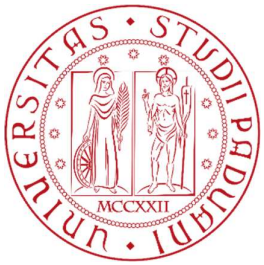


UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia dello Sviluppo e della Socializzazione

Corso di Laurea Magistrale in Psicologia di Comunità, della Promozione del Benessere e del Cambiamento Sociale

Tesi di Laurea Magistrale

Facial similarity and vote attribution:
How inferences from faces guide our judgement.

Relatrice:

Prof.ssa Luciana Carraro

Correlatore:

Prof. Luigi Alessandro Castelli

Laureando: Federico Padovan

Matricola: 2080431

Anno Accademico 2023/24

Abstract

The scientific literature (Bailenson et al., 2006; 2008) reports that increased facial similarity between an elector and a candidate enhances vote intention towards that same candidate. Moreover, scientific evidence (Rule & Ambady, 2009; 2010) suggests that people can accurately guess the political orientation of both politicians and voters at a rate significantly higher than chance.

Therefore, the current research aimed to determine whether other people use facial similarity between an elector and a candidate to infer a person's vote intention even in the absence of ideological cues. Furthermore, it was hypothesized that the gender of the observers, the electors, and the candidates would significantly mediate the effects of facial similarity.

To test the hypotheses, an experimental study was designed, which included a questionnaire consisting of 6 triplets of artificially generated faces, all of the same gender. For each triplet, the participants, a sample of N = 566 (76% women), were asked to indicate the likelihood that the person portrayed in the central photograph, described as an elector, had voted for one of two candidates, whose pictures were displayed at the bottom of the screen. Following the probability attribution task, participants were also asked to report the criteria they used while making these attributions.

The main hypothesis found support in the data, revealing a small yet significant impact of the resemblance between the elector and a candidate on the participants' vote attributions. An effect of the triplet's gender was also detected: while the effect of similarity in female triplets did not emerge, it was significant in male triplets. Furthermore, an effect of the participant's gender was also supported by the data, with male participants likely exhibiting the own-gender recognition bias (Herlits & Lovén, 2013), as the effect of similarity on vote attribution intensified when male participants evaluated male triplets.

Overall, the findings integrate previous literature on the topic and deepen our understanding of the role played by facial similarity in political inferences.

Keywords: Political Psychology, Facial Similarity, Facial Inferences, Faces, Gender differences.

Table of Contents

Introduction

The present dissertation aims to deepen the scientific understanding regarding the role played by the facial similarity between an elector and a candidate in the inferences made by an independent observer. The present study also seeks to bridge a gap in the current literature. While existing research highlights the correlation between the facial resemblance of an elector and a candidate and increased intention to vote for the latter (Bailenson et al., 2006; 2008), as well as the ability of individuals to discern the political orientation of both politicians and voters at a rate notably above chance (Rule & Ambady, 2009; 2010), it fails to address whether the criterion of facial similarity between candidates and electors is utilized in making inferences about others' voting intentions.

The first two chapters will have a theoretical focus, reviewing the existing literature on the topic. Chapter I will address the basic mechanisms of face perception, with particular attention to the neural correlates of such processes and the proposed evolutionary theories that try to explain them. It will also introduce the topic of trait and similarity inferences from faces. Chapter II will focus on these two topics, analyzing them in the more specific context of political psychology, reviewing the main findings on the role played by facial features in the creation of first impressions of political candidates, and explaining the role played by similarity in political inferences. The role of gender and gender prototypicality, two factors closely linked with trait and similarity inferences, will also be introduced and discussed in the chapter.

Chapter III will outline the main study's experimental design, emphasizing the primary experimental hypotheses and detailing the materials utilized. In particular, Chapter III will also describe the pre-testing process for the artificially generated facial stimuli used in the main study.

Chapter IV will focus on analyzing the data collected with the main questionnaire and proposing a series of models to better understand the relationship between the variables and the effects of the experimental manipulation.

Lastly, Chapter V will discuss and propose an interpretation of the findings, explore the study’s limits, and propose possible directions for further studies.

Chapter I – An introduction to face perception

1.1 The specificities of face perception

Faces are possibly the most important stimuli humans perceive throughout their lives. From faces, humans get to automatically infer a variety of important information, such as the identity of a person, their gender, and their belonging to an ethnic ingroup or outgroup. Looking at the direction of the gaze, other humans can understand where the attention of their conspecifics is allocated (Langton et al., 2000). Moreover, thanks to the over thirty facial muscles innervated by the facial and the trigeminal nerves (Westbrook et al., 2018), humans can express various complex feelings and emotions through their faces. These facial expressions of emotions, as well as the ability to recognize them, are regarded as at least partially innate and universal by an extensive line of research (Ekman, 1982; 1994; 2006) stemming from the work of Darwin himself (Hess & Thibault, 2009; Barrett, 2011). Faces are so crucial for our species that we tend to wrongfully perceive them in non-facial stimuli, giving rise to a phenomenon known as face pareidolia (Zhou & Meng, 2020). Although the origin of face pareidolia is still debated, the scientific literature on this topic presents strong indications that it might be at least partially innate. The existence of gender differences in pareidolia, with women displaying a higher tendency to see faces in neutral stimuli (Zhou & Meng, 2020), as well as the fact that other primates seem to exhibit analogous phenomena with Rhesus Monkey (Taubert et al., 2017) and Chimpanzees (Tomonaga & Kawakami, 2023) showing higher visual preference towards objects arranged to look like a face, seems to point out to a phylogenetic origin of such phenomenon. According to the evolutionary interpretation of this phenomenon, pareidolia may have emerged through natural selection because recognizing face-like patterns could have been crucial for detecting dangers in the surrounding environment (e.g., an enemy hiding in the shadows). Such an ability would have been so important and effective in guaranteeing survival that natural selection favored people with higher reactivity to face-like patterns. Thus, responsiveness to these patterns was selected, even at the cost of giving rise to false positives, as happens in face pareidolia (Zhou & Meng, 2020).

Human beings exhibit an innate preference towards faces. Even newborn babies (Umiltà et al., 1996) show more extended fixation patterns towards stimuli that present face-like characteristics, and they seem to prefer those kinds of stimuli even when they are presented simultaneously with other objects having optimal sensory properties for the newborn visual system. Although this finding appears to support the innate nature of the

importance given to faces, it could be argued that even newborns must have been exposed to faces, first and foremost to the ones of their primary caregivers. However, comparative studies conducted in the field of animal psychology point to the preference for faces being innate for primates. For instance, Sugita (2009) tested newborn Rhesus monkeys who had never been exposed to facial stimuli and found that they displayed a strong fixation preference toward faces.

Humans also have an astounding ability to recognize faces. Jenkins et al. (2018) estimate that the average person should recognize about 5,000 different faces; due to the high individual variability found in the study, the authors evaluate that some people may even be able to recognize up to 10,000 faces. The authors noted that if these estimates are reliable, modern humans can identify a much larger number of faces than the number of individuals that ancient humans were exposed to in a lifetime in their evolutionary context.

Developmental research shows that babies seem to be able to recognize individual faces, relying on a variety of cues, from configural properties, such as the distance between the eyes, to specific features, such as the shape of the various parts of the face (Simion et al., 2007). This ability emerges exceptionally early and is already detectable at 1-3 days old (Simion et al., 2007). Research supports the idea that the developmental trajectory of face recognition practically starts at birth and that, according to some researchers, it reaches maturation at an extremely young age of 3-5 years (McKone et al., 2016; Fuhrmann et al., 2016). Other lines of research point to the idea that the ability to remember faces continues to develop until the end of adolescence (Fuhrmann et al., 2016). Face cognition then diminishes during adulthood (Connolly et al., 2021). Such a decline is positively similarly correlated with age to the more general cognitive decline associated with aging (Miles & Miles, 1961). Conversely, Connolly et al. (2021) found that age-related changes in the ability to recognize faces and facial expressions were present even when controlling the changes in general intelligence. However, the scientific community has not reached a consensus on the topic: a recent meta-analysis (Walker et al., 2023) found a moderate and significant association between general intelligence and face-processing abilities, thus showing that the relationship between these variables is still a matter of discussion for the expert in this field.

Further developmental research supports the idea of an extremely early emergence of the mechanisms underlying face cognition. Turati & Sangrigoli (2004) highlight that since the age of four months, babies start to show the well-known inversion effect (Valentine,

1988; Farah et al., 1995). Likewise, Turati et al. (2004) report that when four-month-old infants are habituated to a face, subsequently presented in various poses, they later seem to recognize the target face with higher accuracy when the stimulus is presented upright than when it is inverted. This phenomenon's presence in newborns confirms that the specificity of face perception is innate rather than the result of social learning.

The face inversion effect is a widely reported and extensively studied phenomenon, consisting of the fact that people seem to be far quicker and more accurate at recognizing faces when they are presented upright rather than inverted. Although most objects are more challenging to recognize when presented in unusual positions, such as upside-down, recognizing faces seems to be particularly impaired when presented with an inverted orientation (Valentine, 1988; Farah et al., 1995). This effect confirms the importance and the peculiarity of face perception in the human species, as well as the existence of unique perceptive mechanisms that underlie this phenomenon and distinguish faces from other kinds of visual objects. Specifically, what differentiates face perception from the codification of different visual stimuli seems to be the holistic nature of this phenomenon (Farah et al., 1998). Instead of being based solely on the recognition of the various components of a face, otherwise known as “first-order features” (Rhodes, 2013), such as the shape of the lips, the nose, or the eyes, face perception seems to mainly rely on the recognition of configurational patterns, which the author called “second-order features.” These include spatial relationships between the various first-order features of a face, as well as information on the general shape of the face (Bruce & Young, 1986). When operating tasks involving facial stimuli, such as face recognition, people seem to use second-order rather than first-order features more frequently. In other words, people tend to perceive a face as an undifferentiated visual gestalt rather than focus on its various first-order features (Homa et al., 1976). The holistic nature of face processing is also supported by studies on the social inferences that people can make from faces and individual face characteristics. In this context, Santos & Young (2011) found that, although the various first-order characteristics can be informative of social judgments, their informative value seems to decrease when presented alone compared to when they are perceived in the context of a whole face.

1.2 The neural bases of face cognition

Since their birth at the end of the nineteenth century (Wickens, 2014), the fields of neuroscience and neuropsychology have dedicated themselves to elaborating a functional

map of the brain, connecting specific cortical and sub-cortical regions to various perceptive, motor, cognitive, and emotive processes. The first highly influential model that tried to explain the complexity underlying face recognition was proposed by Bruce and Young (1986). In accordance with the evidence presented by the existing literature about patients with unilateral cerebral lesions, these authors proposed that the cognitive processes of recognizing the identity of a face and identifying a specific facial expression had to be mutually independent. Even more recent studies (Fisher et al., 2017; Bell et al., 2023) seem to support this claim: for example, in developmental prosopagnosia, a congenital and possibly genetic condition that impairs explicitly face cognition, there appears to be a mutual dissociation between the ability to recognize facial expression and the ability to recognize individual faces. The model proposed by Bruce and Young (1986) drew inspiration from pre-existing models for word recognition (Nelson et al., 1977) as well as for object recognition (Warren & Morton, 1982). Such a model described face cognition as the result of two parallel, hierarchical processes, one of which is deputed to analyze the position of the facial muscles and draw conclusions about facial expressions. In contrast, the other is deputed to processing information about the physiognomy of perceived faces and assigning an identity to such faces. Bruce and Young, however, did not advance a proposal on the neural localization of either of those processes (Atkinson & Adophs, 2011). Only in the following decade, thanks to the advancement in the field of neuroimaging (Uğurbil, 2012), the areas thought to be responsible for these processes were identified. Thus, successive theories, such as the one proposed by Haxby et al. (2000), proposed various neural structures as bases for the numerous processes involved in face cognition while maintaining the double pathway structure proposed by Bruce and Young (1986).

The double dissociation between face recognition and the recognition of facial expressions has found broad support in the scientific literature. The two processes involve different neural pathways. For instance, Puce et al. (1998) determined that when people watched the movement of the eyes and the lips of an otherwise motionless face, functional Magnetic Resonance Imaging (fMRI) results showed increased neural activity in a bilateral region centered in the posterior superior temporal sulcus (STS), a finding compatible with the existence of a specific pathway that focuses on the perception and interpretation of facial muscles' movements. On the other hand, Kanwisher et al. (1997) used the fMRI technique to locate an area of the brain that seems to be selectively activated during face recognition. This bilateral region of the extrastriate cortex, located

in the inferior temporal gyrus, has been denominated Face Fusiform Area (FFA). This cerebral area shows similarly strong responses to human faces, as well as to cartoon faces and animal faces (Tong et al., 2000), thus supporting the hypothesis that the human ability to recognize faces is so developed that it responds to various face-like stimuli rather than to human faces alone. Moreover, the FFA seems to respond to whole faces and partially occluded faces alike, while its response to a pair of eyes decontextualized from a face seems far weaker (Tong et al., 2000). Those findings seem to point to the idea that the FFA elaborates face-like patterns as gestalts rather than elaborating the single first-order characteristics of the face one by one. Since the initial studies on the FFA suggested that it could be the cerebral correlate of the proposed face perception cognitive module (Kanwisher et al., 1997), discovering this cerebral area seemed to corroborate Bruce and Young's theory further. Nonetheless, further research on the FFA disproved its specialization for faces. In a very influential study, Gaunthier et al. (2000) recruited people with expertise in recognizing bird species and car models. They asked them to recognize individual bird species and car models from photographs while measuring their brain activity through fMRI. The authors reported that the brain areas previously thought to be specialized for face perception were activated during this task. Furthermore, the intensity of the activation was positively predicted by the level of expertise of the experimental subject within the category, as assessed in an independent behavioral test. Later studies confirmed that the FFA is recruited during other recognition tasks performed by experts in a field. For instance, Bilalić et al. (2011) reported that the FFA was more intensely activated when chess masters looked at common chess positions than novices. These results suggest that the FFA is not exclusively implied in face recognition; on the contrary, the current evidence supports the idea that this cortical area is at least partially responsible for recognizing individual stimuli that are members of a superordinate category with which the individual is familiar. This evidence supports the idea that face recognition is not qualitatively different from the recognition of other objects, with the main difference being the frequency with which people assign a name (and thus, an identity) to individual faces rather than to sub-categories of other stimuli, thus weakening the original theories of face cognition. Furthermore, while Bruce and Young's theory predicted face perception to be a hierarchical, exclusively bottom-up process, recent neuroscientific evidence highlights the critical role of top-down processes in face perception, as already proposed by Haxley et al. (2000). A recent Italian study (Palmisano et al., 2023) examined the effects of a 40Hz transcranial alternating current stimulation

(tACS) over structures associated with the broader face perception network (i.e., the right lateral occipitotemporal cortex and the left pre-frontal cortex) on a sample of 75 healthy volunteers. Results showed that, when the subjects received the tACS, they were more likely to perceive faces in irrelevant stimuli, and thus, they were more likely to be subject to the phenomenon of face pareidolia. The authors interpreted these data as evidence of a top-down process, descending from the stimulated areas down to the primary face-sensitive areas in the visual system, thus supporting the hypothesis that perceptual illusions in healthy subjects may stem from expectations and decisional processes as well as from the intrinsic perceptual properties of the perceived object.

Despite some evidence contradicting earlier theories on face cognition, more recent research seems to reaffirm the peculiar nature of the processes underlying such a phenomenon. Elbich and Scherf (2017) used the Magnetic Resonance (MR) technique to study the activation of various brain areas when performing a process, either face or object recognition. A selective activation was present for numerous cerebral regions, including, but not limited to, the FFA, some specific bilateral cortical areas in the anterior temporal lobe, the fusiform gyri, and the sub-cortical nuclear complex of the amygdala. Moreover, the authors found that processing faces, but not objects, correlated with activating a global face network distributed in numerous secondary areas across the brain. These findings reveal a much more complex relationship between cognition and neural structures than previously hypothesized: far from being localized in a singular cerebral area, processes such as face cognition involve various degrees of activation in several different brain regions. The most recent neuroscientific evidence, however, supports the persisting importance of the FFA in the field of face cognition. Tsantani et al. (2021), for instance, used the fMRI technique to try to identify the difference between the face-related representations encoded in the three main face areas of the brain: the FFA, the superior temporal sulcus (pSTS) and the occipital face area (OFA), a cortical region on the lateral surface of the occipital lobe that previous studies (e.g., Rossion et al., 2012) found to be implicated in extracting information from facial stimuli. The evidence gathered by Tsantani et al. (2021) suggests that, while both the FFA and the OFA are implied in face cognition and both can discriminate the identity of a face, the information represented in the two areas seems to be markedly different. While the cognitive representations encoded in the OFA seem to be driven by low-level differences based on image properties, the representations held in the FFA are much more complex and abstract social information. The information encoded by the FFA seems to include perceived similarity to previously

known faces, the perceived gender of the person depicted, as well as many other social judgments regarding the trustworthiness, dominance, attractiveness, and valence of a face.

In conclusion, despite the numerous neuroscientific advancements since the early theories about face cognition, the FFA seems to maintain a primary role in the current conceptualizations of the processes of face cognition. Not only that, but the area also seems to be specifically responsible for social and similarity judgments, which is crucial for the current dissertation and will be the subject of the following paragraphs.

1.3 Judging similarity between faces

As presented in the previous paragraph, the FFA seems to be responsible for codifying similarity judgments between faces (Tsantani et al., 2021). Moreover, the neural responses of the FFA have been found to be reflective of the perceptual differences between the individual faces perceived. Davidesco et al. (2013) measured neural activation in the FFA in correspondence with the visual presentation of a set of faces by using electrocorticographic (ECoG) recordings, an intracranial electroencephalographic technique much more invasive than the fMRI, requiring the placement of electrodes directly on the cortical surface. Despite the limited cases in which it can be ethically used on human patients, this technique presents a far more accurate spatial resolution than most other techniques, including the fMRI. The use of ECoG was possible because the sample consisted exclusively of patients who have intractable epilepsy, whose brain activity had to be monitored using ECoG due to clinical reasons. Thus, the sample size was small, including only 14 individuals, which did not represent the general population. Despite its limitations, the study found an interesting and significant pattern in the responses of the FFA to a set of various faces: the neural activity in the FFA was positively and significantly correlated ($r=0.52$, $p<0.0005$) with the perceptual distance between the stimuli, as independently assessed by another sample of healthy individuals. In other words, the neural response patterns in the FFA were more similar when the photographs portrayed similar individuals. This evidence suggests that information about similarity is encoded and processed in the FFA, although it cannot prove that this specific brain area is directly responsible for similarity judgments.

Scientific literature in evolutionary biology has broadly discussed the possible adaptive reasons underlying the development of mechanisms of similarity recognition (DeBruine et al., 2008). Indeed, during the late nineteenth century up until 1964, one of the most

critical limits (Dugatkin, 2007) of Darwin's theory of evolution by natural selection (Darwin, 1859) was constituted by the existence of altruism and the presence of selfless acts, in which both human and non-human animals show to be willing to penalize themselves by favoring others, seemingly without return, thus reducing their evolutionary fitness. The problem of altruism was later addressed by Hamilton (1964), who, in a vastly influential paper, differentiated evolutionary fitness into two different metrics of reproductive success: personal and inclusive fitness. In short, Hamilton defined inclusive fitness as the number of offspring equivalents that an individual helps or supports throughout their life. Hamilton's inclusive fitness theory explains the existence of altruism by linking it with the fact that an individual could increase the diffusion of their genes by supporting other individuals with whom they share identical alleles. Thus, the theory of inclusive fitness, still nowadays considered a milestone in the field of evolutionary biology (Marshall, 2016), predicts that highly social species will evolve the ability to recognize their kin and that an individual will change their behavior towards another depending on the perceived degree of kinship (Agrawal, 2001).

Due to the tendency of humans to rely on the sense of sight, faces have been proposed by evolutionary biologists and psychologists as possible visual cues that would help people determine their degree of relatedness to one another (Mateo, 2015; DeBruine et al., 2008). The idea that kinship can be assessed through similarity, besides being particularly intuitive and culturally widespread, is also known as phenotype matching. Evidence for this process has been found in non-human primates, such as the Chimpanzee (Parr & de Waal, 1999) or the Rhesus Monkey (Parr et al., 2010). However, such animals could use phenotype matching only in mother-son and not in mother-daughter pairs of photographs. The scientific literature on the topic seems to support the ideas that a face can convey information about relatedness and that judgments of similarity and kinship are somehow similar, both in the case of children's (Dal Martello & Maloney, 2006) and adults' (DeBruine et al., 2009) faces. By asking their participants to judge whether the children portrayed in a pair of photographs were siblings or not, Dal Martello & Maloney (2006) discovered that, at least when it comes to children's faces, the upper half of the face holds a higher informative value when compared to the lower half. In the experimental condition in which the lower half of the face was masked, the participants' ability to determine kinship was not significantly impaired; on the contrary, when the upper half was masked, the participants' performance worsened significantly. The authors argue that the non-informative nature of the lower half of the face could be due to this part of the

face being more developmentally unstable, changing widely its proportions before adulthood. Moreover, the authors highlighted the importance of the eyes: according to their data, the eyes alone accounted for about half of the upper face's informativeness. On the other hand, selectively masking the mouth did not significantly change the participants' performance; thus, the mouth seems to be non-informative in the field of kinship judgments. Lastly, the authors reported that the similarity ratings of each pair of photos predicted the probability that they were considered siblings and that gender homogeneity was predictive of neither the similarity nor the kinship rating, suggesting that gender might be a confounding variable that people rightfully ignore when assessing kinship. These findings were only partially replicated by DeBruine et al. (2009), who instead used portrayals of adults who were, depending on the condition, siblings or not. In this case, kinship and similarity judgments were consistent only when the pairs of faces were matched for gender and age. Gender and age inhomogeneity predicted lower similarity ratings but not lower kinship ratings, thus suggesting that the information conveyed by these two kinds of judgments is not entirely the same. The studies of Dal Martello & Maloney (2006) and DeBruine et al. (2009) were recently replicated by Hansen et al. (2020), who extended the confounding effect of gender in similarity judgments from adults' faces only to children's and adults' faces alike.

Being an informative cue of kinship, facial resemblance can be crucial in some evolutionarily relevant decisions, such as showing pro-social behavior toward an individual or choosing a romantic or sexual partner. Overall, the psychological literature on the topic supports these hypotheses. Evidence suggests that, at least in an experimental context, participants tend to express more favorable judgments toward individuals with a higher degree of facial similarity to the observer (DeBruine et al., 2008). For instance, higher levels of self-resemblance, artificially obtained through facial morphing techniques, were associated with higher degrees of generosity and cooperation in the context of a group-based economic game (Krupp et al., 2008). Recent research (Nakano & Yamamoto, 2022) showed that people tend to infer a higher degree of trustworthiness from the faces of people of their gender with whom they had a higher degree of facial resemblance. Similarly, Krupp et al. (2012) replicated these results using faces of the opposite gender to their participants. The authors manipulated the degree of self-similarity of the photographs viewed by their participants, both positively (i.e., the resemblance was increased) and negatively (i.e., the resemblance was decreased). The authors' results showed that people were more likely to form a positive impression of

people they resembled, judging them as more trustworthy and attractive. Moreover, an opposing effect was found for negative resemblance, with participants showing decreased appreciation for people whose faces were manipulated to differ from the participants'. The authors argued that these data could be interpreted as evidence of humans' ability to perceive both positive and negative similarities to themselves by looking at other people's facial phenotypes. By being able to do so, humans could actively identify their kin and allocate resources to help them; on the contrary, they would be able to enact spiteful behavior toward people with whom they share less genetic heritage. In doing so, a person would favor people who bear a more significant share of their alleles, thus increasing their inclusive fitness, while disfavoring people who are more likely to bear rival alleles, thus reducing competition. As Krupp et al. (2012) identified, positive judgments toward self-resembling people also include attractiveness judgments. DeBruine (2004) found that the positive effect of facial similarity on attractiveness was stronger for faces of the participant's gender than for faces of the opposite gender. This effect could be linked to the fact that mating between too closely related individuals could be deleterious, and thus, the ability to recognize and perceive as less attractive related people of the opposite gender could have evolved to avoid inbreeding. Other evidence, however, shows that similarity is essential in sexual and romantic attraction. For instance, Tea-Makorn & Kosinski (2020) examined the facial resemblance of 517 heterosexual couples at the beginning of their marriage and 20-69 years later, finding that married people tended to look similar but that their facial similarity did not increase with time, as the convergence hypothesis (Zajonc et al., 1987) postulated. This evidence seems to corroborate Bateson's biological theory of optimal outbreeding (Bateson, 1980; 1983), predicting that, since an excessive level of genetic similarity might result in the dangerous diffusion of recessive genetic diseases, while outbreeding could lead to the loss of genes specifically adapted for a particular environment, individuals would choose a mate with an optimal degree of genetic similarity.

1.4 Inferences from faces

Besides representing a possible neural correlate of similarity judgments between faces (Tsantani et al., 2021), evidence suggests that the FFA might also be responsible for the ability to infer many other relevant biological and psychological characteristics of an individual from their face. Such impressions from faces seem to arise extremely rapidly: Bar et al. (2006) measured that consistent impressions regarding certain personality traits

of the observed person can emerge within 39ms from the visual presentation of a face. In particular, the authors asked to judge faces on perceived threat and intelligence traits: whereas the former judgments reached consistency within 39ms, the latter did not. These findings might indicate that the facial information regarding threats is elaborated preferentially in the first milliseconds after the presentation, most likely because cues of threats are crucial for survival and thus are more salient. In the same year as the study from Bar and colleagues, Millis and Todorov (2006) studied the presentation time that people required to reliably draw inferences from a face stimulus. The authors asked their participants (117 undergraduate university students) to judge a set of faces on a specific psychological trait for each of the five between-subjects conditions: attractiveness, liking, competence, trustworthiness, and aggressiveness. In all five conditions, the visual presentation of the stimulus managed to be nearly fully informative even with a presentation time of only 100ms, corroborating the hypothesis that impressions from faces are particularly fast to form. Later evidence reviewed by Todorov et al. (2015) confirmed that social inferences from faces are incredibly rapid and that there is a widespread agreement between individuals on the content of such inferences, suggesting that some facial features might hold information that external observers can interpret following specific common rules. Despite the remarkable agreement between subjects, the impressions that people draw from faces are not remarkably accurate to the personality of the observed person. In a study by Oliviola & Todorov (2010), the impressions of a vast sample of 1018 subjects were found to be highly unreliable at predicting the characteristics of a person; in fact, the information drawn from faces was detrimental, and the participants would have been more accurate if they completely ignored it, simply guessing the characteristics from their base-rate frequencies among the general population.

While early evidence noticed that many different personality traits of an individual could be inferred from their face, Oosterhof & Todorov (2008) demonstrated through principal component analysis that these traits essentially were explained by two orthogonal dimensions: valence and dominance. A judgment of threat, for instance, could emerge from the perception of an individual's face having both a negative valence, thus being perceived as willing to harm the observer, and a positive dominance, thus being considered as effectively capable of inflicting such harm. In the same paper, the authors showed that these dimensions were well approximated by asking people to evaluate the trustworthiness and dominance of an individual by looking at their face. Sutherland et al.

(2013) replicated Oosterhof and Todorov's analysis, confirming the relevance of the dimensions of trustworthiness and dominance for social judgments. Nonetheless, a third dimension, denominated by the authors as "youthful/attractiveness," also emerged. This factor explicated a series of inferences that the previous two dimensions could not fully account for; it was negatively correlated with perceived age and positively correlated with perceived attractiveness, health, and facial immaturity.

The evidence suggests that, despite the common saying, people are remarkably quick at judging "a book by its cover" when given the chance. That given, many other studies investigated what specific facial cues people use to formulate such quick and consistent judgments. Santos and Young (2011) conducted a series of experiments to determine the informativeness of various facial regions and features. In the first experiment, the authors presented a series of 224 faces to their participants, that is, 32 faces for each of the seven judged characteristics (i.e., sex, age, distinctiveness, attractiveness, approachability, trustworthiness, intelligence). Each participant saw a third of the faces as complete photographs, whereas a third was presented with the internal features removed and a third with the external features (e.g., the hair, the neck, etc.) removed. The authors measured the percentage of agreement between the whole-face condition and the two others, finding that agreement between raters generally decreased when participants were asked to make inferences from the external features of a face only. On the contrary, for most of the characteristics that the participants were required to judge, viewing only the internal features of a face did not significantly impair the agreement rate of the participants. Having determined that the internal region of a face must be considered the most informative, the authors conducted a second experiment in which they compared the informativeness of four main regions of the inner face: the upper head, which was used as a control, as the authors hypothesized that the region could not be particularly informative, the eye, the nose, and the mouth region. The mouth region was determined to be the most influential on the judgment of approachability, with a decrease in the between-subjects percentage of agreement when the region was masked. Similarly, the eye region was the most influential upon judgments of attractiveness. However, none of the other five trait judgments were significantly impaired by masking a particular region; moreover, the percentages of agreement remained above the chance level in every condition in which a region of the face was masked. The authors interpreted the finding as evidence of the holistic nature of face perception. Despite the single-face regions being somewhat informative when decontextualized, they acquire their full significance only

when presented in the context of a whole face. Oosterhof & Todorov (2008) also contributed to the line of research that was interested in determining which features underlie impressions of valence and dominance from faces. The authors randomly generated a series of faces with neutral expressions and then exaggerated their features. The participants of the study were asked to judge the faces on the dimensions to determine which emotion the face depicted, being forced to choose between the tag "neutral" and the six basic emotions of anger, happiness, sadness, disgust, fear, and surprise. The faces were later evaluated on the dimensions of trustworthiness and dominance. The results indicated that the resemblance of a neutral face to an emotional expression was responsible for varying judgments in trustworthiness: for instance, resemblance to happiness was associated with higher judgments of trustworthiness; on the contrary, resemblance to anger was associated with lower judgments of trustworthiness. Moreover, the resemblance of a face with emotional expressions also showed weaker associations with judgments of dominance: fearful-looking faces were judged as submissive. In contrast, neutral and angry-looking faces were more likely to be categorized as dominant. Nonetheless, dominance judgments were better explained by facial maturity and facial masculinity. When the authors manipulated faces to look more mature and masculine, for instance, decreasing the distance between the eyes and the eyebrows, faces were perceived as belonging to more dominant individuals. The findings of Adams Jr. et al. (2012) also contributed to corroborating the hypothesis that both masculinity/femininity and resemblance to emotional expressions are powerful explanatory factors for inferences drawn from faces. The authors conducted a pilot study comparing the inferences from masculine and feminine-looking faces. Their findings indicated that faces perceived as belonging to a woman were more likely to be judged as submissive, affiliative, naïve, honest, cooperative, babyish, fearful, and happy than faces perceived as belonging to a man. Moreover, when neutral faces were manipulated to resemble emotional expressions of anger and fear more closely, the judgments varied similarly: fearful-looking faces were judged as significantly more babyish, fearful, sad, subordinate, affiliative, naïve, honest, cooperative, and intuitive than angry-looking and neutral faces. This pattern of results, the authors argue, is clearly stereotype-driven. More recent studies consistently confirmed the importance of the resemblance to emotional expressions: for instance, Jaeger & Jones (2022) found that this dimension was the most predictive of the impressions drawn from faces, both for trustworthiness and dominance judgments. As evidenced by the vast literature in social psychology, social cognition is based on judgments on two dimensions:

warmth and competence (Fiske et al., 2007), which remarkably resemble the dimensions of valence and dominance individuated by Oosterhof & Todorov (2008) and are crucial in explaining both interpersonal and intergroup social cognition. These two dimensions seem inextricably linked to masculinity/femininity, with femininity being associated with warmth and masculinity with competence (Abele, 2003; Ebert et al., 2014; Fiske, 2010). In recent years, the scientific community seems to have reached a consensus on the overgeneralization hypothesis (Zebrowitz, 2017). According to this hypothesis, the surprising percentage of agreement in the impressions that people draw from faces is essentially due to the overgeneralization of adaptive impressions of various social categories that share some unique physiognomic traits (idem). For instance, people showing immature traits (i.e., a higher degree of baby-facedness), such as larger eyes, higher eyebrows, smaller nose bridges, rounder and less angular faces, thicker lips, and lower vertical placement of features (Zebrowitz et al., 2008), would more closely resemble babies, and thus they would elicit impressions coherent with the stereotype traditionally associated with babies, such as higher levels of honesty, happiness, naivety, etc.. As previously examined, gender and, thus, gender stereotypes constitute another important dimension underlying the process of overgeneralization in the context of face cognition. Recent research (Oh et al., 2020) displayed that observers who more strongly endorsed gender stereotypes were also more likely to judge women as more trustworthy and men as more competent. Similarly, Xie et al. (2021) examined the role of racial and gender stereotypes in the context of the formation of impressions from other people's faces. In a large sample of 4,247 adults from the United States, the authors found that facial impressions were partially dependent on the target's gender and race. Moreover, the impressions' content was associated with the structure of the stereotypes held by the observers, as measured by having them subsequently asked to answer on a Likert scale to a series of questions such as "Please indicate how people in society see [ethnic/gender group] on [psychological dimension]". As explained by Hehman et al. (2019), face impressions are now being studied as phenomena emerging from the complex interaction between the characteristics of the target and a perceiver: the observer's beliefs, their background, and the stereotypes they endorse may play a role as important as the morphological characteristics of the perceived face itself (Xie et al., 2021; Todorov & Oh, 2021).

Chapter II – Inferences from Faces in Politics

2.1 The role of perceived competence and warmth

Due to the increasing personalization of politics (e.g., Caprara, 2007), further amplified by the diffusion of social media (e.g., Metz et al., 2020), elections are nowadays perceived as being more of a competition between candidates rather than between parties or ideologies. Consequently, politicians' perceived personalities are playing an increasingly important role in guiding the electors' choices. Scientific evidence in political psychology corroborates the central role played by first impressions in electoral choices: the impressions that an elector produces online the first time they are presented with a candidate seem to predict better the candidate evaluation than the mix of pro and con information retrieved from memory (Lodge et al., 1989). In the last decades, a significant line of research in political psychology has examined how people form impressions of politicians and the effect that such judgments have on vote intentions. Caprara et al. (2002) discovered that the electors' description of a politician's personality significantly differed from how people usually describe themselves and from how they describe other public figures (Caprara et al., 1997). Whereas other categories of people are usually described using all the five dimensions of the so-called Big Five (i.e., extroversion, openness to experience, agreeableness, conscientiousness, emotional stability; De Raab, 1998; 2000), personality judgments regarding politicians, both Italian and American, seem to cluster around two fundamental dimensions. Such dimensions, which the authors called Energy and Agreeableness, seem to be the driving factors for judging the politicians' personalities, both during electoral campaigns and in the following years. It is evident how these two dimensions closely resemble judgments of warmth and competence (Fiske et al., 2002): similarly to competence, Energy is linked with agency and power; on the other hand, Agreeableness is linked with emotional stability, love, and communion, similarly to warmth.

As examined in the previous chapter, research on face cognition has found evidence that face impressions are organized along two orthogonal dimensions: valence and dominance (Oosterhof & Todorov, 2008). These two dimensions also remarkably resemble the ones proposed by Fiske et al. (2002) in their model of stereotype content: competence and warmth. In their original paper, Fiske et al. asked their participants (a total of 272 undergraduate students across the four studies) to evaluate a series of outgroups, varying in gender, ethnicity, age, class, and disability. To prevent the effects of social desirability,

the participants were asked to evaluate the groups as society views them, instead of being asked to give their personal opinions on the various outgroups. The results showed that stereotype content could be classified into four different quadrants, depending on the two orthogonal dimensions of warmth and competence. When a group was perceived on the lower side both in competence and warmth (e.g., homeless people), it mainly elicited negative emotions, such as contempt. Secondly, when a group was perceived as high in warmth, but low in competence (e.g., elderly people, housewives), it mainly elicited paternalistic prejudice, associated with feelings of pity and sympathy. Thirdly, when a group was perceived as cold, but highly competent (e.g., Asians, Jews, feminists), it elicited envious prejudice, associated with feelings of envy and jealousy. Lastly, the perception of high competence and high warmth was reserved for the ingroup and its close social allies and was associated with feelings of either pride or admiration. Moreover, the authors found that the perception of competence was predicted by the social status of the target social group, with higher levels of status being associated with higher competence. Additionally, estimates of warmth were negatively associated with the perceived competitive pressure exerted by the outgroup toward the ingroup. For instance, housewives were perceived as warm and incompetent because they possessed a low social status, but they posed no threat to the dominant group. On the contrary, feminists were perceived to be a group with high competence, since they struggle for a higher social status, but low in warmth, since they were perceived as a threat to the social *status quo*. Despite theoretically being two separate, orthogonal dimensions, warmth and competence are not always independent from one another. The extensive literature on the topic, both experimental and correlational (Judd et al., 2005; Kervyn et al., 2010), seems to suggest that, when two targets are compared on the two fundamental dimensions of social perception (i.e., competence and warmth), a compensation effect arises. The compensatory effect, however, does not emerge in every situation. For instance, in the context of a large multi-phase study on inferences from faces, Schmitz et al. (2024) did not find any compensatory effect between competence and warmth specifically; on the contrary, compensatory effects were found for other dimensions of social perceptions, such as trustworthiness, warmth, and dominance. Similarly, Castelli et al. (2009) did not fully replicate the effect when asking Italian electors to evaluate pairs of candidates at regional elections and argued that the compensatory effect may only emerge when perceivers compare two social targets which they are also motivated to see as somehow similarly good. Moreover, the authors found that people report that both facial

competence and sociability are crucial qualities for winning an election. However, actual vote results were only positively predicted by judgments of competence, whereas perceived warmth was negatively correlated to the candidate's electoral success.

In conclusion, the reviewed evidence suggests that compensatory effects between competence and warmth may not be common in political judgments, where partisan electors are positively biased towards only one of many candidates, and where participants are often asked to replicate in an experimental setting a similar choice. Despite the complex relationship between the two fundamental dimensions of social judgment, the relationship between both dimensions has been investigated in the scientific literature, and both have been found to be consistently predictive of electoral outcomes. For instance, Bennett et al. (2019), found that warmth and competence were predictive of vote intention, both when people were asked to judge a politician's personality (by using the stereotype content model; SCM) and when people were asked to judge a politician's brand (by using the Brands as Intentional Agents Framework; BIAF), albeit identification with a political party was found to be a more powerful predictor. In the field of inferences from faces, Todorov et al. (2005) similarly found that judgments of competence emerging from an extremely brief (1-second) exposition of a politician's face were strong predictors of the outcome of the 2002 and 2004 United States Senate and House of Representatives electoral races. When the participants were asked to indicate which one of two concurring candidates was the most competent, the results positively correlated to the real electoral outcomes, to the point that such judgments also linearly related to the margin of victory of one candidate on the other. Atkinson et al. (2009) replicated the effect of face-inferred competence. They also estimated the possible influence on the electoral results: the probability of voting for a competent-looking candidate increased by 1-3 percentage points for partisan electors, who strongly identified with their political party, and by almost 4 points for independent electors. Later studies, such as the one by Sussman et al. (2013) continued this line of research. The authors asked a sample of 223 American residents to evaluate the 18 candidates in the 2011 Bulgarian presidential election on a series of personality traits by only looking at the candidates' faces. The assessed personality traits were selected by asking a sample of Bulgarian electors what the most important traits were for a politician to have to be a good president; the list was then restricted by asking another sample of Bulgarian electors to rate the importance of the previously collected list of traits. In the end, five traits were selected to undergo assessment by the American sample: competence, dominance, honesty/incorruptibility,

and likability. The American participants also evaluated the candidates' attractiveness, a variable considered particularly relevant due to its immediateness. The study's design allowed the authors to expand on the previous finding by Todorov et al. (2005) because they tested the predictive value of competence in evaluating a series of individual faces instead of comparing two opposing candidates. The results replicated the original finding: competence, and none of the other traits, predicted the electoral results. This finding is particularly relevant because it generalizes the importance of competence even in a multipolar democratic system, which is more similar to the one currently present in Italy. In recent years, the growing scientific evidence supporting the central role played by facial traits inferred competence has raised concerns for the well-being of democratic systems (Casey, 2022). Although it is undoubtedly worrisome that the appearance of a candidate is consistently used as a heuristic to make a choice as important as the election of a nation's president, the scientific literature offers some reassuring evidence that inferences from faces might not only be consensual between raters but also be, to a certain extent, accurate at an above-chance level (e.g., Rule et al., 2009; Rule & Ambady, 2010). For instance, in the field of political psychology, Casey (2022) presented evidence supporting the hypothesis that competence inferences from candidates' faces may be diagnostic not only of their electoral performance but also of their performance as efficient members of the parliament. The author used photographic portraits of elected candidates from Sierra Leone and recruited the participants in rural communities in the same country. The participants, a total of 399 Sierra Leonean electors, declared to be unfamiliar with the candidates whose faces were used in the study, who all belonged to other constituencies. The experimental procedure mimicked the one proposed by Todorov et al. (2005): the participants were shown pairs of faces and were asked to select the most competent-looking one. A politician's performance was subdivided into three main constructs: lack of corruption, constituency service, and ability to persuade. Each construct was operationalized using a series of publicly available measures. For instance, lack of corruption was measured through the share of funds given to the candidate that was actually and verifiably spent on the constituency's development. Overall, the collected data supported the hypothesis that face inferences of competence predicted the candidate's performance: the candidates whose faces were selected for being more competent showed a higher ability to persuade (as measured by an independent sample who judged their performance in a speech) and lower levels of corruption. No association was found between the facial inferences and constituency service.

Casey's (2022) study also provides evidence for the cross-cultural importance of competence in predicting the outcomes of political races. The literature on this topic, however, has not reached a consensus: although some studies (e.g., Chen et al., 2012; Chen et al., 2016) show that facial inference might be less important in Asian democratic cultures, such as Taiwan, other studies found support for the crucial role played by competence judgments, even in Asian democracies, such as South Korea (Lin et al., 2017). The evidence gathered by Lin et al. (2017) also supports the idea that, in non-western cultures, other kinds of social inferences about a candidate, such as the threat level and open-mindedness, may play a role when choosing a candidate over another from their facial appearance.

Despite the success of inferred competence as an explanatory factor of electoral behavior, the relationship between vote intentions and the other fundamental dimension of social judgments, warmth, has not been investigated as much. Ferreira da Silva and Costa (2019) found that judgments of warmth were positive predictors of a candidate's electoral performance in several European countries, including Italy. It must be noted, however, that the judgments of warmth and competence were extrapolated from national survey data and were not inferred from the politicians' faces.

Castelli et al. (2009) also examined the role of both competence and warmth, as well as other dimensions such as age, morality, pleasantness, and regional prototypicality, in predicting a candidate's electoral success. The participants, a total of 56 Italian electors, were asked to evaluate 10 pairs of candidates on the previously listed dimensions, as well as to choose, for each of the 10 electoral races, who was the most likely winner. Results showed that the participants were more likely to select the candidate that they evaluated as more attractive, sociable, and competent. Conversely, only two dimensions were significant predictors of the actual winner of each electoral race: competence and sociability. However, while perceived competence was a positive predictor of the probability of a candidate winning, sociability, a judgment closely related to the concept of warmth, was found to be negatively associated with the probability of winning an election.

Nonetheless, perceived warmth is not always associated with negative electoral results. On the contrary, a growing body of literature supports the hypothesis that the role played by inferred competence and warmth may depend on the larger social context (Little et al., 2007; Van Vugt & Grabo, 2015). Little et al. (2007), for instance, report that changing the context from wartime to peacetime shifts the electoral preferences from competent,

masculine-looking candidates toward warmer, more feminine-looking candidates. Van Vungt & Grabo (2015) interpreted this finding using the framework of evolutionary anthropology, arguing for the evolved nature of these processes: people may prefer cooperative, warmer leaders in times of peace and prosperity to maintain the positive *status quo*; *vice versa*, people may prefer authoritative, competent looking leaders to effectively guide the ingroup out of a difficult situation. Similarly, younger leaders may be selected to face novel problems, while older, more experienced candidates may otherwise be preferred (idem). This framework may help to contextualize the otherwise incoherent results examined in the present paragraph. The results of certain studies highlighting the centrality of competence may have been exaggerated due to the social and political atmosphere in which they were conducted. For instance, the original findings by Todorov et al. (2005) may have been influenced by the tense social and political climate of the U.S. in the years immediately following the attack on the Twin Towers and the American intervention in Iraq and Afghanistan. Similarly, the results of Castelli et al. (2009), reporting warmth as a negative predictor of electoral success, may have been influenced by the 2008 financial crisis representing a menace to the population's well-being similar to that of a war. On the other hand, the positive impact on electoral results of warmth judgments highlighted by Lin et al. (2017) and by Ferreira da Silva & Costa (2019) may have been due to the more favorable social, economic, and political situation, as well as to the necessity to maintain peace.

2.2 The role of gender and gender prototypicality

The study by Little et al. (2007) reported that facial masculinity and femininity were important predictors of vote intention, although the direction of the correlation changed between war and peacetime. The design of the study, however, utilized only 8 pairs of faces, belonging to the main candidates at recent elections in the U.S. (2 pairs), the United Kingdom (3 pairs), Australia (2 pairs), and New Zealand (1 pair). Of the 8 pairs, only one consisted of two opposing candidates who were both women, the one referring to the New Zealander elections of 1999, which saw the progressive candidate Helen Clark antagonizing the conservative outgoing prime minister, Jenny Shipley. Moreover, the second part of the study, which tested the hypothesis that masculinization or feminization of a face could lead to different inferences in times of peace rather than in war, only used the faces of two male candidates, that is George Walker Bush and John Forbes Kerry. The study's design could thus lead to confusion between the dimensions of facial masculinity

and femininity and the gender prototypicality of a candidate. For this reason, the study's findings may only be valid for male politicians: more masculine, more prototypical and more competent-looking men seem to be preferred as leaders in times of war, while more feminine, less prototypical, and warmer-looking men are preferred in times of peace.

Later studies have further examined the role of gender, and that of gender prototypicality, in facial inferences from candidates' faces. The role played by gender, both in social and political judgments and in the formation of impressions from faces, is crucial, since gender is an extremely salient variable, which is automatically inferred and substantially impossible to ignore. As already mentioned in the previous paragraphs, gender is strongly linked to the two cardinal dimensions on which political choices substantially rely: in fact, men and, more generally, masculinity, are associated with competence and agency, while women and femininity are associated with warmth and communality (Volpato, 2022). This association is also replicated in the field of facial inferences: using both explicit and implicit techniques, Wen et al. (2020) were able to demonstrate that when male and female faces presented more masculine features they were more strongly associated with competence; on the contrary, when faces showed more feminine features they were more strongly associated with warmth. This evidence was further reinforced by a recent study by Cheung and Jintcharadze (2023). The authors digitally manipulated the masculinity/femininity of the faces of candidates in American elections. They then presented the faces to the participants in the form of digital three-dimensional models. The results highlighted that the participants' judgments of competence were positively correlated to higher degrees of facial masculinity. On the other hand, the participants' judgments of warmth were correlated with higher degrees of facial femininity, once again reinforcing the idea that features traditionally associated with femininity and womanhood are linked with warmth and that the same is true for masculine features and competence. Similarly, Quist et al. (2011) highlighted that, in the case of female faces, higher degrees of facial masculinity were associated with higher inferred dominance. Moreover, judgments of masculinity given by participants looking at a set of women's photographs were positively correlated with the dominance of the woman depicted as assessed through the dominance subscale of the International Personality Items Pool (Goldberg, 1999), indicating that facial masculinity may be diagnostic of a person's dominance. The interaction between gender prototypicality, judgments of warmth and competence, and vote intention has also been widely studied in recent years. For instance, an interesting line of research, inaugurated by Johns & Shephard (2007), highlighted how

the characteristics traditionally associated with the two binary genders are somewhat taken for granted by people when they are asked to evaluate a political candidate. In the study of Johns & Shephard (2007) the participants, a total of 368 undergraduate students, were asked to evaluate twelve members of the British parliament from their photographs, equally distributed for political orientation and gender (3 male, conservative politicians; 3 female, conservative politicians; 3 male, progressive politicians; 3 female, progressive politicians). The results confirmed the traditional association between masculine faces and the qualities of competence, leadership, and intelligence, which all correlated to form the factor of strength, as well as the association between feminine faces and the qualities of likeableness, attractiveness, caring, and honesty, which all correlated to form the factor of warmth. Interestingly, however, the impact of trait inferences on electoral preferences was stronger for inferences of strength for female politicians, while it was stronger for inferences of warmth for male politicians. This suggests that when people are asked to evaluate a political candidate in a context in which it is possible to compare them with other candidates of a different gender, the traits stereotypically associated with the gender of the candidate are given less importance, and the candidate is primarily judged on non-stereotypical dimensions. The fact that other previously quoted studies mainly relied on comparisons between candidates of the same gender may explain the apparently contradictory nature of this finding. Johns & Shephard (2007) also found that, compared to female participants, male participants were more likely to judge male candidates as strong and to rely on judgments of strength to cast their voting preference. In line with Johns & Shepard's (2007) findings, the evidence gathered by Carpinella et al. (2013) suggests that, since warmth is strongly connected with the cultural construction of the feminine gender, women are perceived as similarly warm by most electors independently of their assessed levels of facial femininity/masculinity. On the contrary, judgments of feminine and masculine-looking women widely varied in terms of the dimension of competence, depending on the political orientation of the participant. Liberal participants tended to see feminine women as more competent, while conservative participants tended to see them as less competent. A similar and opposite pattern emerged in judgments performed on male faces. Although masculine and feminine male faces were considered similarly competent, more men with more gender prototypical faces were perceived to be less warm by liberal, but not by conservative participants. These findings could be interpreted as an effect of higher reliance on the dimension of warmth by liberal electors and on competence by conservative electors: feminine traits, traditionally associated with

warmth, would enhance the overall judgment of a candidate for liberals, thus explaining why they would give higher judgments to prototypical female and non-prototypical male candidates; on the contrary, masculine traits would enhance the overall judgment of a candidate for conservatives, thus benefitting prototypical male and non-prototypical female candidates.

The research group led by Carpinella also investigated the relationship between gender prototypicality and actual electoral outcomes using the results of the 111th elections for the United States Congress of 2008. Carpinella et al. (2016) used FaceGen Modeller, a software able to keep track of various morphological features of digitalized faces, to operationalize the gender prototypicality of 841 candidates in the 111th House of Representatives elections of 2008. The measure of gender prototypicality was then used as a predictor of the election outcomes of the various constituencies. The results widely differed for Republican and Democrat candidates. While the former benefited from higher degrees of gender prototypicality, the same variable had no significant effect on the electoral success of Democrat candidates. These findings partially contradicted the results of the previous study. In the first study (Carpinella et al., 2013), conservative participants evaluated gender prototypical women as less competent than women with masculine features. On the contrary, the second study (Carpinella et al., 2016) highlighted how gender-prototypical, conservative women candidates obtained a higher share of the votes. This contradicting evidence may be reconciled if we hypothesize that, although conservatives see less prototypical women as more competent, in a real electoral context, they may prefer to vote for women with less competent-looking but more gender-prototypical faces, possibly because they would perceive competent-looking women as more of a threat to the social *status quo* that they aim to conserve. The evidence on the topic, however, is non-conclusive: In a study conducted to analyze the effect of the facial width-to-height ratio on judgments of dominance and vote intention, Maeng & Aggarwal (2022) found that conservative participants were less likely to vote for women in general, and for non-dominant looking women specifically. Since, as already discussed in this paragraph (Johns & Shepard, 2007), facial masculinity is probably used as a cue of dominance, this evidence openly contradicts the results from the study of Carpinella et al. (2016) that were previously discussed. In conclusion, gender and gender prototypicality certainly play an important role in the fields of inferences from faces and politics alike. However, the scientific evidence on many crucial issues is not yet unanimous, and the topic should be addressed and accounted for in future studies.

2.3 Inferences of political orientation from faces

The existing scientific literature highlights how the political orientation of the elector plays a crucial role in the importance given to the perceived traits of competence and warmth. Although the literature on the topic in the field of inferences from faces is lacking, Mignon et al. (2016) reported that, when compared to progressive electors, conservative-leaning electors tend to give greater importance to the dimension of agency and to consider less the morality and sociability of a candidate.

Although political orientation has not been extensively studied in facial inferences, the literature offers some interesting evidence.

For instance, Rule & Ambady (2010) observed that people seem to be able to infer the political orientation of a candidate with a probability significantly higher than chance, just by looking at a candidate's photograph. While in the first study the authors used only photographic portraits of actual politicians, Democrat and Republican candidates from the 2004 and 2006 U.S. Senate elections, a subsequent study published in the same research article extended the finding even to partisan electors. In fact, the participants could accurately predict the political affiliation of partisan college students. The authors also examined the possible basis for the attribution of a face stimulus to a specific party, finding that judgments of power positively predicted the attribution of an elector to the Republican Party, and that, on the other hand, judgments of warmth positively predicted the attribution to the Democrat Party. Following this evidence, the authors interpreted the process of attributing a political affiliation to a person by looking at their face as the effect of a comparison between the target's facial traits and the traits stereotypically assigned to Democrat and Republican partisans. Olivola et al. (2012) also replicated the finding that people can assess a candidate's political affiliation at an above-chance level. Their participants, 60 undergraduate students from Princeton University, were asked to guess which of the two candidates presented belonged to the Democrat and which to the Republican party in a total of 256 pairs of stimuli, each representing a real gubernatorial or Senate electoral competition from the period between 1995 and 2008. Results showed that, although there were some identification biases (for instance, non-Caucasian and female candidates were systematically categorized as Democrats), the participants could correctly categorize a candidate most of the time. Moreover, the probability that a Republican candidate was classified as a Republican was predictive of the share of votes that they collected in the real electoral competition, especially in conservative states and among conservative electors. Surprisingly, a similar effect did not appear for Democrat

candidates, for which the probability that they were categorized as Democrats was unpredictive of their actual electoral success. These findings support the hypothesis that political facial stereotypes are at least partially driven by permanent facial features. Moreover, although the exact reason for this remains unclear, these stereotypes appear more significant or noticeable to conservative-leaning voters than to progressive-leaning ones. Whereas Olivola et al. (2012) found that certain characteristics of a candidate may influence their categorization as either progressive or conservative, Herrmann and Shikano (2016) also identified certain biases in the attribution of political affiliation to a candidate. Their findings, obtained through a procedure similar to that of the studies previously analyzed, point to the fact that, although political affiliation attribution is generally accurate, participants make systematic errors by assigning their political affiliation to the candidates that they evaluated as being more attractive and competent-looking. The authors interpreted these biases as an effect of a projection of the participant's political views onto the candidates that they found more likable, or charismatic. Moreover, this finding may contribute to explaining how competent-looking candidates seem to be almost universally appreciated by both conservative and progressive electors.

Despite the presence of all these biases, even the most recent research attests that people are consistently able to make accurate political orientation judgments by only relying on facial information. In the last few years, a research group led by Kosinski (Kosinski, 2021; Kosinski et al., 2024) reported that the most recent algorithms for facial recognition can make accurate inferences of political affiliation, with an accuracy comparable to that of human raters and with a good predictive value. Although these findings raised some concerns and contributed to sparking a debate on what inferences artificial intelligences should be allowed to make from a person's photographs (Engelmann et al., 2022), they also contributed to the advancement of the comprehension of the morphological correlates of facial political stereotypes.

For instance, Kosinski et al. (2024) reported that the only face characteristic that significantly correlated with political affiliation attributions was the ratio between the higher and the lower part of the face, with conservatives tending to have bigger lower faces when compared to progressives. Nonetheless, the predictions based on this ratio were far worse than the predictions made by humans and artificial intelligence. Moreover, none of the other facial characteristics considered, such as the facial width-to-height ratio (fWHR) or the Body Mass Index (BMI), was predictive of the evaluations of political

orientation. The authors argue that such data suggests that, although morphological hints to a person's political orientation certainly exist, since they are captured by both human participants and artificial neural networks, they generally seem to be far more subtle, complex, and configurational than what has been studied hitherto.

2.4 The Role of Elector-Candidate Similarity

Similarity constitutes one last important predictor of an elector's vote intention. Extensive evidence supports the claim that perceived similarity with a politician increases an elector's vote intention towards them. For instance, Caprara et al. (2007) asked their participants to report both their personality traits and the personality traits of two opposing political candidates (either George W. Bush and John Kerry or Silvio Berlusconi and Romano Prodi, depending on the nationality of the respondent). The authors then assessed that, both in Italy and in the United States, participants were more likely to vote for a candidate with a personality that they assessed to be similar to their own. This effect can be explained as a combination of the participants projecting their positive personality traits onto a politician that they already like, in a similar fashion to that of the study previously discussed by Herrmann & Shikano (2016), and the participants actively selecting politicians with personality traits more similar to their own. Numerous other studies support the hypothesis that electors tend to vote for candidates with a personality like their own. A recent study by Aicholzer & Willmann (2020), for example, determined that people tend to wish for their leaders to be like them, at least on value-based personality traits, in a way resembling the mechanisms of assortative mating, for which people tend to choose partners with matching personality characteristics (Rammstedt et al., 2013). Additionally, the participants of Aicholzer & Willmann (2020), a random sample of 631 people representative of the Austrian population, also generally preferred candidates to be more emotionally stable, extraverted, conscientious, and honest than the average citizen, but also less agreeable. Nonetheless, although the strongest predictor of vote intention was the evaluation of consciousness, none of the other traits were more predictive than self-similarity. Overall, the authors interpreted the results as people desiring to be represented by politicians with personalities similar to their own but also endowed with more positive qualities capable of making them effective leaders.

There are numerous similarities between electors and politicians that can play a role in predicting vote intentions. Sevi (2021), for instance, reports that electors generally tend to vote for candidates with whom they share their socio-demographic profile. Scientific

literature reports that the similarity effect is present for a wide range of socio-demographic characteristics. For instance, Brians (2005) reports that women are marginally more likely to vote for female candidates and that Democrat women candidates benefit from crossover votes from Republican women electors when facing Republican men candidates. Similarly, Besco (2019) reports that the self-identification of ethnic minorities with political candidates of the same ethnic group predicts vote intention, even more than the policies proposed by a candidate and than the candidate's ideology. Lastly, Sevi (2021) demonstrated that age similarity can be a small but significant predictor of vote behavior, with younger electorates generally preferring younger candidates. As a whole, the role of socio-demographic similarity in predicting electoral behavior could be interpreted as an effect of the struggle, especially of marginalized and minoritarian social groups, to find true representation in the modern political landscape, still overwhelmingly dominated by specific demographics (for instance, the current composition of the European Union Parliament still sees a prevalence of male politicians and people over the age of 50; Sabbati, 2020).

Since, as extensively discussed in the previous paragraphs, people are consistently able to infer personality traits from facial cues, and since some of the most important socio-demographic variables, such as gender and race, can be easily inferred from photographs, the evidence hitherto discussed in the present paragraph supports the hypothesis that facial characteristics could lead to vote intention through the mediation of inferences of similarity. Moreover, facial similarity could lead to higher inferences of trustworthiness (Nakano & Yamamoto, 2022) toward strangers. In the field of political psychology, this could lead to stronger vote intention for candidates whose faces resemble the elector's.

Bailenson et al. (2006; 2008) tested this hypothesis by artificially manipulating facial similarity between their participants and a political candidate using the morphing software Magic Morph. In the first pilot study (Bailenson et al., 2006), the authors asked their participants, 76 undergraduate students from Standford University, to provide them with a frontal photograph of their faces with a neutral expression.

For half of the participants, the photograph was then morphed with the portrait of an unknown male Caucasian politician, who was later presented to the participants as Tom Steele, a Democratic candidate for the legislative assembly. In this case, morphing between the participant's and Steele's faces was set to a 40:60 blend, meaning that the participants were presented with a photograph of a person who was still recognizable as Steele but also a significantly increased similarity to each participant's photograph. The

other half of the participants saw the unmanipulated photograph of Tom Steele. After the presentation of the photograph, the participants were asked to express their voting intention toward Steele. The participant's responses were then categorized into two main categories: participants willing to vote Steele and undecided or unwilling participants. The participants were also asked to evaluate their impression of Steele on a "feeling thermometer" from 0 ("extremely cold") to 100 ("extremely warm") and to evaluate Steele himself on various traits, such as intelligence, sincerity, industriousness, and attractiveness. In general, the authors did not find a significant effect of similarity on vote intention for Steele. However, when interacting with gender, facial similarity provided a significant predictor of numerous measures. Male participants showed higher vote intention towards Steele in the similarity condition. In contrast, female participants showed an opposite effect, and their vote intention was reduced when Steele's face resembled their own. Similarly, while male participants provided, on average, better evaluations of Steele in the morphing condition both for attractiveness and on the feeling thermometer, female participants provided lower evaluations of Steele on these dimensions when compared to the control condition. Substantially, while male participants preferred the morphed Tom Steele, female participants preferred the original. The authors offer different interpretations of this effect. Firstly, they hypothesize that morphing between photos of different genders could lead to more unrealistic results, leading the female participants to see an "uncannier" version of Tom Steele, which may have induced worse evaluations of the candidate. Alternatively, cross-gender morphing could lead the female participants to see their traits mixed with traits traditionally associated with the opposite gender. In turn, this may lead to an "unconscious priming" (Barg et al., 1996) that could elicit an inter-group thinking process, leading women to reduce their appreciation for a candidate who is being perceived as a member of an outgroup. Lastly, the authors propose that the effect may rely on self-categorization (Hogg et al., 1987), with women finding it more difficult to process self-relevant traits in the face of a member of the gender outgroup.

An alternative explanation could be offered by the effect of gender prototypicality, previously discussed in paragraph 2.3. Whereas male participants would have seen a more gender-prototypical version of Tom Steele's face, female participants would have seen a less prototypical, more feminine version of the candidate. According to Carpinella et al. (2013; 2016), a more gender-prototypical face could lead conservative people specifically to a greater appreciation of a candidate. This could have been the case for Tom Steele,

who was presented as a Democrat candidate, and, when masculinized, could have attracted votes and better evaluations even from more conservative participants. The data presented by the authors, however, are insufficient to confirm or dismiss this last hypothesis. A subsequent paper, published by the same research group (Bailenson et al., 2008), tried to deepen the scientific understanding of how facial similarity impacts vote intention even in large-scale, gubernatorial and presidential elections. Three studies were conducted. In the first study, the participants had to choose between two unfamiliar candidates for the 2006 Florida gubernatorial elections, Charlie Crist and John Davis. Both Crist and Davis can be identified as Caucasian males. Each participant saw one candidate morphed to themselves, and the other morphed to another random participant. Which candidate was morphed with the participants was decided randomly. The second study followed a similar structure but utilized the photographs of the two main candidates for the 2004 presidential elections, John Kerry and George W. Bush. In this case, too, the candidates were two Caucasian males. In the third study, a larger set of candidates' photographs for the 2008 presidential elections was used, only some of which were widely known. In this case, four of the eight candidates were Caucasian women (i.e., Hillary Clinton, Kate Hutchinson, Elizabeth Dole, and Jennifer Granholm).

In each of the three experiments, the participants were asked to express an affective evaluation of each candidate shown and their vote intention. Moreover, they were asked about their political affiliation, gender, and education level.

Overall, the results from the three studies support the hypothesis that facial similarity can lead to a small, but significant increase in both trait judgments of a candidate and vote intention. Moreover, the authors reported that for the effect to be present the similarity did not need to be particularly intense: the condition with a 40:60 blend and with a 20:80 blend between the participant and the candidate did not show significant differences. Additionally, the effect was reported to be stronger in the case of previously unknown candidates, toward whom the participants did not have strong pre-established attitudes. The effect was also stronger for non-partisan participants when compared to Democrat and Republican partisans, who tended to express higher evaluations of their party's candidate independently of the similarity manipulation.

Interestingly, the peculiar effect reported in the previous study (Bailenson et al., 2006), which saw the female participants diminish their preference towards candidates morphed with themselves, did not replicate in this study. It is worth noting, however, that the study lacked completely non-Caucasian candidates, and excluded completely non-Caucasian

participants, as well as participants with facial hair or glasses. Since only four of the twelve candidates' pictures used across the three studies belonged to women, women were also under-represented in the data. These factors could impede the generalizability of the otherwise fascinating finding.

2.5 Main Gaps in the Existing Literature

The different lines of research presented in this chapter clearly show the important role played by trait and similarity inferences from faces in politics. However, the scientific evidence also presents some gaps.

Firstly, the complex role of gender in the electors' judgments of politicians is still relatively underexplored. In a world gradually progressing toward greater gender equality, even in fields traditionally occupied by men, more research is needed to determine whether judgments of female politicians follow analogous patterns to the ones made toward male politicians or whether they are based on entirely different criteria. As discussed in the previous paragraph, the role of the electors' gender is also under-analyzed when it comes to the effects on vote intentions of both facial (Bailenson et al., 2006; 2008) and personality similarity (Caprara et al., 2007). Future studies need to account for the possible effects of both the individual's and the politicians' gender, giving greater attention to including a fair share of female facial stimuli in the experimental material.

Chapter III – Materials and Methods

3.1 Main Aims and Hypotheses

The present study aims to build and expand on the existing scientific evidence discussed in the previous chapters. The existing research on the topic of inferences from candidates' faces and vote intention highlights the importance of three main factors. Firstly, the most important predictor of vote intention seems to be the perceived facial competence of a candidate (e.g., Todorov et al., 2005; Atkinson et al., 2009; Sussman et al., 2013), although warmth also plays a part according to certain studies (e.g., Little et al., 2007; Castelli et al., 2009). Secondly, both personality (e.g., Caprara et al., 2007) and facial (e.g., Bailenson et al., 2008) similarities between the elector and a candidate seem to play a role in positively predicting the vote intention of the elector. Lastly, the gender of a candidate and their gender prototypicality seem to be significant factors, although their effect is often unclear, and some results are contradictory (e.g., Carpinella et al. 2013; 2016). The confusing effects of a candidate's gender may depend on the fact that, although numerous policies have been implemented, women are still underrepresented in the field of politics, even in Western democracies such as the U.S. and the European Union (Sanbonmatsu, 2020; Sabbati, 2020). This under-representation, combined with traditional stereotypes and prejudices about women, may lead electors to see women candidates as less of leaders compared to men and to apply different rules of judgment to candidates of opposite genders. Be as it may, it appears clear that findings obtained using exclusively or almost exclusively stimuli portraying male candidates may not always generalize to female candidates.

Given that the evidence shows that facial similarity is a criterion that people use, either consciously or unconsciously, to guide their vote intention (Bailenson et al., 2008), the present study asks the following research question: Do people use facial similarity between an elector and a candidate as a criterion to make inferences on the vote intention of other people?

Since, as extensively examined in Chapter I, research in evolutionary psychology has determined that facial similarity is commonly used as a cue of kinship (e.g., DeBruine et al., 2009), people may attribute vote intention based on facial similarity because they would see the elector and the candidate as members of "the same clan". Although the morphological traits linked with being conservative/progressive are still not fully understood (Kosinski et al., 2024), the increased similarity between an elector and a

candidate may lead to the attribution to the same political alignment (Rule et al., 2009; Rule & Ambady, 2010). Moreover, since facial similarity is likely to increase the similarity of the perceived personalities of the elector and the candidate, people could evaluate physically similar electors and candidates as more psychologically and, consequently, more ideologically similar.

In line with the presented evidence, we propose the following main hypothesis:

H1: When asked to express the probability that an elector will vote for a candidate or another, people will assign a greater vote probability to the candidate that looks the most similar to the elector.

Since evidence shows that women perform better in tasks of face recognition (Rehnman, 2007; Herlits & Lovén, 2013) and that people tend to exhibit a bias in face recognition favoring their own gender (Herlits & Lovén, 2013), we also hypothesize that:

H2: Compared to men, women will assign a higher probability of the elector voting for the candidate they look the most similar.

H3: A participant will assign a higher probability of the elector voting for the candidate to whom they look the most similar when the elector and the candidates belong to the participant's gender.

Due to the inconclusive evidence gathered by Bailenson et al. (2006; 2008) on the effects of the participant's gender on vote intention toward a candidate manipulated to resemble the participant, we propose two alternative hypotheses on the possible role played by the electors' and the candidates' gender.

Firstly, since in the current experiment, vote attribution would regard a fictional elector instead of the participant themself, we hypothesize that:

H4a: There will be no differences in the assigned probability to vote for the candidate who was morphed to resemble the elector the most between the condition in which the candidates are male and the condition in which are female.

Alternatively, in line with the results reported for vote intention for female participants by Bailenson et al. (2006), facial similarity between the elector and a candidate may not be used when the elector belongs to the female gender.

H4b: The effect of similarity on the attributed probability of the elector voting for one candidate over the other will be stronger when the judged faces are male than when they are female.

3.2 Participants

3.2.1 Prospective Power Analysis

Before publishing the questionnaire, a prospective power analysis was performed to estimate the number of participants needed to reliably detect the effect of similarity on vote attribution. In the absence of previous studies on the topic, it was conservatively estimated that the effect was comparable to the one detected by Bailenson et al. (2008) regarding the effect of similarity on vote intention (Cohen's d = 0.2). To estimate the sample size, the function pwr.t.test of the R package "pwr" was used (Champely, 2022). In the context of a one-sample t-test with a significance level of 0.05, to reach a *power* = 0.8 the software estimated a sample size of 198.15 ≈ 199 subjects, while to reach a *power* = 0.95 the sample size was estimated at 326.80 ≈ 327 individuals. Following this analysis, it was decided to recruit a sample of at least 500 participants in order to have a sufficient sample size even after the potential elimination of inadequate responses.

3.2.2 Recruitment of the Participants

The questionnaire, estimated to take approximately 5 minutes to complete, was designed to be accessible to a wide range of participants. The pictures used in the questionnaire were considered sufficiently visible even on a smartphone, allowing participants to take the survey on either a computer or a mobile device. To ensure a vast and diverse sample, a snowball sampling method was implemented, primarily involving sharing the questionnaire on various social media platforms (Facebook, Instagram, and Whatsapp). A QR code was also produced to reach additional participants via fliers, further enhancing the inclusivity of the recruitment process.

Participation in the survey was voluntary, uncompensated, and reserved for individuals over the age of 18. Within the informed consent, the confidential and anonymous nature of the collected data was emphasized, as well as the option to withdraw from the study at any time. The research was approved by the ethical committee of the University of Padua and conducted following the ethical norms outlined by the Italian Psychological Association.

3.2.3 Characteristics of the Sample

While a total of 846 people started the survey, only 566 responses were considered adequate. 280 people were excluded from the analysis, including:

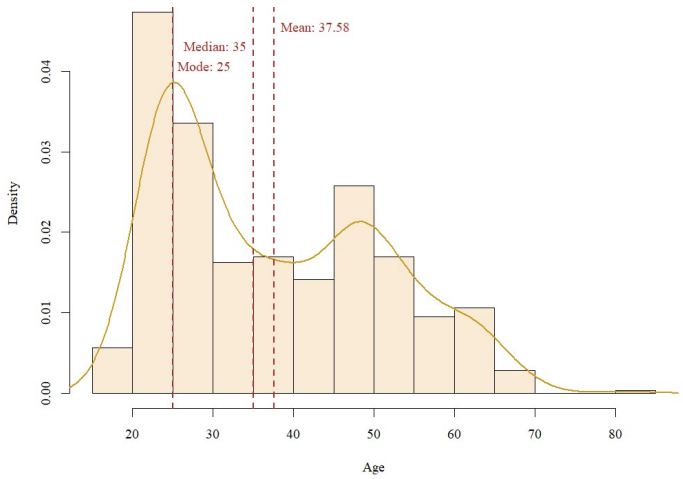
- 12 participants who did not give their consent to participate in the study;
- 188 participants who did not fully complete the questionnaire;
- 80 participants who did not answer the attention check correctly.

The sample was strongly unbalanced for gender, being largely composed of female participants (*N* = 432, 76.3% of the total). The remaining participants identified either with the male (*N* = 128, 22.6% of the total) or the non-binary gender (*N* = 6, 1.1% of the total). Due to the low number of non-binary participants, these participants were excluded from the analysis regarding the role of the variable gender.

The participants had a mean age of 37.58 (*sd* = 13.52). However, the distribution (Figure 4) of the variable age exhibited a positive skewness, with the median age (*Mdn* = 35) being lower than the mean age and the mode (*Mo* = 25) being even lower. This suggests that the observations are concentrated on the left, with fewer observations on the upper end of the distribution, resulting from the fact that elder ages are underrepresented in the sample.

Figure 4

Histogram of the variable "Age"



Note. Distribution of the age variable, with the density curve overlaid. The dashed lines represent the mean, the median, and the mode age.

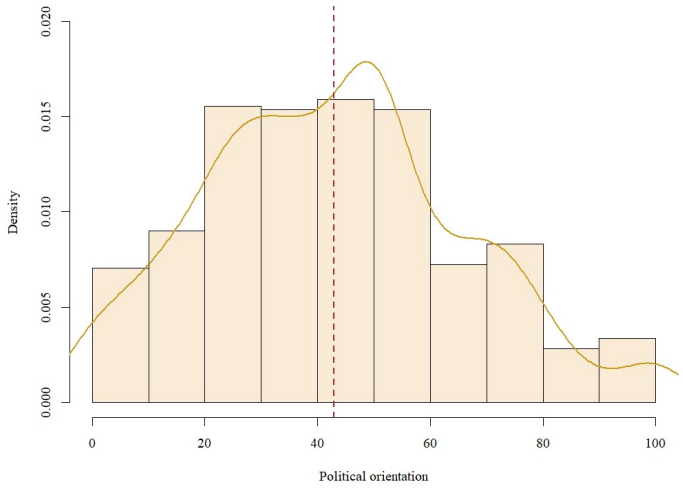
The sample was composed mainly of well-educated people. The vast majority of the participants possess a degree equivalent to or higher than a high school diploma (N = 528,

93.2% of the sample), and 310 people (54.8% of the sample) possess at least a bachelor's degree.

As expected, the sample was overwhelmingly composed of people of Italian ethnicity (N = 546, 96.5% of the total), with the remaining participants being of European/non-Italian origin (N = 12, 2.1% of the sample), African (N = 2, 0.4% of the sample), belonging to another ethnic minority (N = 30, 0.5% of the sample), or refusing to share their ethnicity (N = 3, 0.5% of the sample). Due to the low number of non-white participants, it was decided to exclude them from further analysis since it was evaluated that they could have perceived the faces shown during the study as members of an ethnic outgroup. Lastly, likely as a result of the relatively young mean age, the relatively high average degree of instruction, and the strong presence of women in the sample, the participants' political orientation was significantly shifted, t (565) = -7.3728, p < .001) towards the political left (M = 42.9, Mdn = 42.3, sd = 22.9). However, the participants covered the whole range of the political spectrum, and no political orientation was particularly underrepresented (Figure 5).

Figure 5

Histogram of the variable political orientation



Note. Distribution of the political orientation variable, with the density curve overlaid. The dashed line represents the mean.

3.3 Materials

3.3.1 Artificially generated faces

The study was conducted through an online distributed questionnaire (see Appendix I and paragraph 3.4.1) that required 18 faces, equally distributed for the two binary genders (9

female faces; 9 male faces). The faces of the fictitious politicians and electors needed for the study were generated by a photorealistic model of Artificial Intelligence (AI) through the website This-Person-Does-not-Exist.com (Lopuka, 2021). The website's creator describes the photographs published as the result of the work of two competing AIs. The first AI has been trained on data from true photographs and specializes in producing photorealistic portraits of non-existing individuals. The second AI has been trained to distinguish between artificially generated and real photographs of a person's face. The photographs appearing on the website are generated by the first AI, and survive the selective process of the second, meaning that, at least for the second AI, they are indistinguishable from photographs depicting real human beings. For the present study, a total of 26 photographs were generated and later underwent a further selection following their evaluation on a series of variables relevant to the study by a small sample of participants (see paragraph 3.2.1).

Of the 26 photographs:

- 13 were generated with the "Male" gender option and the "White" ethnicity option enabled, and were assigned an alphanumeric code from MI1 (Man-Ingroup) to MI13;
- 13 were generated with the "Female" gender option and the "White" ethnicity option enabled and were assigned an alphanumeric code from WI1 (Woman-Ingroup) to WI13;

Only white photographs were included in the sample since it was expected that the sample would be predominantly white. The literature on face recognition highlights that people are generally bad at seeing subtle similarities and differences in the morphologies of individuals belonging to a racial outgroup (O'Toole et al., 1994).

This effect, which has been known for over a century (Feingold, 1914), is known as the "other race effect" and it is explained by the fact that the faces of members of an outgroup seem to be processed through a more superficial, local elaboration (Hughes et al., 2019) and to be perceived more like members of a group than as individuals. Studies show that this kind of elaboration of faces is insufficient to produce accurate judgments of similarity between faces (Bicego & Grosso, 2019). Moreover, presenting the participants with both faces of an ethnic ingroup and an ethnic outgroup may have elicited some form of inter-group thinking that may have compromised the results.

Although, as previously noted, the median age of candidates in Western democracies tends to be much higher, all 26 faces tested were generated using the option "26-35 years old", as it seemed to more easily produce high-quality photographs when compared to other age group options, that appeared to produce unusable results more frequently. In particular, the photographs had to be suitable for the morphing procedure (see paragraph 3.2.2) and thus had to possess a series of characteristics listed below, partially inspired by the criteria used by Bailenson et al. (2006; 2008).

1. *Frontal position*

   This position allows for a complete, symmetrical view of the face. Frontal, symmetrical faces are easier to morph without giving rise to unrealistic results. Moreover, when decontextualized from the photograph, a more asymmetrical face could be judged as less attractive (Zheng et al., 2021).

2. *Presence of a smile and visibility of the teeth*

   Most of the research on face inferences, starting with Todorov et al. (2005), has been using photographs of individuals with a neutral expression. However, since ThisPersonDoesNotExist.com tries to replicate realistic photographs, most of them display either positive or negative emotions. It seems like the AI has a strong gender bias since it tends to portray women smiling more often than men. However, the evidence for this bias is anecdotal, and its presence could be easily explained by biases in the training data of the AI. Since it was otherwise nearly impossible to obtain a high enough number of photographs of non-smiling women, it was decided that all the photographs had to adhere to these characteristics, to make the morphing process easier and not to have faces that could be judged as intrinsically warmer than the others. Smiles, in fact, are associated with higher evaluations of warmth (Wang et al., 2017).

3. *Absence of unrealistic elements in the image*

   Despite most of the photographs present on the website being extremely realistic, some of them may present unrealistic elements even at first glance. For instance, earrings, hats, clothing, and facial piercings may be deformed and present incoherent perspectives and distorted and unrecognizable elements in the background, contributing to the impression of an undeniably artificial image. The photographs presenting unrealistic elements covering parts of the face were thus excluded.

4. *Absence of beard*

   Only clear-shaved male faces were deemed acceptable for the study. Evidence shows that bearded men tend to be perceived as more masculine, aggressive, dominant (Addison, 1989), and having a higher social status (Dixson & Vasey, 2012). Moreover, the presence of a beard in only one of two morphed faces would result in an extremely unrealistic picture.

5. *Absence of "heavy" makeup*

   Pictures of women wearing "heavy" makeup were excluded. Specifically, dark and heavy eye makeup and vibrant lipstick led to the exclusion of the photographs, since the aim was to utilize only pictures that could pass for being shots of politicians used in electoral posters. Moreover, evidence in the scientific literature suggests that heavy makeup use around the eye area may lead to higher levels of cognitive objectification (Bernard et al., 2019), which, in turn, may impede the global elaboration of the face.

3.3.2 Selection of the artificially generated faces

Estimates of gender prototypicality, warmth, competence, and morality were also needed for each face to form balanced pairs, in which none of the fictitious candidates were particularly distinct from the other for a certain characteristic that may impact vote attribution. A questionnaire was prepared by using the online platform *Qualtrics* and was distributed from July 14th to July 21st, 2023. The faces were presented sequentially in a random order, and the participants were asked to evaluate each stimulus on the following dimensions:

- *Competence*, on a slider from 0 ("Not competent at all") to 10 ("Extremely Competent");
- *Warmth*, on a slider from 0 ("Not sociable at all") to 10 ("Extremely sociable");
- *Morality*, on a slider from 0 ("Not moral at all") to 10 ("Extremely moral");
- *Masculinity*, on a slider from 0 ("Not masculine at all") to 10 ("Extremely masculine");
- *Femininity*, on a slider from 0 ("Not feminine at all") to 10 ("Extremely feminine");
- The picture's realism, on a slider from 0 ("Not realistic at all") to 10 ("Extremely realistic"). The participants were informed at the beginning of the questionnaire

and later reminded that every face presented was an artificially generated imitation of a human face; thus, they knew that the faces were fictitious.

After evaluating 16 faces on all dimensions, the participants were asked a few demographic questions (gender, age, highest instruction degree, ethnicity, political affiliation). The duration of the questionnaire was estimated at 15 minutes.

The participants were then asked to confirm the consent they previously gave to treat their data and were successively thanked for their participation.

A convenience sample of 98 participants was recruited by diffusing the questionnaire on various social media (Whatsapp, Instagram, Facebook). 96 participants confirmed their consent at the end of the questionnaire (26 male, 3 non-binary participants). The mean age of the sample was extremely low (M=30.3, sd=12.01), most of the participants declared to have obtained a bachelor's degree (n=51) and the overwhelming majority identified with the Italian ethnicity (n=90), with only 2 participants of non-white ethnic origin. On average, the sample presented a remarkably progressive political orientation (M = 3.37 on a range from 0 to 10, sd = 1.93). Overall, the sample can be considered highly non-representative, as it is composed prevalently of young, well-educated, progressive-leaning people.

During the data analysis, an average score for every variable evaluated in the study was calculated for each face. The faces were generally judged to be very realistic (M = 7.99, sd =1.57), sufficiently competent (M = 6.24, sd =1.53), warm (M = 6.58, sd = 1.40), and moral (M = 6.21, sd = 1.43).

It was decided to exclude from further studies:

- Faces whose average scores exceeded the 2 standard deviations above or below the mean for one or more of the dimensions of competence, warmth, and morality;
- Masculine faces that presented excessive average scores of femininity (2sd above the average for masculine faces) and/or deficient average scores of masculinity (2sd below the average for masculine faces);
- Feminine faces that presented excessive average scores of masculinity (2sd above the average) and/or deficient average scores of masculinity (2sd below the for feminine faces);
- All faces evaluated as unrealistic (2 standard deviations below the average realism).

A total of 6 faces were excluded (3 male faces, 3 female faces).

3.4 Methods

3.4.1 The experimental manipulation: Morphing faces to enhance similarity

The previously tested fictitious facial pictures were organized into six triplets, homogeneous for gender. The triplets were composed of two fictitious politicians and an elector. The two candidates' faces were paired to ensure no significant differences in evaluations across all previously tested dimensions (competence, warmth, morality, masculinity, and femininity), as determined by conducting paired t-tests on the scores of each photograph for each variable.

The following pairs were formed:

- MI3-MI12 (from now on the pair will be referred to as Male Triplet 1, MT1);
- MI6-MI9 (MT2);
- MI7-MI13 (MT3);
- WI2-WI8 (Female Triplet 1, FT1);
- WI7-WI11 (FT2);
- WI10-WI12 (FT3).

The electors' faces were randomly selected between the remaining pictures and then matched randomly with an existing pair of candidates.

As visible in Table 1, picture MI11 was assigned to the pair of candidates designated with the code MT1, MI2 to the pair MT2, and MI10 to MT3.

For feminine faces, the picture WI5 was assigned to the pair of candidates FT1, WI9 to FT2, and WI1 to FT3.

Differently from the procedure described by Bailenson et al. (2006; 2008), we chose to manipulate the electors' faces, instead of the candidates'. While in the original study it was necessary to manipulate the candidates' faces, since the goal was to enhance the similarity with the participant, in the present case it was possible to induce similarity without changing the candidates' faces. By doing this it was possible to maintain stable the characteristics of the candidates, thus consenting to use the previously obtained estimates of competence, warmth, and morality for each candidate as alternative predictors of vote attribution during the analyses.

Table 1

Experimental Triplets of Faces Used in the Questionnaire

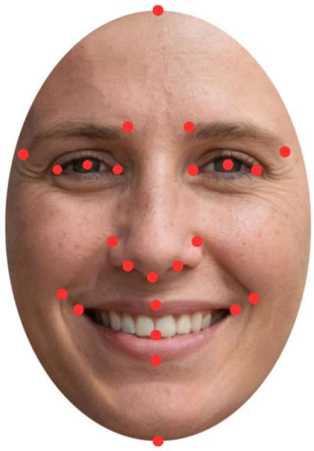
Triplets' code	Elector's face	Candidates' faces	
MT1			
	MI11	MI3	MI12
MT2			
	MI2	MI6	MI9
MT3			
	MI10	MI7	MI13
FT1			
	WI5	WI2	WI8
FT2			
	WI9	WI7	WI11
FT3			
	WI1	WI10	WI12

Note. In the questionnaire, the presentation order of triplets, as well as of the candidates within the triplets, varied randomly.

Each morphing process required manually placing control points first on the face of the elector and then on the face of the candidate, generally on key features of the face, such as the eye, the mouth, and the nose area. At least 24 control points were placed on each face in various key positions, such as the eye, the mouth, and the nose area. The precise positions of the morphing points are represented in Figure 1.

Figure 1

Standard control points positioned on the face of the elector WI9

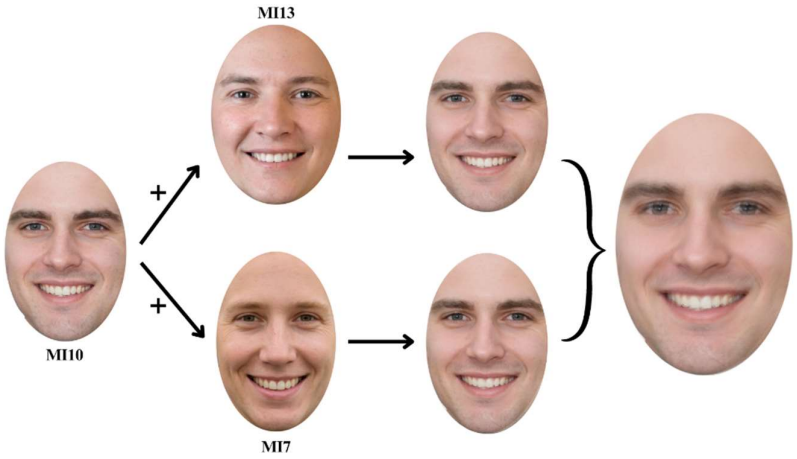


Note. Additional points were sometimes added to increase the morphed face's quality.

For each elector's picture, two different versions were created, each resembling one of the two candidates of the same triplet (Figure 2). The morphing level of the elector's and the candidate's faces was set at an 80:20 proportion, meaning that the resulting face only slightly resembled the candidate. Although Bailenson et al. (2006, 2008) used a higher ratio (40:60) in most of their experiments on the effects of facial similarity between the participants and political candidates, it was chosen to lower the similarity ratio, since in the presented study participants would have been exposed to the elector's and the candidates' photographs at the same time, and higher levels of facial similarity could have made the manipulation too obvious, leading participants to guess the main experimental hypothesis. Moreover, Bailenson et al. (2008) reported that the effects of facial similarity did not significantly vary when the similarity ratio was at 20:80 compared to when it was at 40:60, indicating that the similarity manipulation, although less noticeable, is still effective at the lower ratio.

Figure 2

Illustration of the result of a morphing process between faces



Note. The figure shows how the elector's photo, MI10, is transformed after being morphed either with the candidate MI13 or the candidate MI7. The results, visible on the right, are two faces highly resembling the original MI10, with only slight differences in the shape and the positioning of the facial characteristics. The difference between the two images can be better seen on the last, larger face on the right, which consists of the two morphed faces superimposed one over the other: the blur visible in the picture indicates the areas that have slightly moved or changed shape.

3.4.3 The questionnaire structure

The questionnaire was created on the online platform Qualtrics, and it was composed of three main parts:

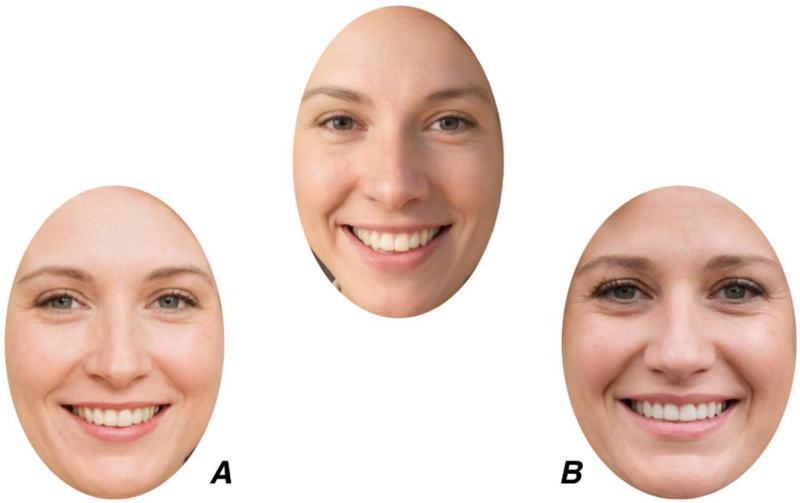
- the experimental part, composed of 10 questions of vote attribution;
- the self-evaluation part, composed of 10 questions in which the participants were asked to evaluate the possible criteria underlying the choices they previously made;
- the demographic section, composed of 7 questions on the participant.

Firstly, the participant would have been asked for their consent to participate in the study. If a participant denied their consent, they were thanked for their time and the questionnaire would end. On the contrary, if the participant accepted to participate in the survey, they could access the rest of the questionnaire.

In the first section, the participants were shown 10 triplets of fictitious faces, each described as being composed of one elector, presented on the top part of the image, and two opposing candidates, at the bottom left and the bottom right of the image, who were facing each other in a fictitious electoral race (Figure 3).

Figure 3

The experimental triplet FT1



Note. In this case, the elector's face (WI5, on the top) was manipulated to resemble the candidate "A" (WI8).

Of the 10 triplets, 6 (MT1, MT2, MT3, FT1, FT2, and FT3) were experimental, while 4 were confounding triplets that had not been manipulated to enhance the similarity of the elector to one candidate or another. These four confounding triplets were composed of the remaining pre-tested faces that went unused as experimental stimuli, as well as of some new faces generated from Thispersondoesnotexist.com. Each participant saw all 10 triplets, and the presentation was presented in a random order.

For each triplet, the participants were asked to use a slider to assess the probability that the elector had voted for one candidate or the other, on a scale from 0 ("The elector certainly voted for candidate A") to 100 ("The elector certainly voted for candidate B"). The slider was positioned by default at 50, the middle point between the two extremes of the scale.

To avoid a possible association between the positioning of a candidate's face and the political concepts of left and right, it was decided to balance each triplet by randomly presenting a specific candidate's face on the right part of the screen for 50% of the participants, and on the left for the other 50%. This balance may also help to avoid other biases in the data, such as a possible tendency to move the cursor in a certain direction over the other and the spatial-agency bias (Maas et al., 2009), according to which people represented on the left side of a picture tend to be perceived as more agentic than people represented on the right side. Moreover, it was decided to present the elector morphed

with a candidate to half of the participants, and with the other candidate to the remaining half. During the analysis, the data were then re-coded to be always higher when the participant moved the cursor in the direction of the candidate whom the elector was manipulated to resemble the most. In this section, the timing of the participant answering each question was also recorded.

After the presentation of all 10 triplets, the second section of the questionnaire was shown to the participants. This section comprised 10 questions, 8 of which aimed to induce the participants to reflect and report on the criteria used to assign the probability of the electors voting for one candidate or the other. Each question was formulated as a first-person affirmation, and the participant had to judge whether the sentence accurately described their thought process by moving the cursor on a slider from 0 ("Absolutely false") to 100 ("Absolutely true").

The questions addressed the following criteria:

- Competence ("I assigned higher vote probability to the candidate that I considered more competent");
- Warmth ("I assigned higher vote probability to the candidate that I considered more sociable");
- Morality ("I assigned higher vote probability to the candidate that I considered more moral");
- Attractiveness ("I assigned higher vote probability to the candidate that I considered more attractive");
- Maturity ("I assigned higher vote probability to the candidate that I considered more mature");
- Physical similarity with the elector ("I assigned higher vote probability to the candidate that I considered more physically similar to the elector");
- Self-similarity ("I assigned higher vote probability to the candidate that I considered more similar to me");
- Personality similarity with the elector ("I assigned higher vote probability to the candidate that I considered more similar to the elector in regard to their personalities").

Moreover, the participants were asked to assess if they tend to rely on first impressions ("I consider myself a person who lets myself be guided by the first impressions I have of other people"). Lastly, the section included an attention check in the form of a question

asking the participant to move the cursor completely toward "absolutely true" if they were paying attention. To prevent possible effects of the order of presentation, all the questions of the section were presented in a random order.

After completing the second section, the participants were introduced to the third and final section with a brief message indicating that they would be asked some questions about themselves. The section included questions about the following information about the participants:

- Gender, a categorical variable assessed through a close-ended question with 4 levels ("female", "male", "non-binary", and an additional level for people who did not want to disclose their gender);
- Highest obtained degree of instruction, a categorical variable assessed through a close-ended question with 7 possible answers, ordered from "no educational qualifications" to a "PhD degree or equivalent title";
- Ethnic origin, a categorical variable assessed through a close-ended question with 5 levels ("Italian", "European/non-Italian", "African", "other ethnic minority", and an additional level for people who did not want to disclose their ethnicity);
- Age, a numerical variable assessed through a text entry. To exclude people who had not reached the age of majority and to prevent errors such as entering the same number twice, which might have resulted in recording unrealistic ages, the question was set to accept a minimum value of 18 and a maximum of 99;
- Political orientation, a numerical variable assessed through a slider, from 0 ("Far left") to 100 ("Far right");
- Political sophistication, a variable measured using two questions:
    - "How interested are you in politics?"
    - "How informed do you consider yourself about politics (current events and political figures; party platforms; etc.)?

Each question had to be answered with a slider, from 0 ("Not at all") to 100 ("Very much").

Chapter IV – Data Analysis

The present chapter analyzes the data from the responses obtained through the questionnaire. All the analyses were conducted using the open-source software R (R core team, 2024) and its supplementary packages. In particular, Readxl (Wickham & Bryan, 2023) was used to import the data downloaded from Qualtrics, Corrplot (Wei et al., 2021) to graphically represent correlations, sjPlot (Lüdecke, 2023) and ggplot2 (Wickam et al., 2016) to implement most graphical representations, Generalized Linear Mixed Models using Template Model Builder (glmmtb; Brooks et al., 2017) to produce beta-regression models and rcompanion to calculate the goodness-of-fit for regression models (Mangiafico, 2024).

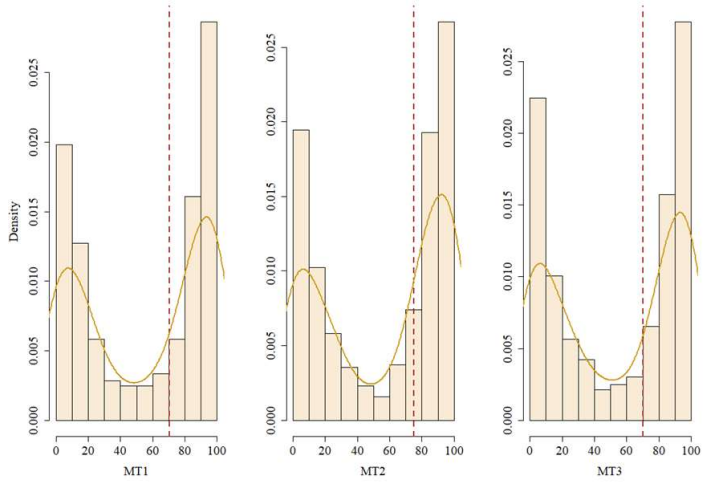
4.1 Analyzing Vote Attributions for each Triplet

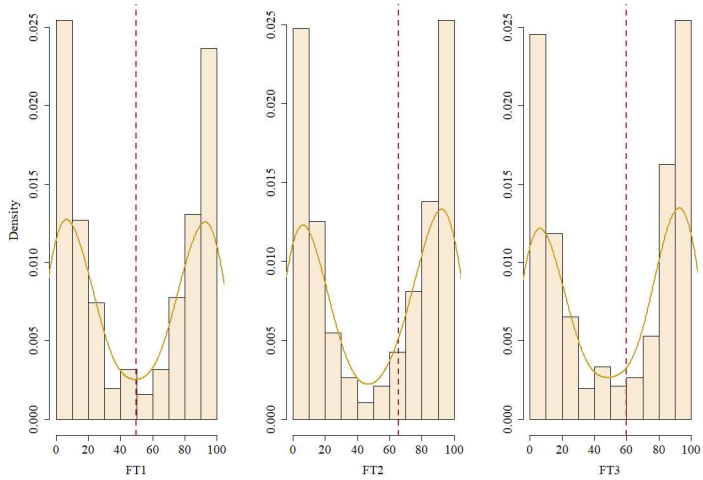
To properly analyze the effect of the manipulation on the participants' vote attribution tendencies, the data of the six experimental questions of vote attribution were recoded to show higher values when the participant moved the cursor toward the candidate whose face resembled the elector the most.

As visible in Figure 6, all six triplets show a remarkably non-normal distribution of the answers. The shape of the distributions seems to suggest that the participants prefer giving decisive and polarized vote attributions to either a candidate or the other rather than expressing uncertainty and remaining "neutral" by keeping the cursor near the center.

Figure 6

Histograms of the vote attribution variable for the six experimental triplets





Note. Distributions of the vote attribution variable in the dataset, with density curves overlaid. The dashed lines represent the median of each distribution.

The responses' non-normal nature was confirmed by performing the Shapiro test of normality (Shapiro & Wilk, 1965) on each of the three items: the test proved significant in each case, confirming the non-normal distribution for each triplet (Table 2).

Table 2

Results of the Shapiro test of normality for each of the six triplets

	W	*p*
MT1	.83	< .001
MT2	.83	< .001
MT3	.83	< .001
FT1	.83	< .001
FT2	.83	< .001
FT3	.83	< .001

Note. $N = 566$

As visible in Figure 6 and reported in Table 2, the mean and median of the answers of certain triplets noticeably shift in the direction of the hypothesis. Particularly, this is the case for all male triplets. Due to the peculiar shape of the distributions, performing t-tests to falsify the null hypothesis would have been inadequate. It was preferred to use the Wilcoxon Test (Wilcoxon, 1992), also known as Mann-Whitney's test (Mann & Whitney, 1947), a non-parametric test commonly used to determine if two samples belong to the same population or two populations with different distributions when the assumption of normality of the data is not met. The test was performed for all triplets, confronting the

recorded data with a distribution with a median of 50. The one-tailed alternative hypothesis was that the medians were shifted towards the candidate whose face was morphed with the elector, and thus the true medians were higher than 50.

Although the effect size (*r*) was small, all three male triplets' medians were significantly shifted in the direction of the hypothesis, indicating that the participants in the study were more likely to attribute vote intention to the male candidate whom the elector was manipulated to resemble the most. Conversely, no significant effects were found for the triplets composed of female faces, suggesting a possible role of the electors' and the candidates' gender in determining the participants' vote attributions. Spearman's correlation tests were performed to assess whether Vote Attribution toward the manipulated candidate co-varied with the reported use of the Facial Similarity Criterion. As reported in Table 3, depending on the Triplet, the correlation varied between small and negligible and reached statistical significance only for triplets MT1, MT2, and FT2.

Table 3

Descriptive Statistics and Wilcoxon Test Results for the Six Experimental Triplets

	M	*Mdn*	*sd*	*r*	*rho* *Similarity*
MT1	55.31	70.5	39.47	.118**	.13**
MT2	57.06	75.2	39.05	.144***	.10*
MT3	54.64	70.0	39.62	.102**	.05
FT1	49.72	49.9	39.90	-.02	.03
FT2	51.68	65.4	39.94	.02	.09*
FT3	51.77	59.7	40.09	.02	.03

Note. $N = 566$. * = $p < .05$; ** = $p < .01$; *** = $p < .001$

To determine a possible reason for which certain triplets show weaker effects of the manipulation, the data were recoded to reveal potential tendencies to attribute vote intention to one candidate over the other, independently from the manipulation itself. As Table 4 shows, the participants tended to attribute higher vote intention to a candidate over the other in four of the six experimental triplets. Specifically, in triplet MT2, candidate MI9 was preferred to candidate MI6.

As reported in Table 4, a significant but extremely weak correlation was found between vote attribution toward MI9 and the reported use of the Warmth Criterion. Surprisingly, all three Female Triplets presented a small but significant correlation between the vote

toward one of the two candidates independently from the experimental manipulation and the reported use of the Facial Similarity criterion. The possible reasons behind these correlations will be discussed in the next chapter.

Table 4

Descriptive Statistics, Wilcoxon Test Results and Correlation with the Criteria of Warmth and Similarity for the Six Experimental Triplets Recoded

	M	*Mdn*	*sd*	*r*	*rho*	
					Warmth	*Similarity*
MT1-MI12	51.72	56.4	39.79	.005	-.04	.06
MT2-MI9	56.46	74.0	39.15	.130***	.08*	.04
MT3-MI7	58.63	77.6	38.95	.210***	.02	.03
FT1-WI8	53.19	68.8	39.77	.060	-.02	.10*
FT2-WI7	65.73	82.6	36.74	.370***	-.05	.15***
FT3-WI12	68.12	83.8	35.80	.404***	-.01	.11**

Note. The table shows descriptive statistics, Wilcoxon test results, and Spearman's correlation with the use of the criteria of warmth and similarity for the six recoded experimental triplets.

$N = 566$. * = $p < .05$; ** = $p < .01$; *** = $p < .001$

4.1.1 Response Time for each Triplet

As visible in Table 5, the average response time did not vary between the six experimental triplets. When performing a non-parametric-ANOVA (Friedman, 1937), the differences in the timing of the participants' first clicks do not reach statistical significance ($\chi^2 = 1.75$; $df = 5$; $p = .88$). Similar results emerged for the timing of the participants' last clicks across the six experimental triplets ($\chi^2 = 1.75$; $df = 5$; $p = .88$).

Across all triplets, Response Time was not predictive of Vote Attribution nor significantly correlated with any other variable. Subsequent paragraphs will thus not include Response Time as a possible predictor.

Table 5
Response Time for Each Triplet

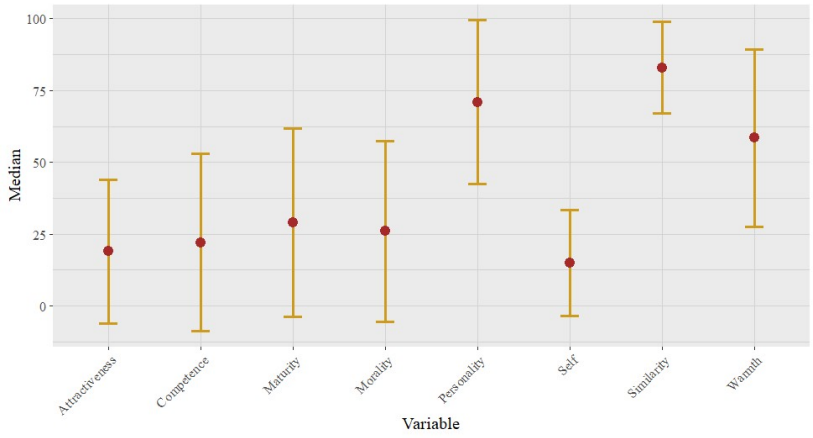
	M		*Mdn*	
	First Click	Last Click	First Click	Last Click
MT1	7.74	10.6	5.25	6.90
MT2	7.76	11.3	5.30	7.75
MT3	7.22	10.3	5.30	6.99
FT1	9.56	10.5	5.60	7.22
FT2	7.61	12.6	5.47	7.47
FT3	7.07	12.1	5.20	7.24

Note. Response times for each experimental triplet.

4.2 Analyzing the Reported Use of the Criteria

As described in the previous chapter, the second section of the questionnaire included eight questions regarding a set of possible criteria used by the participants while assigning the probability that the elector voted for one candidate over the other. Since most of the variables did not conform to a normal distribution, a non-parametric-ANOVA (Friedman, 1937) was performed to assess if there were significant differences in the use of the various criteria. The ANOVA indicated significant differences between the various criteria ($\chi^2 = 1175$; $df = 8$; $p < .001$), as visible in Figure 7.

Figure 7
The median reported use of the eight criteria.

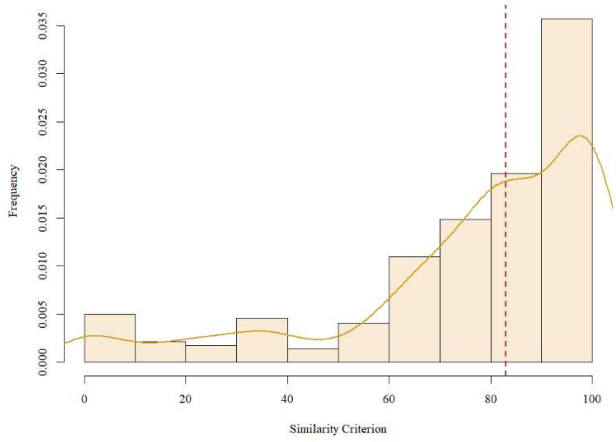


Note. The plot displays median values for and associated uncertainties for the eight criteria present in the dataset. Points represent medians, while vertical bars depict uncertainties.

In line with the first hypothesis, the most used criterion was Facial Similarity (Figure 8) between the elector and the candidate (M = 76.38, Mdn = 83, sd = 26.38). Since the data were not normally distributed, as determined by performing a Shapiro test (W = .82, p <.001), the possible effect of gender was tested by performing a one-way ANOVA on ranks (Kruskal & Wallis, 1952) instead of a regular one-way ANOVA. However, no significant differences were found between genders (χ^2 = 2.28, df = 2, p = .32).

Figure 8

Reported use of the facial similarity criterion

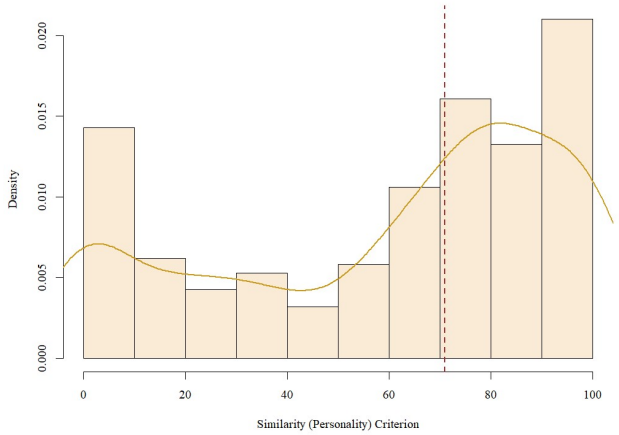


Note. Distribution of the facial similarity criterion variable in the dataset, with the density curve overlaid. The dashed line represents the median.

The second most used criterion was the similarity between the inferred personalities (Figure 9) of the elector and the candidate (M = 6.54, Mdn = 71, sd = 33.85), suggesting that the inferences of personality may have played an essential role in the participants' answers. No significant differences were found between genders (χ^2 = 0.95, df = 2, p = .62).

Figure 9

Reported use of the personality similarity criterion

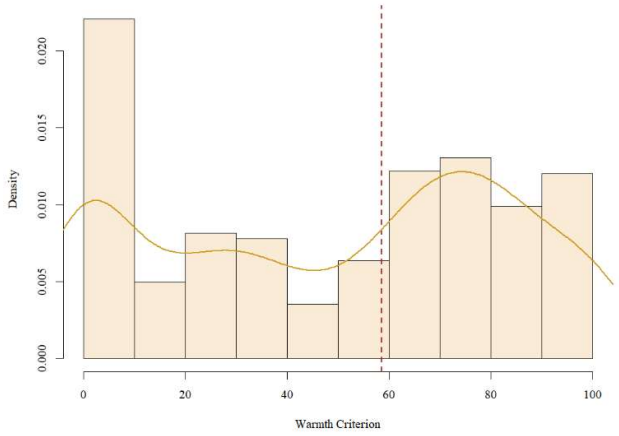


Note. Distribution of the personality similarity criterion variable in the dataset, with the density curve overlaid. The dashed line represents the median.

The third most used reported criterion was the candidates' perceived warmth (Figure 10; $M = 49.65$, $Mdn = 58.5$, $sd = 34.16$). Once again, no significant differences were found between genders ($\chi^2 = 3.21$, $df = 2$, $p = .21$).

Figure 10

Reported use of the candidates' warmth criterion



Note. Distribution of the candidates' warmth criterion variable in the dataset, with the density curve overlaid. The dashed line represents the median.

No other criterion exceeded a median of 50, the middle point of the scale in which the cursor was placed by default. From the most to the least used, the remaining criteria were the following: the candidates' perceived maturity ($M = 36.10$, $Mdn = 29$, $sd = 32.47$), their morality ($M = 34.06$, $Mdn = 26$, $sd = 32.89$), their perceived competence ($M = 33.30$, *Mdn*

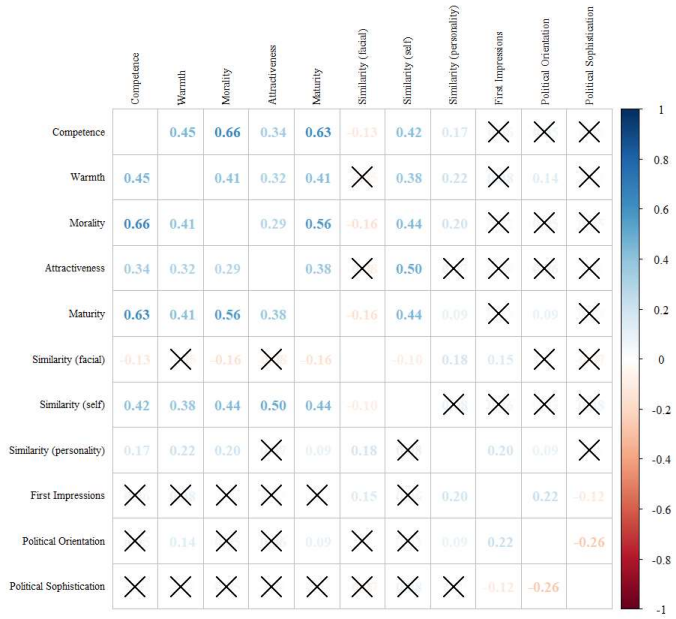
= 22, *sd* = 33.17), their attractiveness (*M* = 27.97, *Mdn* = 19, *sd* = 29.23), and lastly the candidates' similarity with the participant (*M* = 24.53, *Mdn* = 15, *sd* = 29.22).

4.2.1 Correlation Analysis Between the Criteria

To better comprehend the respondents' behavior, correlations were investigated between the eight reported criteria and other key variables, such as the participants' reported reliance on first impressions, political orientation, and political sophistication. The correlation matrix is reported in Figure 11.

Figure 11

Correlation plot between the eight criteria and the reported tendency to rely on first impressions



Note. Values colored in progressively more intense shades of blue indicate a positive correlation, while values colored in progressively more intense shades of red indicate negative correlations. The cells marked with an "X" indicate that the correlation is not significant.

As visible in Figure 11, the use of the Facial Similarity criterion showed a small but significant correlation with the reported score of how much the participants tend to rely on first impressions (*rho* = .148, $p < .001$), indicating that participants who state to rely on first impressions also tend to use the facial similarity criterion. No correlation was found between the reported use of this criterion and the political orientation of the participant (*rho* = -.047, *p* = .27). Similar to what was reported for the criterion of facial similarity, the personality similarity criterion showed a weak but significant correlation

with the reported tendency to rely on first impressions (*rho* = .197, $p < .001$). A significant correlation was found between the use of this criterion and political orientation. However, the strength of this correlation was extremely weak (*rho* = .085, $p = .04$) and did not indicate substantial differences between progressive and conservative participants. Surprisingly, a weak, positive correlation was found between the use of the third most used criterion, the candidates' perceived Warmth, and political orientation (*rho* = .138, $p < .001$), indicating that conservative-leaning participants might have been more prone to use the candidates' warmth as a criterion. A slightly stronger correlation was found between the use of the Warmth criterion and the respondent's age (*rho* = .234, $p < .001$), indicating that older participants were more likely to rely on this criterion.

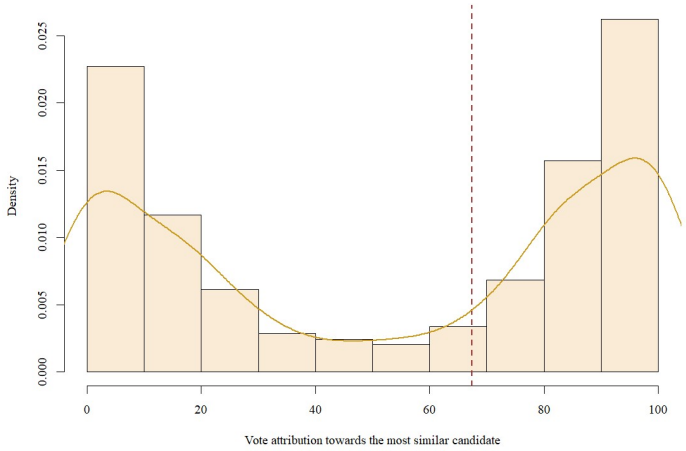
As visible in Figure 11, the five least-used criteria all show significant correlations with each other, with the strongest correlation being between the use of the criteria of competence and morality (*rho* = .66, $p < .001$) and between competence and maturity (*rho* = .63, $p < .001$). On the other hand, the correlations between these criteria and facial similarity are consistently weak and negative. Similarly, the correlation between the reported use of warmth and the reported use of facial similarity is negative and not significant (*rho* = -.04, $p = .26$). This consistently negative or non-significant correlation between the reported use of facial similarity and the reported use of the inferences of competence, warmth, morality, attractiveness, and maturity seem to indicate that people who relied on similarity were less likely to use the candidates' inferred traits as criteria to attribute the elector's probability to vote for them.

4.3 Analyzing the Stacked Data

To better analyze the effect of various variables on the vote attributions of the participants, the data were stacked, creating a new dataset with a variable "Vote Attribution" reporting the attributed values of all six experimental triplets (Figure 12) and a variable "Triplet" reporting the specific triplet for which the attribution was made. An additional variable, "Triplet's Gender" was added due to the hypothesized importance of the faces' gender.

Figure 12

Histogram of the vote attribution variable across all six experimental triplets



Note. Distribution of the vote attribution variable in the stacked dataset, with density curve overlaid. The dashed line represents the median.

As visible in Figure 12, the distribution of the Vote Attribution variable presented a slight shift in the direction of the hypothesis (M = 53.36, Mdn = 67.4, sd = 39.72). Overall, the cursor had been moved toward the direction of the most similar candidate 1840 times on a total of 3396 observations, indicating that the participants preferred the most similar candidate to the elector at an above-chance level (P = 54.18%, p <.001), as assessed by performing a proportion test after recoding the data to fit a binomial distribution (*value* < 50.1 = 0, *value* > 50 = 1).

The validity of hypothesis 1 was confirmed through a rigorous Wilcoxon test on the stacked data of vote attribution. The alternative hypothesis, that the values would have a median higher than 50, was proven significant (p < .001). However, the effect size was found to be minimal (r = .06), a factor that is generally considered negligible. Despite this, further analyses were conducted to uncover the potential roles of key variables. For all subsequent analyses, the vote attribution variable was recoded to fit the criteria needed to perform beta regressions, resulting in a distribution between .001 and .999 instead of between 0 and 100. The non-normal distribution of the vote attribution variable led to the determination that linear regression models, which assume a normal distribution of their output, were unable to capture the complexity of the data. Instead, beta regression models were deemed more appropriate because they are specifically designed to handle variables that are bounded between 0 and 1, such as proportions or probabilities (Cribari-Neto & Zeileis, 2010). Furthermore, models of beta regression can accommodate the heteroscedasticity and skewness often present in such data that also characterizes the vote

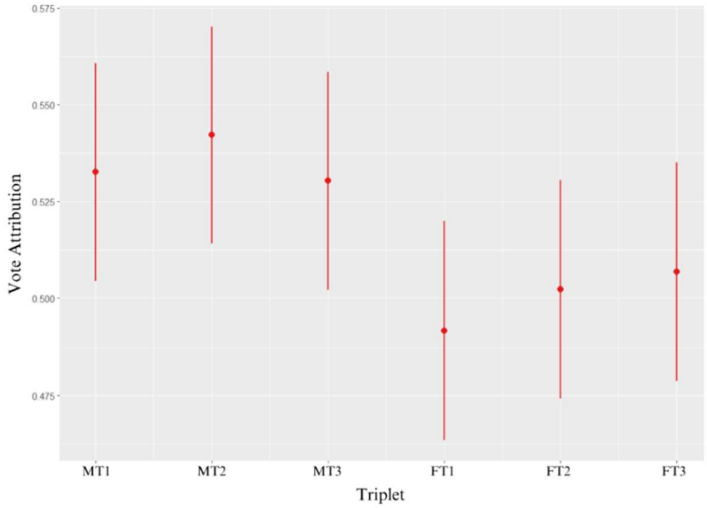
attribution variable, which was not normally distributed, as confirmed by performing the Shapiro test (W = .83, p < .001). By using beta regression, the analysis can more accurately capture the nuances of the vote attribution variable's distribution, leading to more reliable and valid inferences.

4.3.1 The role of the triplets' gender

Firstly, a multi-level regression model (Figure 13) was constructed to predict the vote attribution toward the most similar candidate based on the triplet, while accounting for the potential correlation within each participant (Fit1). A random intercept was predicted for each participant to capture individual differences in baseline vote attribution. This multi-level component of the model, although it will not be mentioned in the next paragraphs, will be maintained for each subsequent model. The model summary revealed that FT1 had a significantly different intercept than all other triplets (z = -2.01, p = .04). Although the difference was not significant for FT2 (z = -1.49, p = .13) and FT3 (z = -1.26, p = .21), as visible in Figure 13, the female triplets tended to have lower intercepts than the male triplets.

Figure 13

Marginal effect of Triplet levels on Vote Attribution



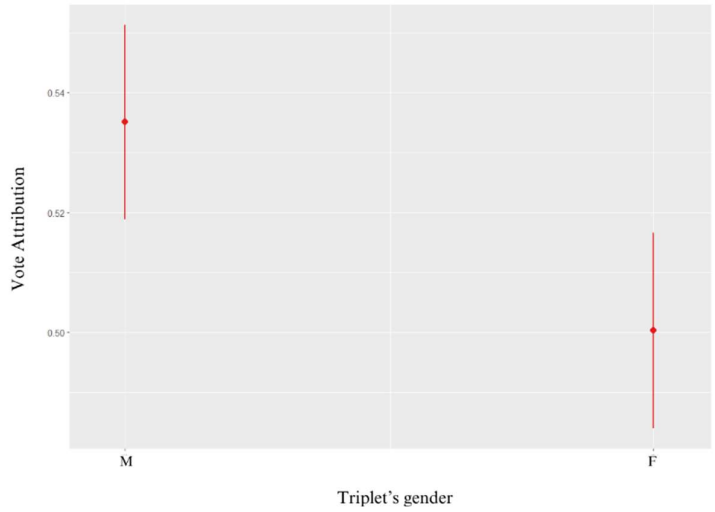
Note. Predicted values of vote attribution are lower for the female triplets (FT1, FT2, FT3) than for the male triplets (MT1, MT2, MT3).

A second model of beta-regression (Fit2; Figure 14) was constructed to test the effect of the gender of the candidates and the elector depicted in a triplet on the participants' vote attribution. The model attested a significant difference between the two gender conditions. In particular, the participants showed a higher tendency to attribute vote

intention toward the target candidate when the experimental triplets were composed of male faces than when they were composed of female faces (z = -2.96, p = .003). Similarly, when recoding the values of Vote Attribution to fit a binomial distribution, Male Triplets presented a higher probability of the participant moving the cursor toward the most similar candidate (P_M= 56.61%, P_F= 51.55%, p = .003).

Figure 14

The marginal effect of Triplet's gender levels on Vote Attribution



Note. Predicted values of Vote Attribution are higher for male triplets than female ones.

4.3.2 The role of the participant's gender

Since the scientific literature attests that women tend to perform better in tasks of facial recognition (Rehnman, 2007; Herlits & Lovén, 2013), a third model (Fit3) was constructed to test the hypothesis that women would also more likely attribute a higher vote intention to the candidate that resembled the elector the most. The model did not show any significant effect of the participant's gender on the variable Vote Attribution (z = 1.13, p = .25), indicating that men and women attributed vote intention largely in the same way.

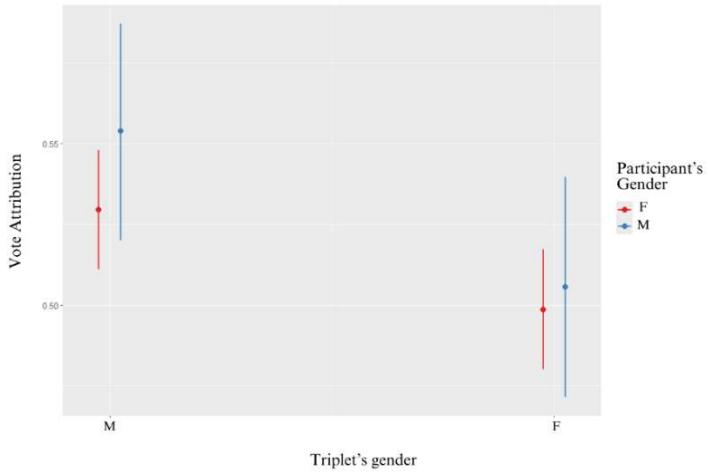
Nonetheless, since the scientific literature reports that people tend to recognize better faces of people of their gender (Herlits & Lovén, 2013), an additional model (Fit4) was tested to assess the possible interactions between the participants' and the triplets' gender. Fit4 was constructed as a nested model of Fit2, maintaining the variable Triplet's Gender as its first predictor and adding the Participant's Gender as a second predictor in interaction with the first.

As visible in Figure 15, the predicted values of Vote Attribution are generally similar between genders. Only when the participant is male and the triplet is composed of male faces, does the model presents a significant interaction between the Participant's and the Triplet's gender (z = 2.30, p = .02). This suggests that, when compared to female participants, male participants were more likely to attribute higher vote intention to the candidate whose photo was used in the morphing process which increased the similarity between the elector and a candidate.

Surprisingly, the model does not support the existence of a corresponding effect for female participants when evaluating triplets composed of female faces (z = .62, p = .53). On the contrary, male and female participants alike appear to rely less on facial similarity when attributing vote intentions of female electors for female candidates. Nearly identical results are found when recoding the values of Vote Attribution to fit a binomial distribution. In this case, the probability of a male participant moving the cursor in the direction of the hypothesis is significantly higher than the probability of a female participant doing the same (P_M= 57.73%, P_F= 55.87%, p = .02). The difference between participants of opposite gender practically disappears when evaluating triplets of faces belonging to the female gender (P_M= 51.82%, P_F= 51.47%, p = .58).

Figure 15

Predicted values of Vote Attribution in the Fit4 model



Note. As visible in the Figure, the participant's gender is treated as a binary variable. All the models constructed, discussed, and compared in paragraph 4.3 exclude the data from the six non-binary participants because their number is excessively small for the effect of this class to be effectively compared to those of the two binary genders.

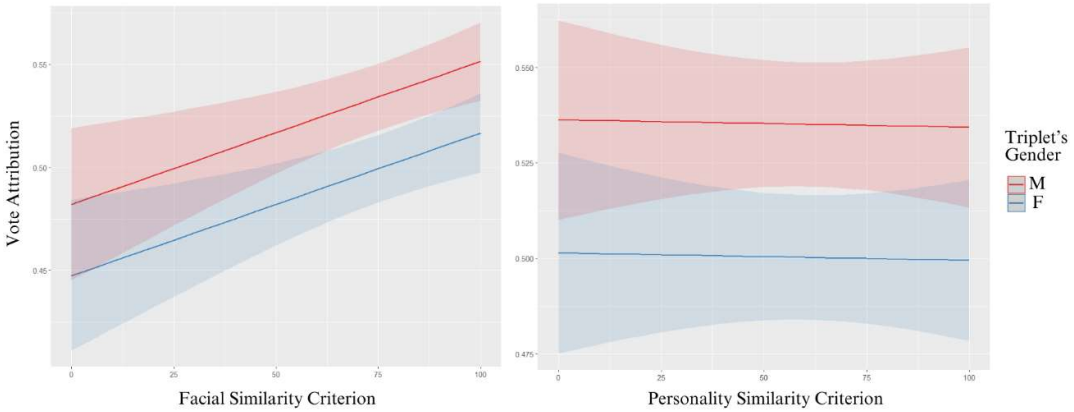
4.3.3 The role of the reported Criteria

As previously examined in paragraph 4.1, the participants reported using similarity criteria, specifically facial and personality similarity between the candidate and the elector, more than any other criterion. Thus, the models Fit5 and Fit6 were constructed to respectively include the criterion of facial similarity and personality similarity as predictors of Vote Attribution. Although, as previously analyzed, the two criteria presented a weak, positive correlation (*rho* = .18, $p < .001$) with each other, the two models were widely different. As visible in Figure 16, a higher reported use of the Facial Similarity criterion predicted higher Vote Attribution values. Moreover, the relation between the two variables in the model's summary was significant ($z = 3.12$, $p = .002$). On the contrary, the reported use of the criterion of Personality Similarity did not significantly predict Vote Attribution ($z = -.11$, $p = .91$). These data seem to indicate that people are at least partially aware of the role played by physical similarity specifically in their judgments. They seem to anchor their electoral attributions on physical traits more than on inferred personality traits.

Nonetheless, the explicative power of the model Fit5 and the criterion of Facial Similarity remains low. Firstly, although positive and significant, the correlation between the assessed use of the criterion and the values of vote attribution is extremely weak (*rho* = .07, $p < .001$). Moreover, Efron's pseudo-R-squared (Efron, 1978), a statistic used to evaluate the goodness-of-fit for regression models commonly used in logistic regression, is extremely low for Fit5 (*Efron's* $R^2 = .005$), indicating that the reported use of the criterion only explains a minimal fraction of the data's variability.

Figure 16

Predictions of the models Fit5 and Fit6



Note. Comparison between the predictions of the models Fit5 (on the left) and Fit6 (on the right).

An additional model, Fit7, was constructed to account for a possible interaction between the use of the criterion of Facial Similarity and the Triplet's gender. Such interaction, however, proved to be non-significant (z = .42, p = .67), and the model predicted results remarkably similar to Fit5, with a nearly identical pseudo-R-squared (*Efron's* R^2 = .005). Similarly, no significant interaction emerged between the use of the Personality Similarity criterion and the Triplet's gender (z = .02, p = .98) in model Fit8.

4.3.4 Other models

Other models were constructed to explore the possible predictive roles of different variables. The model Fit9 tested the hypothesis that people who reported relying more strongly on first impressions would also attribute higher vote intention to the candidate resembling the elector the most. However, the predictive role of the reported reliance on First Impressions proved non-significant (z = -.27, p = .78). Similarly, the model Fit9 tested the possible effect of political orientation, which also proved non-significant (z = -.57, p = .56). Lastly, the variable age was tested as a possible predictor but once again proved non-significant (z = -1.49, p = .15).

4.3.5 Model selection

To select the best model, the models were compared for the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Efron's pseudo-R-squared (Table 6). Overall, Fit5 can be considered the best model, having the lowest AIC and BIC of all models. Consequently, the reported use of the facial similarity criterion seems to be the variable most predictive of Vote Attribution. Nonetheless, as represented by the extremely low value of Efron's R-squared across all models, this variable is largely unable to explain the variability of the responses. The possible reasons behind the low explanative power of the models constructed hitherto will be discussed in the following chapter.

Table 6

Comparison between the Beta-regression Models

	AIC	*BIC*	*Efron's* R^2
Fit0	-9557.479	-9539.120	-.001
Fit1	-9557.233	-9508.276	.002
Fit2	-9564.262	-9539.783	.001
Fit3	-9556.756	-9532.277	-.001
Fit4	-9561.928	-9525.209	.002
Fit5	-9572.046	-9541.448	.005
Fit6	-9562.275	-9531.676	.001
Fit7	-957.226	-9533.508	.006
Fit8	-956.275	-9523.557	.001
Fit9	-9555.554	-9531.075	-.002
Fit10	-9555.805	-9531.327	-.001
Fit11	-9557.543	-9533.065	.000

Note. Efron's pseudo-R-squared can sometimes assume negative values.

Chapter V – Discussion

5.1 Main results

The main aim of the current research was to verify whether similarity would be used as a criterion of vote attribution even in the absence of political cues, such as the candidates' political orientation, while controlling other confounding similarity variables, such as the candidates' and the electors' genders, ages, and ethnicities.

Overall, the data collected seem to support the study's main hypothesis (H1): the participants tended to attribute a fictitious elector a greater probability to vote for the candidate manipulated to look the most similar to the elector themselves. As analyzed in the previous chapter the effect was, although significant, small. However, this effect size must be considered in the context of the manipulation, which was minimal and comparable in size with random, naturally occurring similarities.

The fact that the main hypothesis was confirmed integrates the results from Rule & Ambady (2010) and Bailenson et al. (2006; 2008). On one hand, as reported by Rule & Ambady (2010), facial similarity may be diagnostic of ideological similarities, hence the ability of people to predict a person's political orientation by only assessing their facial traits. However, it is essential to remember that despite conservative and progressive people showing facial differences sufficient to differentiate them, this does not necessarily imply a biological base for political orientation. In addition, it is essential to highlight that facial differences between people of different ideologies may be mediated by their personalities. As examined in Chapter I, a neutral face's resemblance to a specific emotional expression (Adams Jr. et al., 2012; Jaeger & Jones, 2022) seems to predict how warm and competent it would be judged to be. Moreover, faces judged to be warmer are also more likely to belong to more progressive people (Rule & Ambady, 2010). In this case, a neutral face's resemblance to an emotional display is not necessarily biologically determined and instead is possibly due to micro-expressions that seem to honestly signal a person's stable emotional dispositions (Adams Jr. et al., 2016), and thus their personality, which in turn is complexly determined by a mixture of biological and psychosocial factors. Whether it is biologically or socially determined, the fact remains that people's naïve theories seem to account for the relation between facial and ideological similarity and thus seem to be able to decisively guess for which candidate an elector voted in a matter of a few seconds.

The importance of facial similarity was confirmed by the participants themselves, who reported the Facial Similarity Criterion as the most used in guiding their vote attribution processes. Not only that, but as demonstrated in the model selection process discussed in paragraph 4.3.5, Facial Similarity also resulted in being the most predictive criterion of the participants' vote attribution. Moreso, in the triplets in which the participants systematically preferred a candidate over the other independently from the manipulation, Vote Attribution towards the most voted candidate was positively and significantly correlated with the reported use of the criterion of Facial Similarity, possibly indicating that the candidate was preferred due to their higher baseline resemblance with the elector. These results suggest that the participants are generally aware of using facial similarity as a criterion to predict the vote intention of other people.

On the other hand, it is also interesting to note how the criterion of Personality Similarity, which is reportedly also widely used, is not correlated at all with the variable vote attribution (*rho* = 0.01, *p* =.49). This may indicate that, while the participants reported that Personality Similarity constituted a critical criterion, the complex process of inferring the electors' personality traits and then comparing them to the ones extrapolated from the candidates may have been heuristically substituted (Kahneman & Frederick, 2002) with the much simpler and quicker process of just comparing the three people's facial appearances. Heuristic substitution may also constitute an essential key to interpreting the effect of similarity in political choices. As reported by Todorov et al. (2005), inferences from faces are most likely a product of the fast, unreflective, and effortless "System 1" (Kahneman, 2013), which aims to produce quick, simple, and acceptable answers to complex problems without dispending excessive energy. Electoral decisions are complex and adequately processed would require access to a vast amount of information. For instance, to appropriately evaluate a candidate's competence, an elector should research the candidate's educational background, experience as an administrator, the feasibility of the program, and its appropriateness to face the current political issues. To properly evaluate all these dimensions, an elector should use the slow, logical, and energy-expensive "System 2" (Kahneman, 2013). However, people often lack the time and motivation to activate this system and thus tend to rely on the heuristic processes of System 1. Since the process of extracting information from faces has evolved to occur automatically, facial inferences often offer an easy alternative to much more complex, and often impossible, judgments, thus constituting an example of the heuristic of attribute substitution (Kahneman & Frederick, 2002).

The process of vote intention attribution to another, previously unknown person may follow a similar pattern. To appropriately evaluate a person's political orientation, an observer should have access to a lot of difficultly available information, such as their socioeconomic background, religious and ethical beliefs, *et cetera*. Hence, when asked to guess a person's political orientation, people tend to rely on surprisingly accurate heuristics, such as the heuristic of representativeness (Kahneman & Tversky, 1974). As reported by Olivola et al. (2012), people tend to substitute an extremely difficult question ("What party did this person vote for?") with a much simpler one ("Does this person resemble a Republican or a Democrat more?"), which answer is immediately evident when comparing their facial appearance to the stereotypical prototype of a conservative or a progressive person. In the case of the current study, a similar heuristic substitution seems to have taken place. In the absence of sufficient information about the candidates' political stances and the elector's beliefs, the participants tended to answer a much simpler question: "Which of the two candidates resembles the elector the most?".

As discussed, the use of heuristics in the field of political choices is commonly found throughout the scientific literature. Furthermore, the implementation of such strategies may be favored by the artificial context of most experiments. Although people may use heuristics to decide who to vote for and to assess the political orientation of other people, experiments often remove alternative information sources and pose some time constraints, thus favoring the activation of a "System 1" mode of thinking. Moreover, people may be less motivated to waste cognitive energy on an experimental task than in real-life situations.

5.1.1 Alternative Strategies of Vote Attribution

Although Hypothesis 1 was confirmed, it should be noted that participants did not report to exclusively use Similarity as a Vote Attribution criterion. Warmth, for example, was also assessed to be an important criterion by most participants, even if it did not show a significant correlation with the participants' answers. The reported importance of this variable specifically, which was often chosen over competence, a predictor of vote intention considered more important across the scientific literature (e.g., Todorov et al., 2005), may be due to its salience in the stimuli used. As described and motivated in Chapter III, all the faces implemented in the experiment showed a smiling facial expression. The presence of smiles across all the experimental faces may have motivated the participants to feel that warmth should have been an important factor to consider when

making their vote attribution choices. However, the substantial absence of significant correlations of any kind between the reported use of the variable warmth and the variously recoded data seems to suggest that, although participants felt that Warmth should have played an important role, and perhaps guessed that it could have been part of the hypotheses of the study, they did not effectively use the criterion.

The other, less used criteria, as indicated by the non-significant or negative correlations with the criteria of similarity already discussed in paragraph 4.2.1, seem to be used as a possible alternative for judging vote attribution, possibly by people who found it more difficult to evaluate facial similarities between the elector and the candidates. However, the data are insufficient to prove definitively that using other criteria of vote attribution correlates with a lower ability to judge similarity. Hence, it only remains a hypothesis.

5.1.2 The Role of Gender

The analysis revealed complex and occasionally unexpected interactions between the gender of the photographs and that of the participants. Contrary to Hypothesis 1', female participants did not attribute a higher probability of an elector voting for the candidate with whom the elector's picture was morphed. This suggests that the experimental manipulation affected both male and female participants equally, despite women typically performing better in face recognition tasks (Rehnman, 2007). Consequently, the data do not support Hypothesis 2.

The discrepancy between the hypothesized effect and the data may stem from previously unconsidered differences between the experimental task and standard face recognition tasks. Face recognition typically engages short- or long-term memory, requiring the ability to distinguish a previously observed face from others in various contexts. Women tend to remember faces better than men because they spend more time observing them and focusing on key features such as the mouth, nose, and eyes (Heisz et al., 2013). In contrast, the experimental task did not necessitate memorizing facial features beyond working memory. In the experiment, the candidates' and elector's faces were displayed simultaneously and remained on screen until the page was submitted, making the ability to memorize faces unnecessary when comparing similarities.

An interesting effect emerged from the interaction between the triplets' and participants' gender. The own-gender bias (Herlits & Lovén, 2013) was evident only in male participants, who were more likely to attribute vote intention to the candidate resembling the elector in the case of male triplets. This aligns with existing literature, indicating that

it is easier to spot similarities within ingroup faces than outgroup faces (Herlits & Lovén, 2013; O'Toole et al., 1994). Conversely, both men and women performed worse when judging feminine faces. Thus, Hypothesis 3 has only partially found support in the data.

The absence of own-gender bias in women can be attributed to the lower tendency of men and women alike to attribute vote intention to the most similar candidate in female triplets. Since the similarity effect on vote attribution was nearly absent for female triplets, vote attribution was practically random. Therefore, the experimental manipulation seems to not have affected female triplets, and the lack of own-gender bias in female participants could be a consequence. Overall, the evidence supports Hypothesis 4b rather than the opposing Hypothesis 4a, which predicted no differences in vote attribution towards the most similar candidate between male and female triplets. This effect is partially in line with the findings of Bailenson et al. (2006), who reported that morphing a male candidate's face with the face of a female participant produced a boomerang effect, lowering the participant's vote intention towards the candidate. However, in the original study, the morphing was effectuated between photographs of different genders, while in the current study, the morphing only involved faces of the same gender.

There are multiple possible reasons behind the absence of effect for female triplets in the previously analyzed data.

Assuming that the morphing procedure is equally effective in producing similarity for male and female faces, it is possible to offer other alternative explanations for the absence of the effect on Vote Attribution for female triplets in the main questionnaire. For instance, as evidenced in Table 4 (Paragraph 4.1) and visible in Table 11, the participants showed a preference for one face over the other in all three female triplets. The vote attribution toward the participants' favorite face was also significantly correlated in all three cases with the use of the criterion of perceived facial similarity.

Table 7

Preferred Candidate's faces across the three Female Triplets

Triplets' code	Elector's face	Preferred Candidate	Other Candidate
FT1	 WI5	 WI2	 WI8
FT2	 WI9	 WI7	 WI11
FT3	 WI1	 WI12	 WI10

Note. The entity of vote attribution preference is reported in Table 4.

The idea that, independently from the experimental manipulation, WI2, WI7, and WI12 may resemble their elector more than the other candidate is also supported by the data collected in the manipulation check questionnaire, which consistently showed WI2, WI7, and WI12 as the faces which were more likely to be selected as the one resembling the elector the most. However, the variable vote attribution significantly shifted only for the triplets FT2 and FT3; thus, only WI7 and WI12 can be considered similar to their respective electors. It is thus possible to hypothesize that when a candidate already significantly resembles the elector more than the other candidate does, such similarity may lead to more robust vote attribution toward the most similar candidate, and, in turn, this may mask the effect of the morphing manipulation.

As reported in Table 10 (Paragraph 4.4.3) and visible in Table 8, triplets MT1 and MT2 also presented a candidate significantly more similar to the elector (MI12 and MI9, respectively). However, in both cases, the effect of the morphing manipulation was still detectable.

Table 8

Preferred Candidate's faces across the two Male Faces showing the effect

Triplets' code	Elector's face	Preferred Candidate	Other Candidate
MT1	 MI11	 MI12	 MI3
MT2	 MI2	 MI9	 MI6

Note. The entity of similarity attribution preference is reported in Table 10.

This difference between the two male triplets showing similarity independently from the manipulation and the three female triplets is fascinating since it poses the question of why similarity again shows different effects when the triplets are composed of either male or female photographs.

Since the morphing procedure slightly alters the shape but especially the position of the main facial features, the procedure may enhance similarity by making the configurational properties, or second-order features (Rhodes, 2013), of the morphed face more similar to the target face. Bernard et al. (2019) highlight that individuals tend to perceive them through a less holistic and more analytical approach, particularly when feminine faces have heavy makeup applied. The authors argue that this effect is due to the fact that women wearing makeup tend to be sexualized and, as a result, more cognitively objectified by the observer (Bernard et al., 2018). In the case of the previously analyzed data, the fact that both exclusively male and exclusively female triplets were subsequently shown to each participant may have elicited forms of inter-group thinking and prejudices against women, specifically against women in politics. As broadly examined throughout the present dissertation, despite women actively participating in democratic politics for over a century in large portions of the Western World, women are still generally seen as less capable leaders than men, and female candidates are still judged through the lenses of gender stereotypes, being typically seen as more capable only on issues traditionally associated with femininity, such as education and health (Anzia & Bernhard, 2022).

Smiling could also have contributed to the women's photographs being objectified: an extensive line of research stemming from the work of Henley (1973) supports the idea that when women smile, they are perceived as more accommodating, friendly, approachable, and conforming to traditional gender roles, and are thus more objectified (Deutsch, 1990). If these factors led the participants to objectify the women candidates, the participants could have been led to elaborate women's faces in a less configurational and more analytical way. This kind of elaboration is considered insufficient to adequately perform similarity judgments, such as kin recognition tasks (Bicego & Grosso, 2019).

In other words, it is possible that, due to the participants perceiving the faces of the female candidates as more objectified, they could have ignored the configurational similarities between the elector and the candidate. However, other kinds of similarity, possibly relying on more simple, analytical, and superficial traits, may have been used, thus explaining the preference for one candidate over the other independently from the manipulation. For instance, as visible in Table 7, in the Triplet FT1 candidate WI2 could have been preferred due to her face having a more similar height-to-width ratio to the elector than the other candidate, WI8. Similarly, and perhaps most notably, in the triplet FT3 candidate WI12 may have been preferred to candidate WI10 because of her slightly darker skin tone, more similar to the elector, WI1.

Although fascinating, the idea that configurational facial similarities played a more important role in the vote attribution process for male triplets, while more superficial, analytical facial similarities were primarily used in the process of vote attribution for female triplets is undemonstrable given the currently available data.

Perhaps more simply, the female faces could have been matched poorly, with too strong naturally occurring similarities between the elector and certain candidates. If that were the case, the absence of the experimental manipulation's effect could have been due to the specific faces used rather than their feminine gender. Further research is necessary to dispel these doubts and improve our understanding of the phenomenon.

It is possible to hypothesize that the morphing procedure may be less effective in producing detectable similarities between faces when applied to women's faces. To verify this hypothesis, an additional study was performed, the results of which are further discussed in Appendix II.

5.2 Limitations

The current study presents a series of important limitations. Firstly, the use of artificially generated faces constitutes a problem that should not be underestimated for this and possibly for future studies. The artificial nature of the faces may have reduced the effects of the experimental manipulation since the artificial faces used could have exhibited a reduced variability when compared to natural human faces. In other words, the AI may have created faces that are more similar to each other by following the same recurring patterns when generating them. In turn, this may have led to a smaller effect of the experimental manipulation, which may have been masked by pre-existing similarities between faces.

Moreover, artificially generated photographs carried over into the study certain biases of the A.I., such as its tendency to portray people, specifically women, smiling. On the contrary, scientific research on the topic prefers to use faces with neutral expressions. Future studies should aim to replicate the current findings using more ecologically valid stimuli, such as the faces from the Chicago Face Database (Ma et al., 2015; 2021).

A second important limitation is constituted by the unrepresentativeness of the main sample, which largely underrepresents elder people, conservatives, and, perhaps most importantly, men. To better understand the role of the participant's gender, future studies should balance better the sample between the two binary genders, as well as include enough non-binary people to be able to make inferences about other gender categories. The sample was also, as expected, overwhelmingly composed of ethnically European people. Due to the significant lack of non-white participants, we could not study the potential effects of the other-race effect on inferences of vote attribution. Future studies may benefit from using a more ethnically diverse sample to analyze the role of ethnicity in these types of inferences.

Thirdly, certain variables may have been measured in a sub-optimal way. Given the distinctively non-normal nature of the distributions of the vote attribution variables, it is possible to infer that the participants practically refused to communicate uncertainty and almost always decisively select one candidate over the other. Presenting vote attribution as a binary choice rather than a continuum could have simplified the participants' answering process, simplified the analysis, and allowed the creation of more reliable models. Moreover, although the section on the criteria was overall very informative, as it offered insight into the participants' attribution processes, the questions were presented

in a way that prevented the participants from reporting possible differences in the use of the criteria between the male and the female triplets. The idea of measuring the use of the criteria as a continuum on a slider could have made the answering process excessively difficult and could have led to unnecessary random variability in the answers. The use of Likert scales may have led to more informative results.

Lastly, although the low number of triplets used was effective in making the questionnaire quick, thus making it more accessible to a greater sample of people, it also constituted an important limitation. With only three triplets for each binary gender, it is difficult to make definitive inferences on the effect of the photographs' gender, since the variability could be attributed to the unique characteristics of each triplet. To enable more reliable comparisons between genders, future studies on this topic should use a larger number of triplets, even at the cost of slightly reducing the sample size.

5.3 Conclusions and Future prospectives

The current research has contributed to expanding the comprehension of the role played by facial similarity in the field of politics. This topic seems to still be underexplored in the scientific literature. The findings from the present study suggest that vote attribution based on facial cues is possible, even without clear ideological information about the candidates, thus contributing to this line of research. These results integrate the previous literature, which only reported that facial similarity was predictive of a person's vote intention (Bailenson et al., 2006; 2008) and that facial cues could be used to infer a person's political orientation from their face (Rule & Ambady, 2010). Although a person's political orientation generally corresponds with their political intention in a strong two-party system such as the one present in the United States, it is still unclear if inferences from faces may be able to accurately infer vote intention in a more complex, multi-party system such as the one present in most European countries. Since the current study supports the hypothesis that people use the criterion of facial similarity, it is possible also to hypothesize that people should also be able to guess a person's vote intention by relying on facial similarity inferences even in a multi-polar political system. Even though the current study was not designed to analyze the potential diagnostic value of vote inferences from faces, future research might pursue this direction, trying to verify whether the vote attribution inferences made from faces are significantly predictive of the actual vote intentions of an elector. Studies on this topic would be particularly interesting if they focused on multi-polar or primary elections, where candidates are generally ideologically

aligned. In such contexts, the similarity between the voter and the candidate may play a crucial role that should be further explored.

Another essential finding of the current study regards the role of gender in inferences of similarity. Further research on this topic should primarily aim to replicate the current study's findings, as the effect of similarity manipulations on vote attribution appears weaker for female candidates compared to male candidates. If the current results are confirmed, verifying whether a similar effect is present in other fields would be interesting. The substantial absence of the similarity effect for female faces in this particular domain may stem from the association of politics with masculinity, which may, in turn, make the feminine faces perceived as less relevant in the context. It is possible that fields that are perceived as less stereotypically masculine could exhibit even more pronounced effects of facial similarity, particularly for feminine faces. For example, suppose the absence of the similarity effect among female faces is genuinely unique to the political realm or at least to traditionally masculine fields. In that case, the effect should manifest when participants are questioned about a less gendered domain, such as friendship. For example, in a scenario in which participants were asked the question, "Which of these two women is more likely to be friends with the woman depicted above?" considering that facial resemblance has been demonstrated to enhance trust (DeBruine, 2002) and cooperation (Giang et al., 2012), it could hypothesize that people would be more inclined to select the face most similar to the target face as the "friend."

Further research should also focus on the possible interactions between gender prototypicality and similarity effects, particularly on the effects on both perceived similarity and vote intention of morphings between faces of different genders. It is still unclear, in fact, why the original study from Bailenson et al. (2006) found a peculiar, negative effect of enhanced similarity on vote intention towards male candidates who had been morphed with female participants. Deepening the scientific understanding of gender prototypicality in the field of politics and of its different effects on male, female, and gender non-conforming candidates is nowadays as urgent as ever. The last twenty years have witnessed the progressive emergence of women as political leaders in European politics, with notable figures such as Angela Merkel, the former German Chancellor; Ursula von der Leyen, the current President of the European Commission; Marine Le Pen, the leader of the most voted party in France during the last European Elections; and Giorgia Meloni, the Italian Prime Minister. While the presence of these figures has undoubtedly contributed to a more balanced gender representation in politics and may

have inspired more women to engage in the political process (Ladam et al., 2018), there remains a need to explore why women appear to attain leadership positions and political success more frequently in conservative-leaning parties than in progressive-leaning ones. A deeper understanding of facial inferences, similarity inferences, and the role of gender is therefore crucial to establishing a scientific foundation upon which greater gender equality and more effective representation in politics can be built.

Bibliography

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: findings from a prospective study. *Journal Of Personality And Social Psychology*, *85*(4), 768.

Adams Jr, R. B., Garrido, C. O., Albohn, D. N., Hess, U., & Kleck, R. E. (2016). What facial appearance reveals over time: When perceived expressions in neutral faces reveal stable emotion dispositions. *Frontiers in Psychology*, 7, 986.

Adams Jr, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation?. *Cognition & Emotion*, *26*(3), 431-441.

Addison, W. E. (1989). Beardedness as a factor in perceived masculinity. *Perceptual and Motor Skills*, *68*(3), 921-922.

Agrawal, A. (2001). Kin recognition and the evolution of altruism. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1471), 1099-1104.

Anzia, S. F., & Bernhard, R. (2022). Gender stereotyping and the electoral success of women candidates: New evidence from local elections in the United States. *British Journal of Political Science*, 52(4), 1544-1563.

Atkinson, M. D., Enos, R. D., & Hill, S. J. (2009). Candidate faces and election outcomes: Is the face–vote correlation caused by candidate selection?. *Quarterly Journal of Political Science*, *4*(3), 229–249.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal Of Personality And Social Psychology*, *71*(2), 230.

Barrett, L. F. (2011). Was Darwin wrong about emotional expressions?. *Current Directions in Psychological Science*, *20*(6), 400-406.

Bateson, P. (1980). Optimal outbreeding and the development of sexual preferences in Japanese quail. *Zeitschrift für Tierpsychologie*, *53*(3), 231-244.

Bateson, P. (1983). Optimal outbreeding. *Mate Choice*, pp. *257*, 277.

Bell, L., Duchaine, B., & Susilo, T. (2023). Dissociations between face identity and face expression processing in developmental prosopagnosia. *Cognition*, *238*, 105469.

Bernard, P., Geelhand, P., & Servais, L. (2019). The face of sexualization: Faces wearing makeup are processed less configurally than faces without makeup. *International Review of Social Psychology*, *32*(1), 16.

Bernard, P., Gervais, S. J., & Klein, O. (2018). Objectifying objectification: When and why people are cognitively reduced to their parts akin to objects. *European Review of Social Psychology*, 29(1), 82-121.

Besco, R. (2019). *Identities and Interests: Race, Ethnicity, and Affinity Voting*. UBC Press.

Bicego, M., & Grosso, E. (2019). On the importance of local and global analysis in the judgment of similarity and dissimilarity of faces. *Image and Vision Computing*, *92*, 103813.

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *Journal of Neuroscience*, *31*(28), 10206-10214.

Brians, C. L. (2005). Women for women? Gender and party bias in voting for female candidates. *American Politics Research*, *33*(3), 357–375.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, *9*(2), 378-40.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal Of Psychology*, *77*(3), 305–327.

Caprara, G. V. (2007). The personalization of modern politics. *European Review*, *15*(2), 151-164.

Caprara, G. V., Barbaranelli, C., & Zimbardo, P. G. (1997). Politicians' uniquely simple personalities. *Nature*, *385*(6616), 493-493.

Caprara, G. V., Barbaranelli, C., & Zimbardo, P. G. (2002). When parsimony subdues distinctiveness: Simplified public perceptions of politicians' personality. *Political Psychology*, *23*(1), 77-95.

Caprara, G. V., Vecchione, M., Barbaranelli, C., & Fraley, R. C. (2007). When likeness goes with liking: The case of political preference. *Political Psychology*, *28*(5), 609-632.

Carpinella, C. M., & Johnson, K. L. (2013). Politics of the face: The role of sex-typicality in trait assessments of politicians. *Social Cognition*, *31*(6), 770-779.

Carpinella, C. M., Hehman, E., Freeman, J. B., & Johnson, K. L. (2016). The gendered face of partisan politics: Consequences of facial sex typicality for vote choice. *Political Communication*, *33*(1), 21–38.

Casey, K. (2022). Snap judgments: Predicting politician competence from photos. *The Journal of Politics*, *84*(2), 698–713.

Castelli, L., Carraro, L., Ghitti, C., & Pastore, M. (2009). The effects of perceived competence and sociability on electoral outcomes. *Journal of Experimental Social Psychology*, *45*(5), 1152-1155.

Champely, S. (2022, October 14). *PWR: Basic functions for power analysis*. Package 'pwr.' https://cran.r-project.org/web/packages/pwr/pwr.pdf

Chen, F. F., Jing, Y., & Lee, J. M. (2012). "I" value competence but "we" value social competence: The moderating role of voters' individualistic and collectivistic orientation in political elections. *Journal of Experimental Social Psychology*, *48*(6), 1350–1355.

Chen, F. F., Jing, Y., Lee, J. M., & Bai, L. (2016). Culture matters: The looks of a leader are not all the same. *Social Psychological and Personality Science*, *7*(6), 570–578.

Cheung, O. S., & Jintcharadze, D. (2023). Effects of masculinity vs. femininity on competence judgment of politician faces and election outcome prediction. *Scientific Reports*, *13*(1), 16825.

Connolly, H. L., Young, A. W., & Lewis, G. J. (2021). Face perception across the adult lifespan: evidence for age-related changes independent of general intelligence. *Cognition and Emotion*, *35*(5), 890–901.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal Of Statistical Software*, 34, 1-24.

Darwin, C. (1859). On the origin of species: facsimile of the first edition.

Davidesco, I., Zion-Golumbic, E., Bickel, S., Harel, M., Groppe, D. M., Keller, C. J., ... & Malach, R. (2014). Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cerebral Cortex*, *24*(7), 1879-1893.

De Raad, B. (1998). Five Big Five issues: Rationale, structure, status and cross-cultural assessment. *European Psychologist*, 3,113–124

De Raad, B. (2000). *The big five personality factors: the psycholexical approach to personality*. Hogrefe & Huber Publishers.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1498), 1307-1312.

DeBruine, L. M. (2004). Facial resemblance increases the attractiveness of same-sex faces more than other–sex faces. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1552), 2085-2090.

DeBruine, L. M., Jones, B. C., Little, A. C., & Perrett, D. I. (2008). Social perception of facial resemblance in humans. *Archives Of Sexual Behavior*, pp. *37*, 64–77.

Deutsch, F. M. (1990). "Status, sex, and smiling: The effect of smiling on men and women." *Personality and Social Psychology Bulletin*, 16(3), 531-540.

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behavioral Ecology*, *23*(3), 481–490.

Dugatkin, L. A. (2007). Inclusive fitness theory from Darwin to Hamilton. *Genetics*, *176*(3), 1375–1380.

Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so competent?—Contemporary implicit stereotypes of women and men in Germany. *Sex Roles*, *70*, 359-375.

Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, *73*(361), 113–121.

Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique.

Ekman, P. (Ed.). (2006). *Darwin and facial expression: A century of research in review*. Ishk.

Elbich, D. B., & Scherf, S. (2017). Beyond the FFA: brain-behavior correspondences in face recognition abilities. *Neuroimage*, *147*, 409-422.

Engelmann, S., Ullstein, C., Papakyriakopoulos, O., & Grossklags, J. (2022). What people think AI should infer from faces. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 128-141).

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect?. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 628.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception?. *Psychological review*, *105*(3), 482.

Feingold, G. A. (1914). Influence of environment on the identification of persons and things. *J. Am. Inst. Crim. L. & Criminology*, *5*, 39.

Ferreira Da Silva, F., & Costa, P. (2019). Do we need warm leaders? Exploratory study of the role of voter evaluations of leaders' traits on turnout in seven European countries. *European Journal of Political Research*, *58*(1), 117-140.

Fisher, K., Towler, J., & Eimer, M. (2017). Face identity matching is selectively impaired in developmental prosopagnosia. *Cortex*, pp. *89*, 11–27.

Fiske, S. T. (2010). Venus and Mars or down to Earth: Stereotypes and realities of gender differences. *Perspectives on Psychological Science*, *5*(6), 688–692.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends In Cognitive Sciences*, *11*(2), 77–83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. In *Social Cognition* (pp. 162–214). Routledge.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*(200), 675–701.

Fuhrmann, D., Knoll, L. J., Sakhardande, A. L., Speekenbrink, M., Kadosh, K. C., & Blakemore, S. J. (2016). Perception and recognition of faces in adolescence. *Scientific Reports*, *6*(1), 33497.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197.

Giang, T., Bell, R., & Buchner, A. (2012). *Does facial resemblance enhance cooperation?*.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology In Europe*, *7*(1), 7-28.

Hamilton, W. D. (1964). The genetical evolution of social behavior. II. *Journal Of Theoretical Biology*, *7*(1), 17–52.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends In Cognitive Sciences*, *4*(6), 223–233.

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*, *13*(2), e12431.

Heisz, J. J., Pottruff, M. M., & Shore, D. I. (2013). Females scan more than males: A potential mechanism for sex differences in recognition memory. *Psychological science*, 24(7), 1157-1163.

Henley, N. M. (1973). Power, sex, and nonverbal communication. *Berkeley Journal of Sociology*, pp. 1–26.

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9-10), 1306-1336.

Herrmann, M., & Shikano, S. (2016). Attractiveness and Facial Competence Bias Face-Based Inferences of Candidate Ideology. *Political Psychology*, *37*(3), 401-417.

Hess, U., & Thibault, P. (2009). Darwin and emotion expression. *American Psychologist*, *64*(2), 120.

Hogg, M. A., & Turner, J. C. (1987). Intergroup behavior, self-stereotyping and the salience of social categories. *British Journal of Social Psychology*, *26*(4), 325–340.

Homa, D., Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition*, *4*(2), 176-185.

Hughes, B. L., Camp, N. P., Gomez, J., Natu, V. S., Grill-Spector, K., & Eberhardt, J. L. (2019). Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception. *Proceedings of the National Academy of Sciences*, *116*(29), 14532-14537

Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation?. *Social Psychological and Personality Science*, 13 (2), 553–561.

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know?. *Proceedings of the Royal Society B*, *285*(1888), 20181319.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *Journal Of Personality And Social Psychology*, *89*(6), 899.

Kahneman, D. (2013). *A perspective on judgment and choice: Mapping bounded rationality. Progress in Psychological Science around the World*. Volume 1 Neural, Cognitive and Developmental Issues., pp. 1–47.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. Heuristics and biases: The psychology of intuitive judgment, 49(49-81), 74.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception.

Kervyn, N., Yzerbyt, V., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology*, *21*(1), 155-187.

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific Reports*, *11*(1), 100.

Kosinski, M., Khambatta, P., & Wang, Y. (2024). Facial recognition technology and human raters can predict political orientation from images of expressionless faces, even when controlling for demographics and self-presentation. *American Psychologist*.

Krupp, D. B., Debruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior*, *29*(1), 49–55.

Krupp, D. B., DeBruine, L. M., Jones, B. C., & Lalumière, M. L. (2012). Kin recognition: evidence that humans can perceive both positive and negative relatedness. *Journal of Evolutionary Biology*, *25*(8), 1472-1478.

Ladam, C., Harden, J. J., & Windett, J. H. (2018). Prominent role models: high-profile female politicians and the emergence of women as candidates for public office. *American Journal of Political Science*, *62*(2), 369-381.

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends In Cognitive Sciences*, *4*(2), 50–59.

Lin, C., Adolphs, R., & Alvarez, R. M. (2017). Cultural effects on the association between election outcomes and face-based trait inferences. *PloS one*, *12*(7), e0180837.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, *28*(1), 18-27.

Lodge, M., McGraw, K. M., & Stroh, P. (1989). An impression-driven model of candidate evaluation. *American Political Science Review*, *83*(2), 399–419.

Lopuka, S. (2021, April 13). *Thispersondoesnotexist - random AI generated photos of fake persons*. This Person Does Not Exist - Random Face Generator. https://this-person-does-not-exist.com/en

Lüdecke, D. (2023). sjPlot: Data Visualization for Statistics in Social Science. *R package version 2.8.15*, https://CRAN.R-project.org/package=sjPlot.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. Behavior research methods, 47, 1122-1135.

Ma, D. S., Kantner, J., & Wittenbrink, B. (2021). Chicago face database: Multiracial expansion. *Behavior Research Methods*, 53, 1289-1300.

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. (2009). Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology*, *45*(3), 496-504.

Maeng, A., & Aggarwal, P. (2022). The face of political beliefs: why gender matters for electability. *Journal of the Association for Consumer Research*, *7*(3), 360–370.

Magic Morph 1.95. Magic Morph---Morphing Software. (n.d.). https://www.effectmatrix.com/morphing/

Mangiafico, S. S. (2024). rcompanion: Functions to Support Extension Education Program Evaluation. Version 2.4.36. *Rutgers Cooperative Extension*. New Brunswick, New Jersey. https://CRAN.R-project.org/package=rcompanion

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals Of Mathematical Statistics*, pp. 50–6.

McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2016). A critical review of the development of face recognition: Experience is less important than previously believed. *Understanding Cognitive Development*, pp. 174–212.

Metz, M., Kruikemeier, S., & Lecheler, S. (2020). Personalization of politics on Facebook: Examining the content and effects of professional, emotional and private self-personalization. *Information, Communication & Society*, *23*(10), 1481-1498.

Mignon, A., Mollaret, P., Rohmer, O., & Bagès, C. (2016). Effect of Political Orientation on Judgment of Agency, Competence, Morality, and Sociability. *Swiss Journal of Psychology*.

Miles, C. C., & Miles, W. R. (1961). The correlation of intelligence scores and chronological age from early to late maturity.

Nakano, T., & Yamamoto, T. (2022). You trust a face like yours. *Humanities and Social Sciences Communications*, *9*(1), 1-6.

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal Of Experimental Psychology: Human Learning And Memory*, *3*(5), 485.

O'toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*, 208-224.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*(2), 323.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315-324.

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer Republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science*, *3*(5), 605-613.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Parr, L. A., & de Waal, F. B. (1999). Visual kin recognition in chimpanzees. *Nature*, *399*(6737), 647-648.

Parr, L. A., Heintz, M., Lonsdorf, E., & Wroblewski, E. (2010). Visual kin recognition in nonhuman primates:(Pan troglodytes and Macaca mulatta): inbreeding avoidance or male distinctiveness?. *Journal of Comparative Psychology*, *124*(4), 343.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal Of Neuroscience*, *18*(6), 2188–2199.

Quist, M. C., Watkins, C. D., Smith, F. G., DeBruine, L. M., & Jones, B. C. (2011). Facial masculinity is a cue to women's dominance. *Personality and Individual Differences*, *50*(7), 1089–1093.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rammstedt, B., Spinath, F. M., Richter, D., & Schupp, J. (2013). Partnership longevity and personality congruence in couples. *Personality and Individual Differences*, *54*(7), 832-835.

Rehnman, J. (2007). *The role of gender in face recognition* (Doctoral dissertation, Psykologiska institutionen).

Rhodes, G. (2013). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, *42*(11), 1179–1199.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain And Cognition*, *79*(2), 138-157.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*, *45*(6), 1245–1251.

Rule, N., & Ambady, N. (2010). First impressions of the face: Predicting success. *Social and Personality Psychology Compass*, *4*(8), 506-516.

Sanbonmatsu, K. (2020). Women's Underrepresentation in the US Congress. *Daedalus*, *149*(1), 40-55.

Santos, I. M., & Young, A. W. (2011). Inferring social attributes from different face regions: Evidence for holistic processing. *Quarterly Journal of Experimental Psychology*, *64*(4), 751-766.

Schmitz, M., Vanbeneden, A., & Yzerbyt, V. (2024). The many faces of compensation: The similarities and differences between social and facial models of perception. *PloS one*, *19*(2), e0297887.

Sevi, S. (2021). Do young voters vote for young leaders?. *Electoral Studies*, p. *69*, 102200.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591-611.

Simion, F., Leo, I., Turati, C., Valenza, E., & Dalla Barba, B. (2007). How face specialization emerges in the first months of life. *Progress In Brain Research*, *164*, 169-185.

Sugita, Y. (2009). Innate face processing. *Current Opinion In Neurobiology*, *19*(1), 39–44.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118.

Taubert, J., Wardle, S. G., Flessert, M., Leopold, D. A., & Ungerleider, L. G. (2017). Face pareidolia in the rhesus monkey. *Current Biology*, *27*(16), 2505-2509.

Tea-Makorn, P. P., & Kosinski, M. (2020). Spouses' faces are similar but do not become more similar with time. *Scientific Reports*, *10*(1), 17001.

Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. In *Advances In Experimental Social Psychology* (Vol. 63, pp. 189-245). Academic Press.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review Of Psychology*, 66, 519-545.

Tomonaga, M., & Kawakami, F. (2023). Do chimpanzees see a face on Mars? A search for face pareidolia in chimpanzees. *Animal Cognition*, *26*(3), 885-905.

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). FFA and OFA encode distinct types of face identity information. *Journal of Neuroscience*, *41*(9), 1952-1969.

Turati, C., Sangrigoli, S., Ruely, J., & de Schonen, S. (2004). Evidence of the face inversion effect in 4-month-old infants. *Infancy*, *6*(2), 275-297.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.

Uğurbil, K. (2012). Development of functional imaging in the human brain (fMRI); the University of Minnesota experience. *Neuroimage*, *62*(2), 613-619.

Umiltá, C., Simion, F., & Valenza, E. (1996). Newborn's preference for faces. *European Psychologist*, *1*(3), 200-205.

Van Vugt, M., & Grabo, A. E. (2015). The many faces of leadership: An evolutionary-psychology approach. *Current Directions in Psychological Science*, *24*(6), 484-489.

Volpato, C. (2022). *Psicosociologia del maschilismo*. Gius. Laterza & Figli Spa.

Wang, Z., Mao, H., Li, Y. J., & Liu, F. (2017). Smile big or not? Effects of smile intensity on perceptions of warmth and competence. *Journal of Consumer Research*, *43*(5), 787-805.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology*, *73*(1), 117–129.

Wei, T., Simko, V. R., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2021). package "corrplot": Visualization of a Correlation Matrix. 2017. *Version .84*.

Westbrook, K. E., Nessel, T. A., Hohman, M. H., & Varacallo, M. (2018). Anatomy, head and neck, facial muscles.

Wickens, A. P. (2014). *A history of the brain: from stone age surgery to modern neuroscience*. Psychology Press.

Wickham, H., Bryan, J. (2023). readxl: Read Excel Files. *R Package Version 1.4.3*, https://CRAN.R-project.org/package=readxl.

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version*, *2*(1), 1–189.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in statistics: Methodology and distribution* (pp. 196–202). New York, NY: Springer New York.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592-598.

Zajonc, R. B., Adelmann, P. K., Murphy, S. T., & Niedenthal, P. M. (1987). Convergence in the physical appearance of spouses. *Motivation and Emotion*, *11*, 335-346.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*(3), 237-242.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517.

Zheng, R., Ren, D., Xie, C., Pan, J., & Zhou, G. (2021). Normality mediates the effect of symmetry on facial attractiveness. *Acta Psychologica*, p. *217*, 103311.

Zhou, L. F., & Meng, M. (2020). Do you see the "face"? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology*, *14*, e2.

Appendix I – The Experimental Questionnaire

Informed_Consent

Gentile partecipante,

le proponiamo di aderire ad uno studio on-line il cui scopo è quello di: identificare quali criteri vengono usati per assegnare la probabilità che un elettore o un'elettrice abbia votato un determinato candidato o candidata. Di volta in volta Le verrà mostrato, al centro dello schermo, il volto di una persona (maschile o femminile) che ha partecipato come elettore/elettrice ad un'ipotetica tornata elettorale. Sotto al volto dell'elettore saranno presenti altri due volti che rappresentano i due candidate (o candidate) rispetto ai quali l'elettore (o elettrice) ha espresso la propria preferenza. Le verrà chiesto di "indovinare" per quale candidato/a ciascun elettore (o elettrice) ha secondo Lei votato.

Le chiediamo di esprimere risposte rapide in quanto siamo interessati alle prime impressioni spontanee.

DESCRIZIONE

Il questionario è diviso nelle seguenti parti principali:

1. Presentazione dei volti di 10 persone (elettori/elettrici), per ciascuno dei quali verranno presentati anche i volti di 2 candidati/candidate. Il suo compito sarà indovinare per chi, secondo lei, l'elettore/elettrice ha votato (10 domande).
2. Informazioni demografiche (genere, età, provenienza etnica, autocollocazione politica, livello di informazione sulla politica).

Il tempo previsto per la compilazione è di circa 5 minuti.

TRATTAMENTO DATI

Tutte le informazioni raccolte in questa ricerca saranno trattate nel rispetto delle vigenti leggi D.Lgs.196/2003 sulla privacy novellato dal D.Lgs. n. 101/2018, UE GDPR 679/2016 sulla protezione dei dati personali e dell'art. 9 del Codice Deontologico degli Psicologi Italiani. I suoi dati saranno analizzati in modo /anonimo e con tutti i criteri che garantiscono la massima riservatezza, utilizzati unicamente ai fini della ricerca medesima. La responsabile della ricerca è Luciana Carraro, Professoressa Associata, afferente a

DPSS. Indirizzo: Via Venezia 8, 35131 - Padova, tel. 0498276525; email: luciana.carraro@unipd.it

La responsabile della ricerca si impegna ad adempiere agli obblighi previsti dalla normativa vigente in termine di raccolta, trattamento e conservazione di dati sensibili. Ogni partecipante ha in ogni momento facoltà di esercitare i diritti di cui all'art. 7 del D.Lgs.196/2003 novellato dal D.Lgs. n. 101/2018. I dati, raccolti ed elaborati in forma aggregata e anonima, potranno essere inseriti in pubblicazioni e/o presentati a congressi o seminari scientifici. Il trattamento dei suoi dati sarà avviato solo con la sottoscrizione di tale consenso.

DICHIARO:

- Di essere maggiorenne;
- Di aderire volontariamente alla realizzazione della ricerca in qualità di partecipante
- Di essere a conoscenza degli obiettivi e delle finalità di tale progetto di ricerca
- Di essere a conoscenza che i dati ricavati, nell'assoluto anonimato, saranno trattati esclusivamente per fini didattici e di ricerca
- Di essere consapevole che non è prevista la possibilità di ottenere la restituzione dei dati raccolti una volta inviati.

Per eventuali chiarimenti è possibile contattare il Dott. Federico Padovan, e-mail: federico.padovan@studenti.unipd.it.

Grazie mille per il tuo prezioso contributo!

Proseguendo nella compilazione del questionario esprimo il consenso a partecipare alla ricerca.

Proseguendo nella compilazione del questionario esprimo il consenso apartecipare alla ricerca.

○ Acconsento

○ Non acconsento

Introduction_Triplets

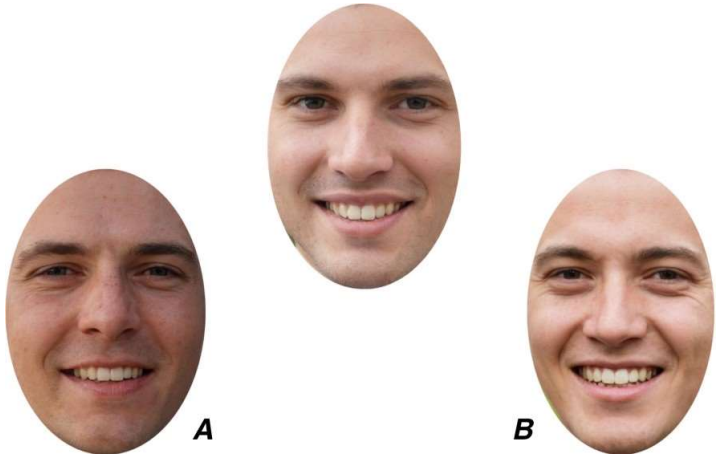
Grazie per aver acconsentito a partecipare al presente studio.
Nella prossima sezione Le verranno mostrate **dieci ipotetiche tornate elettorali**.

In ciascuna di esse, Le verrà chiesto di **esprimere la probabilità che un elettore o elettrice**, il cui volto sarà presentato nella **parte superiore dello schermo**, **abbia votato per uno dei due candidati o candidate** i cui volti saranno presentati nella **parte inferiore dello schermo**.

MT1

Secondo lei, **per quale dei due candidati politici** (A e B, presentati inbasso) **ha votato l'elettore** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

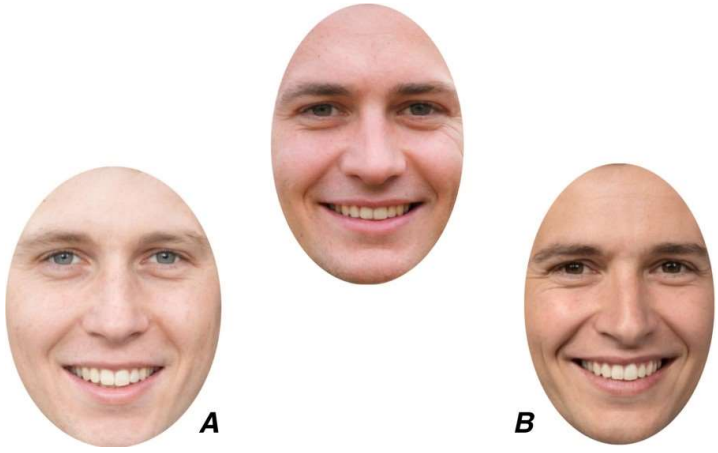


Sicuramente per il candidato A — Sicuramente per il candidato B

MT2

Secondo lei, **per quale dei due candidati politici** (A e B, presentati inbasso) **ha votato l'elettore** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

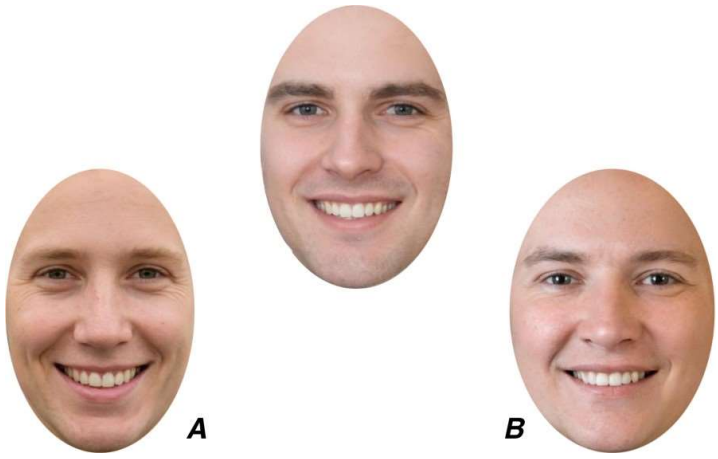


Sicuramente per il candidato A — Sicuramente per il candidato B

MT3

Secondo lei, **per quale dei due candidati politici** (A e B, presentati inbasso) **ha votato l'elettore** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

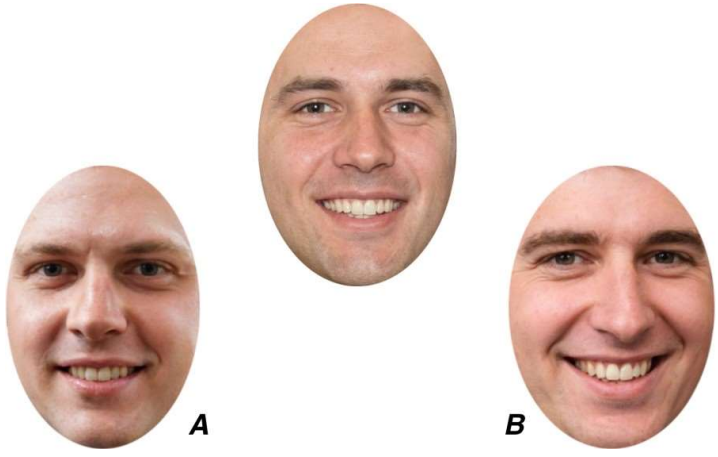


Sicuramente per il candidato A — Sicuramente per il candidato B

MC1

Secondo lei, **per quale dei due candidati politici** (A e B, presentati inbasso) **ha votato l'elettore** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

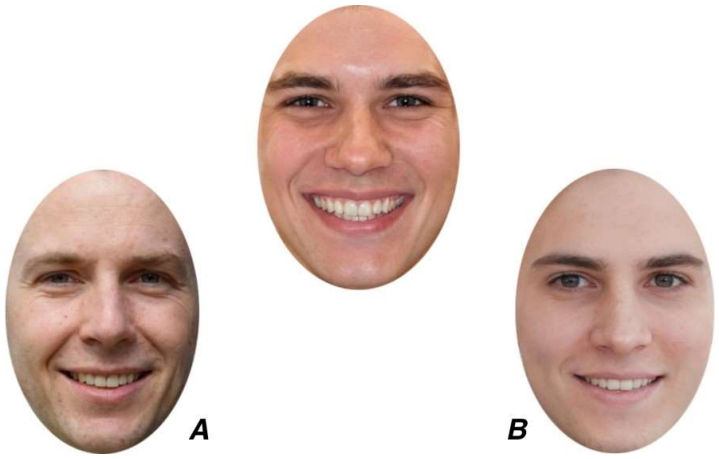


Sicuramente per il candidato A — Sicuramente per il candidato B

MC2

Secondo lei, **per quale dei due candidati politici** (A e B, presentati inbasso) **ha votato l'elettore** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

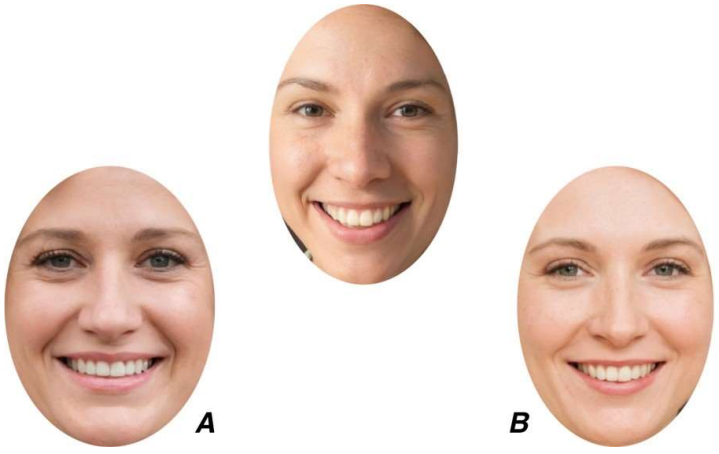


Sicuramente per il candidato A — Sicuramente per il candidato B

FT1

Secondo lei, **per quale delle due candidate politiche** (A e B, presentati inbasso) **ha votato l'elettrice** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

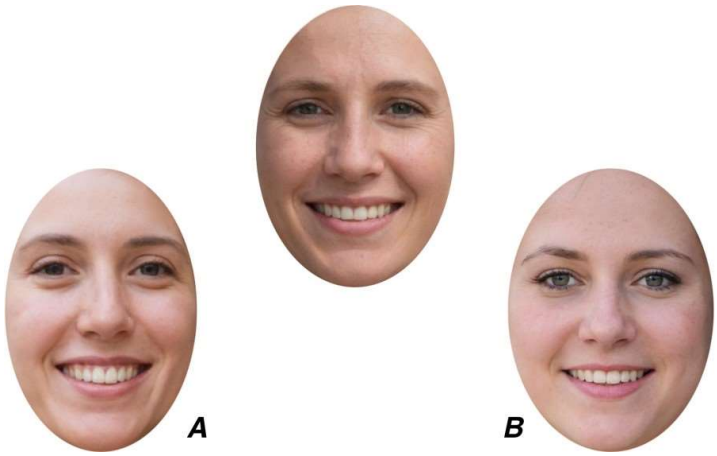


Sicuramente per la candidata A — Sicuramente per la candidata B

FT2

Secondo lei, **per quale delle due candidate politiche** (A e B, presentati inbasso) **ha votato l'elettrice** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

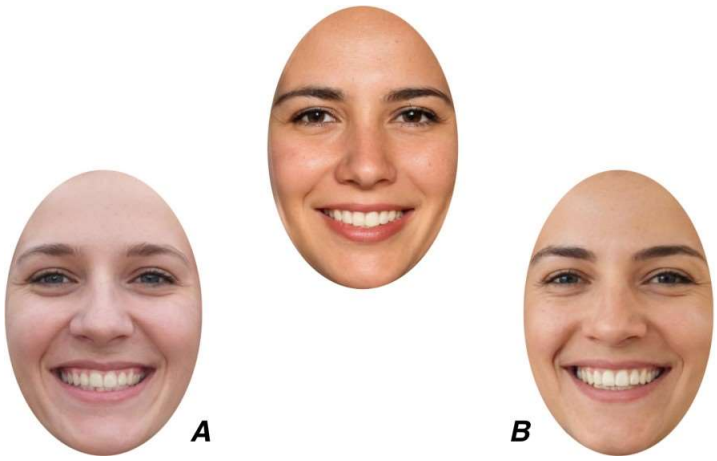


Sicuramente per la candidata A — Sicuramente per la candidata B

FT3

Secondo lei, **per quale delle due candidate politiche** (A e B, presentati inbasso) **ha votato l'elettrice** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

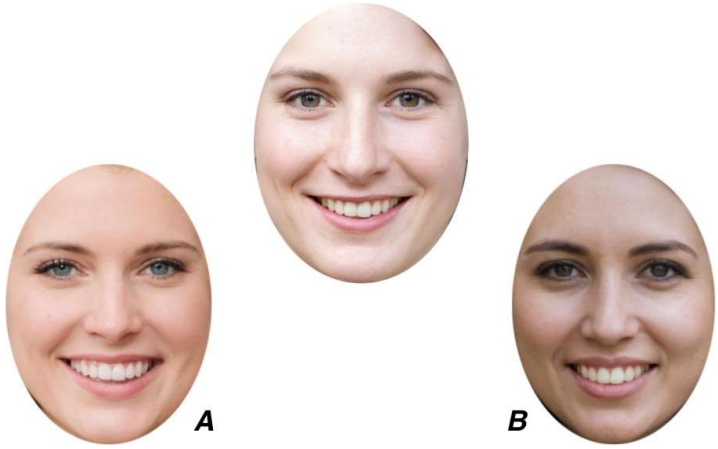


Sicuramente per la candidata A — Sicuramente per la candidata B

FC1

Secondo lei, **per quale delle due candidate politiche** (A e B, presentati inbasso) **ha votato l'elettrice** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:

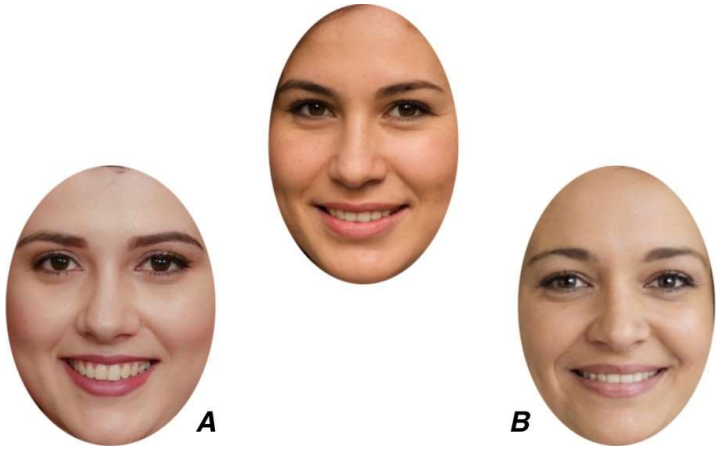


Sicuramente per la candidata A Sicuramente per la candidata B

FC2

Secondo lei, **per quale delle due candidate politiche** (A e B, presentati inbasso) **ha votato l'elettrice** il cui volto è presentato in alto?

Esprima la sua opinione trascinando il cursore in basso verso il volto di uno dei candidati:



Sicuramente per la candidata A Sicuramente per la candidata B

Introduction_Criteria

Il questionario è **quasi terminato**!

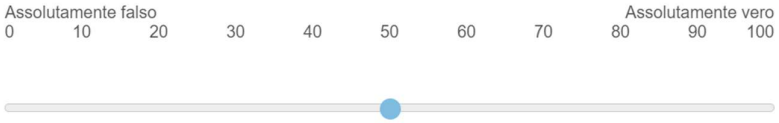
In questa breve sezione, le chiederemo che criteri ha utilizzato nell'assegnare leprobabilità di voto.

Siamo consapevoli che alcuni processi avvengono in maniera automatica e non del tutto consapevole, ma Le chiediamo comunque di esprimere il Suo giudizio a riguardo. Le chiediamo di rispondere alle domande che seguiranno in base a **quanto è vero per lei**.

Criteria

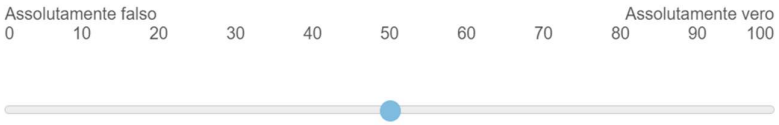
Criterion_Competence

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più competente**:



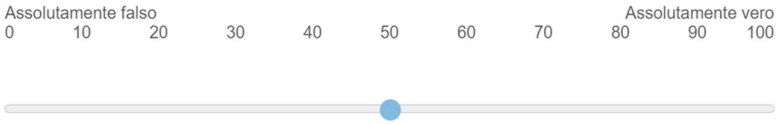
Criterion_Warmth

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più socievole**:



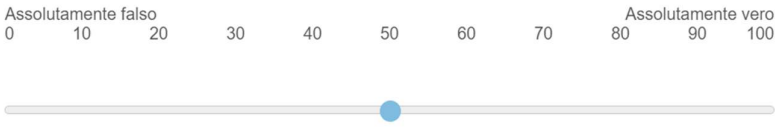
Criterion_Morality

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più morale**:



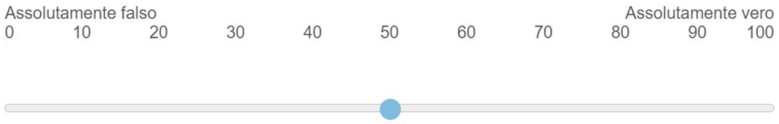
Criterion_Attractiveness

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più attraente**:



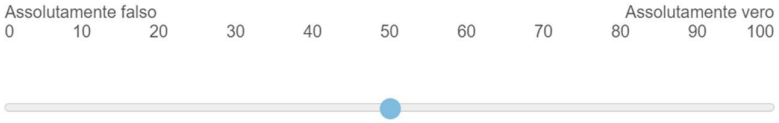
Criterion_Maturity

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più maturo/a**:



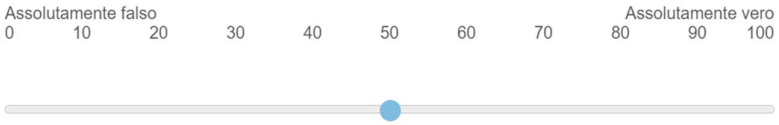
Criterion_Facial_Similarity

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più simile fisicamente all'elettore o elettrice:**



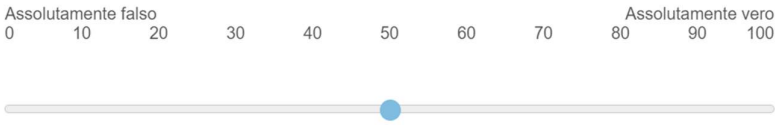
Criterion_Self_Similarity

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più simile a me:**



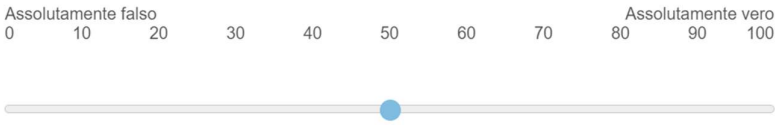
Criterion_Personality_Similarity

Ho assegnato **maggiore probabilità di voto** al **candidato o candidata** che personalmente ritenevo **più simile all'elettore o elettrice dal punto di vista della personalità:**



Attention_Check

Se sta **prestando attenzione**, Le chiediamo di spostare lo slider completamente verso "**Assolutamente vero**".



Introduction_Demographics

Grazie per aver espresso la sua opinione.

In ultimo, le chiediamo alcune brevi **informazioni su di sé**.

Demographics

Gender

Con quale genere si identifica?

- Femminile
- Maschile
- Non-binario/Altro
- Preferisco non specificare

Age

Indichi con un valore numerico la sua età (in anni compiuti): ☐

Instruction_Degree

Qual è il grado di istruzione più alto che ha conseguito?

- Nessuno
- Diploma di Scuola Primaria (Elementare)
- Diploma di Scuola Secondaria di Primo Grado (Media)
- Diploma di Scuola Secondaria di Secondo Grado (Superiore)
- Laurea Triennale
- Laurea Magistrale
- Master universitario, Specializzazione o Dottorato di ricerca

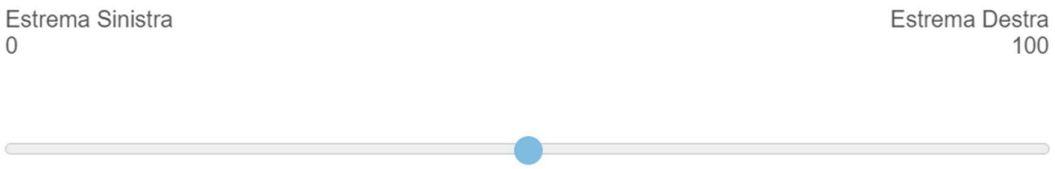
Ethnicity

Quale di queste etichette descrive meglio la sua **origine etnica?**

- Italiana
- Europea non italiana (Albanese, Romena, ecc.)
- Africana
- Altra minoranza etnica
- Preferisco non specificare

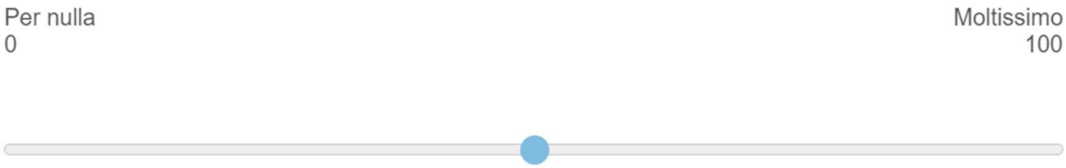
Political_Orientation

Come si posiziona politicamente?



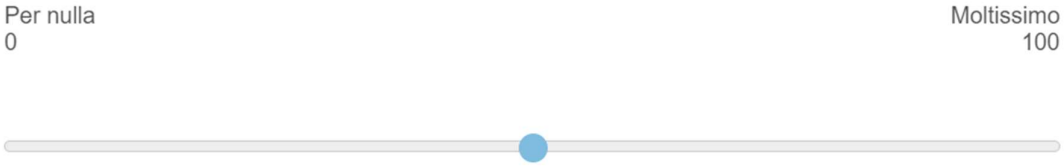
Political_Sophistication_1

Quanto si ritiene interessato alla politica?



Political_Sophistication_2

Quanto si ritiene informato sulla politica (vicende e personaggi politici attuali; programmi di partito; ecc.)?



Ending

Grazie per la partecipazione!

Il Suo contributo è prezioso per la ricerca.
Per concludere la compilazione ed uscire dal questionario, **clicchi sul pulsante in basso a destra**.

Appendix II – The effect of morphing on similarity

Although the morphing process was identical for all experimental triplets, results from the main experimental questionnaire appear to suggest that similarity may not be equally detectable across them. In particular, since the data show a significantly lower vote attribution toward the most similar candidate across the three experimental female triplets, it is possible that the effect of the manipulation might have been particularly feeble and difficult to detect in these triplets, either due to their gender or to the characteristics of the specific faces used. Thus, an additional study was constructed in order to assess the experimental manipulation's effectiveness and to test the following opposing hypotheses:

H5a: The morphing procedure will lead to a similar effect of similarity in male and female triplets.

H5b: The morphing procedure will lead to a higher effect of similarity in the male rather than in the female triplets.

1. The Questionnaire's structure

Similar to the two questionnaires discussed in the present dissertation, this study was created on the Qualtrics online platform and distributed on social media websites such as Facebook, Instagram, and WhatsApp.

The questionnaire reproduced the experimental section of the Vote Attribution Questionnaire, as described in paragraph 3.3.2, and was thus constituted of 10 questions of similarity attribution (six experimental, morphed triplets and four non-morphed, confounding triplets). Compared to the Vote Attribution Questionnaire, the current one changed the main question asked to the participants for each triplet. Instead of being asked to express the probability that the elector voted for one of the two candidates, each participant was asked to move the cursor on a slider to indicate which of the two faces represented on the bottom part of the screen resembled the most the face on the upper part. At the beginning of the study, the participants were informed that the face at the center of the screen was manipulated to resemble one of the two other faces. Nonetheless, they could express uncertainty about which face resembled the target face the most by not moving the cursor completely toward a face.

Since the questionnaire was unrelated to the political choices, it lacked a section relative to the criteria of vote attribution that, on the contrary, was present in the previously discussed questionnaire. However, the question which asked the participants to report how much they tended to rely on first impressions was maintained.

Lastly, the questionnaire also included four demographic questions regarding the participant's gender, age, instruction degree, and ethnicity. The structure of all the questions was analogous to that reported in paragraph 3.3.2.

Since the effect size was expected to be significantly higher (*Cohen's d* = .40) than the effect size of similarity on vote attribution, the dimension of the needed sample was estimated to be 53 to detect the effect with a *power* = .90 and a significance level of p = .05 (Champeley, 2022).

2. Participants

71 people took part in the questionnaire. However, only 61 satisfied the criteria, and 10 people were excluded: 1 person did not agree to the initial consent, 7 did not complete the entire questionnaire, and 3 were of non-European ethnic origin.

The sample was largely unrepresentative since it presented a relatively low mean age (M = 31.97, *Mdn* = 26, *sd* = 13.15), was composed primarily of female participants (N = 42), and presented a generally high degree of instruction (37 people had achieved a degree equal to or higher than a Bachelor's degree).

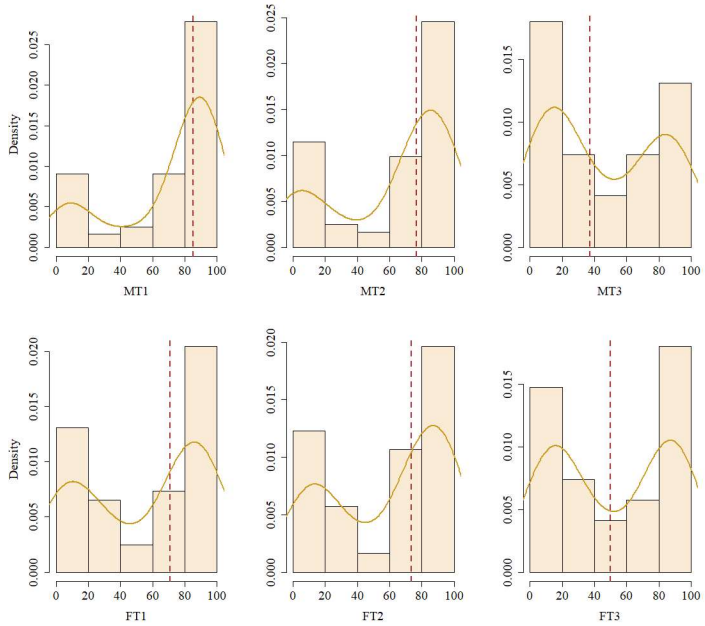
3. Results

As Figure 17 shows, the shape, mean, and median of the distributions of the variable Similarity varied widely across the experimental triplets.

Since it was expected that the data would be skewed rightward, Wilcoxon tests were performed to verify that the manipulation worked and that the similarity between faces would have been enhanced by the morphing procedures. The results of the tests, as well as the descriptive statistics for each triplet, are reported in Table 9.

Figure 17

Histograms of the Similarity Attribution variable for the six experimental triplets



Note. Distributions of the similarity attribution variable in the dataset, with density curves overlaid. The dashed lines represent the median of each distribution.

Table 9

Descriptive Statistics and Wilcoxon Test Results for the Six Experimental Triplets

	M	*Mdn*	*sd*	*r*
MT1	69.49	85.2	33.47	.457***
MT2	62.33	76.7	36.10	.214*
MT3	46.12	37.3	35.25	-.134
FT1	54.78	70.8	37.30	.067
FT2	59.20	73.6	35.87	.196*
FT3	51.47	50.0	36.25	.030

Note. $N = 61$. * = p < .05; ** = p < .01; *** = p < .001

Performing a non-parametric repeated measures ANOVA (Friedman, 1937) also confirmed that the mean significantly varied across the six triplets ($\chi^2 = 21.0$, $df = 5$, The Durbin-Conover pairwise comparison (Durbin, 1951; Conover, 1971) revealed significant differences in the mean of the triplet MT3 compared to the other two male triplets MT1 (S = 4.07, p <.001) and MT2 (S = 2.92, p = .004). These findings are further supported by the significant differences observed in the three female triplets compared to the mean of both MT1 and MT2, as shown in Table 10.

Table 10

Durbin-Conover pairwise comparison results

	S	*p*
MT1 – MT2	1.149	0.252
MT1 – MT3	4.072	< .001
MT1 – FT1	2.532	0.012
MT1 – FT2	1.853	0.065
MT1 – FT3	3.393	< .001
MT2 – MT3	2.924	0.004
MT2 – FT1	1.383	0.168
MT2 – FT2	0.705	0.481
MT2 – FT3	2.245	0.026
MT3 – FT1	1.540	0.125
MT3 – FT2	2.219	0.027
MT3 – FT3	0.679	0.498
FT1 – FT2	0.679	0.498
FT1 – FT3	0.861	0.390
FT2 – FT3	1.540	0.125

Note. Pairwise comparisons between the means of the Similarity Attribution variable for the six experimental triplets. $N = 61$.

4. Conclusions

In general, as visible in Table 9, the analyses support the existence of a strong effect of similarity for MT1. On the contrary, the effect of the morphing appears weak but still significant for MT2 and FT2. The effect is undetectable for triplets FT1 and FT3, while it is in the opposite direction for MT3. The data also suggests (Table 10) that recognizing similarity may have been more complex in the three experimental female triplets than in the first two male triplets. This lower effect of the manipulation could have led to lower vote attributions to the candidates in the female triplets in the main questionnaire. Hypothesis 5b thus seems to be only partially supported by the data, which revealed that the effect of the experimental manipulation on similarity was significantly weaker for all three female triplets than for triplets MT1 and MT2. The case of MT3, however, is

problematic and seems to support 5a. The effect of similarity detected in the triplet MT3, in fact, is similar to the one detected in the triplets FT1 and FT3, thus suggesting that random characteristics of the faces used may lead to differences in the perceived effect of the experimental manipulation and that gender may not be the main explaining factor. Moreover, despite MT3 presenting the lowest effect of similarity across all triplets in the present questionnaire, MT3 exhibits an effect of similarity on Vote Attribution comparable with the other two male triplets in the data from the main study.

This suggests that, due to the small and unrepresentative sample recruited for this last questionnaire, the estimated effects of the experimental manipulation may be unreliable and that the sample size might be insufficient to draw definitive conclusions.

Lastly, the present questionnaire allowed us to assess whether the size of the manipulation was comparable with naturally occurring similarities among faces. The data show that participants overall attributed higher similarity to the correct candidate 59.56% of the time. Although this percentage is significantly higher than chance, it is not higher than the percentage of people who judged the faces MI2 and MI6 ($P = 63.93\%$), WI5 and WI8 ($P = 72.13\%$), WI9 and WI7 ($P = 77.05\%$), and WI3 and WI12 ($P = 78.69\%$) to be similar. These data support the idea that the experimental manipulation was subtle and not immediately detectable. This also suggests that the experimental manipulation may also have a certain ecological validity since it matches the effect size of naturally occurring similarities between random faces.

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Luciana Carraro, for her unwavering support throughout the writing process and for her constant and expert feedback.

I would also like to thank my co-advisor, Professor Luigi Castelli, for his significant contributions to the structure of the study and his insightful advice, as well as Professor Massimiliano Pastore for his competent statistical support.

I wish to extend my special thanks to the staff of MyES, especially to Michael Kavanagh and Thomas O'Neill, who helped me improve my English skills to meet this academic challenge.

Lastly, I am deeply grateful to all my friends and family for their patience, understanding, and encouragement. This work is as much yours as it is mine.