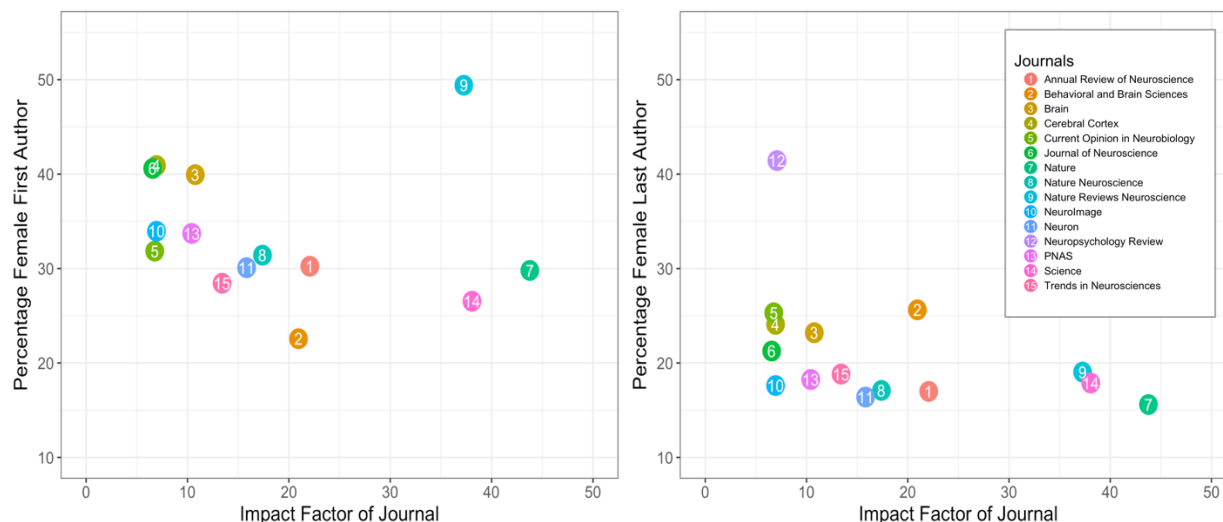


Figure 1. Percentage of female first and last authors between 2005-2017 vs. Journal’s 5-year impact factor, subsetting to authors with very high (>90%) gender certainty.

We selected only papers with either highly gendered³ first author names, or highly gendered last author names, and reran our correlation analysis. The results remained qualitatively the same. There is a negative relationship between a journal’s 5-year impact factor, and the percentage of female authors (first author $r_s = -0.55$, $p < .05$, last author $r_s = -0.50$, $p = .05$).

³ High-gendered is defined as higher than 0.9 probability that the person with a given first name has the predicted gender in the genderize.io database (Wais, 2016).

VITA

Yiqin Alicia Shen was born in Shanghai, China. She graduated from Fudan University with honors in 2011 and came to the U.S. to pursue research in psychology. In 2016, she completed a quantitative minor, and in 2018, she earned a Doctor of Philosophy in Psychology at the University of Washington. Currently, she resides in Seattle, WA with her family, pursuing a career in data science.