Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Psicologia

Ciclo XXXII

**Settore Concorsuale: 11/E3**

**Settore Scientifico Disciplinare: M-PSI 05**

# Prima Facie:
# four facets of gender bias in personnel selection.

**Presentata da:** Sara Pireddu

**Coordinatore Dottorato** **Supervisore**

**Prof. Monica Rubini** **Dr. Michela Menegatti**

**Esame finale anno 2020**

*A zia Wilma*

*"E gli uomini, in generale, iudicano più agli occhi che alle mani; perché tocca a vedere a ognuno, a sentire a pochi. Ognuno vede quello che tu pari, pochi sentono quello che tu se'..."*

N. Machiavelli, Il Principe.

*"It is only shallow people who do not judge by appearances."*

O. Wilde, The Picture of Dorian Gray.

## Abstract

Recent findings have highlighted a 'perfection bias', that is women being evaluated on more criteria than men in the workplace (Moscatelli et al., 2020; Prati et al., 2019). However, these studies have not considered faces as stimuli, even if facial first impressions can affect several real-world outcomes (Todorov et al., 2015). On this basis, the present research aimed to verify the presence of a perfection bias at face perception level, employing for the first time all the four facets of the fundamental dimensions of social judgments (i.e., competence, dominance, morality, sociability; Abele et al., 2016) and attractiveness (Hosoda et al., 2003) as evaluation criteria of applicants' hireability. Four experiments were conducted (total $N$ = 645), employing a gender-neutral position (Study 1) as well as managerial positions (Study 2, 3, 4) and recruiting Italian and British students (Study 1, 2) as well as British workers (Study 3, 4). Results of Study 1 confirmed that male applicants were evaluated only on their facial competence, while female applicants were evaluated on all the other facial traits. However, the other three studies showed a different and unexpected pattern: besides facial attractiveness and competence considered equally important for both male and female applicants, facial dominance was considered as more important in evaluating women, while facial morality and sociability were considered as more important in evaluating men. Hence, results highlighted a sort of 'deficit bias', so that counter stereotypic traits in which men and women are believed weak (Fiske, 1998) were more relevant for their hireability.



**Keywords**: Facets, Facial First Impressions, Gender Stereotypes, Personnel Selection

**TABLE OF CONTENTS**

## Introduction

Even if women's condition in the workplace have been improving over time, the gender gap persists. Indeed, "in comparison to men, women still tend to be employed less, are employed in lower-paid sectors, work on average 6 hours longer per week than men in total (paid and unpaid) but have fewer paid hours, take more career breaks and face fewer and slower promotions" (European Commission, 2018, p. 9). Moreover, looking at the upper echelons of the organizations, the glass ceiling is still present (Barreto, Ryan, & Schmitt, 2009), with a percentage of women that has increased at the higher levels but still ranges from 16% to 33% (European Institute for Gender Equality, 2017).

Gender bias has been usually explained referring to the influence of gender stereotypes, which are represented by the two fundamental content dimensions of social judgment: agency or competence and communion or warmth (e.g., Abele et al., 2016; Eagly & Karau, 2002; Fiske, 1998). In particular, men are seen as competent but not so warm, conversely women are seen as warm but not so competent (Fiske, 1998). For these reasons, women need to provide more evidence of their competence in order to have the same chances as men in the workplace (Biernat, 2012).

Recent evidence has specified four different facets of the two fundamental content dimensions (e.g., Abele et al., 2016; Leach, Ellemers, & Barreto, 2007): competence and dominance as facets of the agency dimension, morality and sociability as facets of the communion dimension (but about the labels, see Abele et al., 2016). In considering different facets of the agency dimension, recent research has uncovered a few changes in the stereotypes content (Eagly, Nater, Miller, Kaufmann, & Sczesny 2019; Hentschel, Heilman, & Peus, 2019). In a similar vein, two works have highlighted an original pattern in the investigation of gender bias in the workplace considering morality and sociability, as the two facets of communion/warmth dimension (Leach et al., 2007), along with competence (Moscatelli, Menegatti, Ellemers, Mariani, & Rubini, 2020; Prati et al., 2019). In particular, results showed that women were evaluated on all the criteria considered, while men were evaluated only on their competence. However, these works did not consider all the

four facets of the fundamental content dimensions and they did not test the so called 'perfection bias' for a leadership position. Moreover, they did not investigate whether this 'perfection bias' could work at face perception level.

In fact, the role of appearance in the workplace is usually associated to attractiveness, since it is broadly recognized its beneficial influence in affecting diverse job-related outcomes (Hosoda, Stone-Romero, & Coats, 2003). However, over the last decade, an increasing number of studies have attested the influence of facial first impressions on people's decisions and behaviours (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Specifically, literature has shown that social and personality traits inferred from faces have consequences on many different real-world outcomes. For example, the level of competence perceived from politicians' faces can affect their likelihoods to win elections (Olivola & Todorov, 2010).

Given the diffuse habit to attach a photo in the CV and the fact that Human Resources (HR) professionals commonly check applicants' social networks accounts (Hoffman & Casnocha, 2012; Napolitano, 2010; Soumitra, 2010), the present research investigated how gender differences in facial first impressions could affect hiring decisions setting experiments as mock hiring processes. Specifically, on the basis of recent findings emerged from three different research fields, four studies tested the presence of the perfection bias at face perception level for a part-time position as well as for leadership positions and employing all the four facets of the fundamental content dimensions.

Below it is firstly reported a review of literature on the three above-mentioned lines of research that represent the pillars under which the research project was built: the fundamental content dimensions of social judgments, facial first impressions and their consequences, gender stereotypes and their influence in the workplace context. At the end of Chapter I, the aim and hypothesis of the present research are elucidated. Chapter II presents Study 1 which tested and confirmed the presence of a perfection bias at face perception level. Chapter III reports Prestudy 1 in which, examining the relevance of the four facets for succeeding as Area Sales Manager, the job

description subsequently employed in the main experiment, the perfection bias hypothesis was supported. Then, it is reported Study 2 which verified for the first time the perfection bias at face perception level for a leadership position considering all the four facets of the fundamental content dimensions. In rejecting the perfection bias hypothesis, results highlighted a puzzling pattern. Hence, the experiment was replicated with Study 3, reported in Chapter V, recruiting adult workers instead of students. In Chapter V it is presented Prestudy 2, which investigated the relevance of the four facets and attractiveness for a Finance Manager position, and Study 4, in which this job description was employed. Across the last three studies, results consistently showed a sort of 'deficit bias', so that men and women were evaluated on those facial facets on which they are stereotypically seen as weak. In Chapter VI, an internal or mini meta-analysis conforms these results. Finally, these findings are broadly discussed in Chapter VII; implications for theory and practice, as well as limitations and future directions are also outlined.

# CHAPTER I

## Theoretical background

The first chapter proposes a review of evidence from which originated the research aim and hypotheses of the present dissertation. Taking sometimes a historical perspective, the chapter also provides a wide-ranging view of the classical literature on the three lines of research considered: the fundamental dimensions of social judgments, facial first impressions and their consequences, gender stereotypes and their influence in the workplace.

As a first step, research on the two fundamental dimensions of social judgments is reviewed. Next, the two dimensions underling facial first impressions are described along with evidence on how social inferences from faces affect people's behaviours and choices. Then, overlapping aspects between these two different lines of research are highlighted. Finally, gender stereotypes are presented in terms of their contents, origins, and how they generate gender discrimination in the workplace.

### The two fundamental content dimensions of social judgment

There is a general consensus on considering agency or competence and communion or warmth as the two fundamental content dimensions of social judgment[1] (Abele et al., 2016; Abele & Wojciszke, 2014; Bakan, 1966; Cuddy, Fiske, & Glick, 2008). The agency/competence dimension encompasses traits related to relevant qualities for goal-achievement, such as being ambitious or intelligent. The communion/warmth dimension includes traits related to qualities relevant for creating and maintaining social relationships, such as being likeable and trustworthy.

Agency/competence and communion/warmth labels come from different lines of research. The agency and communion label was coined in the field of personality psychology by Bakan (1966), who defined them as the two basic modalities of existence, as an individual and as a part of some larger organism respectively. The presence of communion and agency is pervasive in recent

[1] The term "content dimension" means a category of contents that has related semantic meaning, since the dimensions are usually investigated and operationalized as written personality and social traits/descriptions (Abele et al., 2016).

research on the self or on gender (Abele, 2003; Abele & Wojciszke, 2007) and they were adopted in the 'Dual Perspective Model of Agency and Communion' (DPM-AC; Abele & Wojciszke, 2014). The competence and warmth label is well known in stereotype research as the two core dimensions of the 'Stereotype Content Model'[2] (Cuddy et al., 2008; Fiske, Cuddy, & Glick, 2007). These two models have been reliably found in many different countries (Abele, Uchronski, Suitner, & Wojciszke, 2008; Cuddy et al., 2009).

Beyond agency/competence and communion/warmth, many other different labels have been used for the two dimensions (Abele, Cuddy, Judd, & Yzerbyt, 2008): expressiveness and instrumentality (Bales & Parsons, 1955); other-profitability and self-profitability (Peeters, 1992); socially and intellectually good-bad (Rosenberg, Nelson, & Vivekananthan, 1968); social desirability and social utility (Dubois & Beauvois, 2005); socio-morality and task-ability (Ybarra et al., 2008); agreeableness and extroversion in the Wiggins's interpersonal circumplex of behaviours (Wiggins, 1979) to name just a few. Moreover, the two dimensions have been used in studies on person perception (Abele & Bruckmüller, 2011; Asch, 1946; Rosenberg et al., 1968), judgments about political candidates USA (Abelson, Kinder, Peters, & Fiske, 1982; Kinder & Sears, 1981; Wojciszke & Klusek, 1996), and National stereotypes (Alexander, Brewer, & Hermann, 1999; Phalet & Poppe, 1997; Poppe & Linssen, 1999). In sum, the variety of labels reflects the range of research areas and focus from which the two dimensions emerged.

However, despite different labels and across different fields of psychology, these dimensions have similar content conceptualizations and their translation into concrete measure tends to converge. Indeed, aiming to demonstrate the redundancy of these different labels, Abele and Wojciszke (2007) collected ratings for three hundred traits on communal - agentic, masculine - feminine, competence - morality, and so on. Results of the factorial analysis highlighted a two-dimensional structure that explained 90% of the variance. The first factor included labels such as communality, collectivism, and morality; whilst the second factor included labels such as agency,

[2] The Stereotype Content Model is better described below in the gender stereotypes section.

individualism, and competence. Hence, researchers in different fields define and operationalize these two dimensions very similarly, especially the communion/warmth dimension (Abele, Cuddy, et al., 2008). The similarity is less strong between agency and competence because of the components of 'competence' and 'efficiency'; agency, in fact, also covers traits that refer to motivation and assertiveness (e.g., ambitious, goal-oriented) (Abele & Wojciszke, 2014; Cuddy et al., 2008).

The universality of agency/competence and communion/warmth as fundamental dimensions of social judgments has been proven in different U.S., European, and Asian samples, as well as in collectivistic and individualistic cultures (Abele, Uchronski, et al., 2008; Cuddy et al., 2009). According to Fiske and colleagues (2007), the relevance of these two dimensions derives from the fact that they assess two critical issues that are both basic and adaptive to survive and prosper in a social context (Cuddy et al., 2008). Firstly, a social actor needs to establish and anticipate the beneficial or harmful intentions of others; it is therefore sensible to evaluate another person or group's morality, sincerity, kindness, and so on. These inferences about others' beneficial or harmful intentions trigger approach-avoidance behavior. Secondly, it is functional to determine their ability to act on those intentions; to this end, it makes sense to estimate a person or group's intelligence, competence, effectiveness, and so on.

As mentioned above, Rosenberg and colleagues (1968) found that two dimensions characterize first impressions: socially and intellectually good-bad. In their seminal research though, a third dimension emerged from the analyses. In failing to denote it, they posited that this third dimension existed "within the context of the other two" (p. 291). In other words, they anticipated the existence of sub-components or facets within the two main dimensions underling first impressions.

**The prominence of morality**

The so called 'Big Two' (Paulhus & Trapnell, 2008) is a well-validated theoretical model, and evidence has shown that communion/warmth is more important than agency/competence when

forming first impressions about others (Abele & Wojciszke, 2014; Brambilla & Leach, 2014). For instance, people are more interested in gathering information on communion than agency (De Bruin & Van Lange, 2000) and global impressions are better predicted from communion than agency traits (Wojciszke, Bazinska, & Jaworski, 1998).

However, the prominence of communion/warmth in the impression formation process can be better understood by distinguishing between morality and sociability as two distinct components of communion (Brambilla & Leach, 2014; Leach et al., 2007). Whereas sociability pertains to being benevolent to people in ways that favour warm relations with them (e.g., friendliness, likeability, kindness), morality refers to being benevolent to people in ways that favour correct and ethical relations with them (e.g., honesty, trustworthiness, sincerity). Leach and colleagues (2007) pointed out the prominence of morality in group perception showing that traits designed to indicate morality (e.g., sincere, trustworthy), competence (e.g., intelligent, skilled), and sociability (e.g., friendly, likeable) constitute different group characteristics and that morality is more important for a positive evaluation of the in-group than sociability and competence.

At interpersonal level, people are more interested in gathering information about other's morality rather than other's sociability or competence (Brambilla, Rusconi, Sacchi, & Cherubini, 2011). Specifically, participants looked for negative information about targets' morality, this because evidence of immorality is considered more diagnostic of people's moral character than evidence of morality (Skowronski & Carlston, 1987; Trafimow & Trafimow, 1999). Beyond the information gathering process, morality is also prominent in affecting certain people' behaviours, such as to approach and help or avoid others (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013; Brambilla, Sacchi, Menegatti, & Moscatelli, 2016; Iachini, Pagliaro, & Ruggiero, 2015; Pagliaro, Brambilla, Sacchi, D'Angelo, & Ellemers, 2013).

Although morality plays a prominent role in person judgment, it is important to note that contexts and interaction goals affect the type of information that people consider useful to gather (Wojciszke, 2005). Indeed, when the goal is to invite a person to a party, information concerning a

person's sociability is considered more relevant than information concerning that person's morality and competence (Brambilla et al., 2011). On the other hand, when the goal is to hire a person for a research programme, information concerning candidates' competence is considered more relevant than information concerning candidates' morality, which in turn is more important than sociability.

As stated above, two dimensions emerged from different fields with different labels. In face perception research, they are called: dominance and trustworthiness.

**The two underlying dimensions of facial first impressions**

One of the first works on first impressions was the Solomon Asch's article "Forming Impressions of Personality" published in 1946. This research examined how perceivers form global first impressions about a target's personality from fragmented pieces of information, provided as a list of traits like warm, polite, and so on.

Despite the undoubtable importance of this seminal work in uncovering information processing mechanisms that lead to first impressions formation, as pointed out by McArthur and Baron (1983), it is important to consider the role of the stimulus information. While people rarely form impressions of others on the sole basis of a list of traits, they usually form impressions of others based on their facial appearance. Indeed, faces are recognized as one of the most powerful and meaningful communication tools in social interactions (Jack & Schyns, 2015).

In social psychology, Paul Secord was one of the first researchers who directly investigated face-based first impressions (Todorov, 2017). In the 1950s, Secord and his colleagues found that there were high levels of consensus on many personality traits perceived from faces, such as friendliness, honesty and intelligence, and this consensus was cross-cultural (Secord, 1958; Secord, Bevan, & Dukes, 1953; Secord, Bevan, & Katz, 1956; Secord, Dukes, & Bevan, 1954). They also found that faces with comparable rating levels in related personality traits, such as honesty and trustfulness, shared similar facial features patterns, even if people were not able to specifically identify what physical features influenced their impressions. Curiously, these findings have been corroborated by more recent research on facial first impressions (Todorov et al., 2015).

In the 1980s, Leslie Zebrowitz and her colleagues started more systematic research on social inferences from faces (e.g., Berry & McArthur, 1986; McArthur & Baron, 1983; Montepare & Zebrowitz, 1998; for a recent review, see Zebrowitz, 2017). They found that specific features, such as larger eyes, higher eyebrows, thicker lips, short chin, higher forehead, and rounder and less angular faces, make faces appear as more 'babyish', and adults with these childlike facial features are perceived as naïve, submissive, physically weak, warm, kind and honest.

In the last decade, Todorov and colleagues have been conducting an extensive and compelling work on facial first impressions (e.g., Oosterhof & Todorov, 2008; Willis & Todorov, 2006; for an overview, see Todorov, 2017). In their article, Oosterhof and Todorov (2008) found that two dimensions underlay first impressions from faces, combining classical and original methodological approaches. Firstly, the two investigators collected 1,100 unconstrained descriptions of photos retrieved from the Karolinska Directed Emotional Faces (Lundqvist, Flykt, & Öhman, 1998). The photos are standardized for background, dimensions and faces depicted exhibit neutral expressions (i.e., no emotions). The descriptions were then collapsed into 14 trait dimensions that accounted for 68% of the description variance and then used as judgment criteria to rate the same faces. The ratings collected were eventually entered in a principal component analysis: the first principal component accounted for 63.3% of the variance. It was labelled the 'trustworthiness' dimension, because this trait had the higher loading, or 'valence' dimension, because all the positive traits (e.g., attractive, caring) had positive loading while all the negative traits (e.g., aggressive, unhappy) had negative loading. The second principal component accounted for 18.3% of the variance and it was called the 'dominance' dimension because this trait had the highest loading. Thus, trustworthiness and dominance emerged as the two-dimensional model of face judgment.

Employing a data-driven methodological approach (i.e., no a priori hypothesis), Oosterhof and Todorov (2008) have subsequently used the ratings collected to create computer models, in order to investigate how facial features could vary accordingly to the levels of social traits

perceived from faces. Looking at these computer-generated faces ranging from high to low levels of trustworthiness or dominance, it was possible to see a gender characterization. Indeed, trustworthy faces appeared as more feminine while untrustworthy faces look more masculine. The other way around was true for dominance, with dominant faces appearing as more masculine and submissive faces appearing as more feminine.

Sutherland and colleagues (2013) replicated these findings using different material, that is 1,000 photos retrieved from the Internet and highly variable in terms of facial characteristics (e.g., age, expressions) and image characteristics (e.g., lighting, background), and methodology, such as image averaging and morphing techniques. Results highlighted two dimensions (i.e., approachability and dominance) almost identical with trustworthiness and dominance.

A similar pattern about gender characterization of the two dimensions was also found (Sutherland et al., 2013). Indeed, averaging the 20 highest and lowest rated faces for dominance and trustworthiness, the high dominant face averages looked younger and more feminine, while low dominant face averages looked older and more masculine. The high trustworthy face averages, on the other hand, appeared older and more feminine, whereas low facial trustworthy face averages looked younger and more masculine.

Beyond dominance and trustworthiness, Sutherland and colleagues (2013) also found a third dimension: ‘youthful-attractiveness’. This original finding was probably due to the material: indeed, contrary to Ooostherof and Todorov (2008), in this research photos varied along age perceived from faces. Overall, many different findings have sustained the relevance of attractiveness on first impressions.

### *Facial attractiveness*

One of the most referenced articles on attractiveness was published by Dion, Berscheid, and Walster (1972), from which originates the famous sentence "what is beautiful is good" (p. 285). Investigators found that people attribute positive qualities to attractive strangers and negative qualities to unattractive strangers. These findings have been replicated showing that positive

qualities, such as being sociable, friendly, warm and intelligent, are more related to attractive people than less attractive individuals (Dion et al., 1972; Langlois et al., 2000). The effect of individual attractiveness has been defined as a 'halo effect', so that the perception of attractiveness lead to biased judgments of other positive and desired personality qualities (e.g., Costa & Corazza, 2006; Talamas, Mavor, & Perrett, 2016), but also as a stereotype, so that being attractive activates stereotyped expectations on which the target is evaluated (e.g., Heilman, 1983).

The importance of attractiveness has been explained referencing the evolutionary theory which links it with health and reproductive ability (e.g., Buss & Schmitt, 1993; Little, 2014). Specifically, all animals need to face two major issues, survival and reproduction, which lead to reproductive success, which is the key to evolution. As Darwin pointed out, sexual selection cues can have no specific function for survival, such as peacock tails, bird song, and attractiveness in humans, but they could indicate the possession of good genes. Hence, these cues are important for reproductive purposes, since proper mate-choice decisions mean more fecund and healthier offspring.

Evidence also suggests that attractiveness has a stronger link with women than man. According to evolutionary psychology, in fact, although it is important for both men and women, being attractive should be more important for women than for men, because of the criteria used in mate-selection (Buss, 2005). In fact, attractiveness is a criterion used in mate selection by both men and women, but while women use multiple criteria and place more attention on the earning capacity of a potential mate, men mostly based their mate choices on physical attractiveness (Shackelford, Schmitt, & Buss, 2005; Wiederman & Allgeier, 1992). Moreover, in face perception research, a feminine shape in both men and women's face leads to perceived attractiveness (Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000; Said & Todorov, 2011).

Overall, being attractive seems to have numerous benefits not only in mate-selection but in the everyday life. For example, mothers are more caring to attractive babies than unattractive babies (Langlois, Ritter, Casey, & Sawin, 1995), adults with attractive faces live longer (Henderson

& Anglin, 2003) and have more children (Pflüger, Oberzaucher, Katina, Holzleitner, & Grammer, 2012).

Most of all, attractiveness affects different job-related outcomes (Hosoda et al., 2003). For example, the same curriculum was evaluated more positively and having more probability to be selected when there was a highly attractive facial photo attached than a low attractive facial photo, or no photo at all (Watkins & Johnston, 2000). Furthermore, relative to less attractive people, attractive individuals tend to be perceived as having better job qualifications (e.g., Quereshi & Kay, 1986), more likelihood to be successful in a job (e.g., Morrow, McElroy, Stamper, & Wilson, 1990), and higher compensation (e.g., Frieze, Olson, & Russell, 1991).

While it is not difficult to accept that attractiveness affects real outcomes, it could be difficult to imagine that social traits inferred from faces, such as competence, can affect real-life outcomes, such as political elections.

***Consequences of facial first impressions***

In one of the most striking works on the consequences of facial first impressions, Todorov and his colleagues firstly collected competence ratings of politicians' faces running for American gubernatorial races, Senate races, and House of Representative races (Ballew & Todorov, 2007; Hall, Goren, Chaiken, & Todorov, 2009; Todorov, Mandisodza, Goren, & Hall, 2005). Then, they used those ratings to predict the results of the elections: surprisingly, they found that the more competent looking faces had more probability of winning elections, with competence ratings being able to predict around 70 % of results. These findings have been replicated by numerous studies (for a review, see Olivola & Todorov, 2010), even employing (i.e., Swiss) children ratings on foreigners (i.e., French) politicians to predict National elections (Antonakis & Dalgas, 2009).

The fact that people's behaviours and choices are affected by social attributions from faces has been proven by many lab and real-world experiments in different domains (Todorov et al., 2015), ranging from sentencing decisions (e.g., Porter, Brinke, & Gustaw, 2010) to economic transactions (e.g., Rezlescu, Duchaine, Olivola, & Chater, 2012). For instance, Porter and his

colleagues (2010) found that defendants with less trustworthy-looking faces were more likely to receive a guilty verdict.

However, the selection and the salaries of leaders is one of the most addressed research issues (Todorov et al., 2015). As mentioned above, numerous studies have confirmed the relevance of facial competence in the selection of leaders in politics, but other facial traits have been found to have a predictive power in political elections, such as dominant (e.g., Chiao, Bowman, & Gill, 2008; Little, Burriss, Jones, & Roberts, 2007), sociable (Castelli, Carraro, Ghitti, & Pastore, 2009), and threatening (Mattes et al., 2010; Spezio et al., 2008). Studies realized in the military domain have highlighted that a dominance appearance in cadets can predict their future rank attainment (Mazur, Mazur, & Keating, 1984; Muller & Mazur, 1997). Within the workplace domain, managers and CEOs with more competent and dominant faces tend to work for more successful companies and receive larger salaries (Fruhen, Watkins, & Jones, 2015; Graham, Harvey, & Puri, 2016; Rule & Ambady, 2008b, 2009).

There are a few factors that can interact with facial first impressions and their consequences, such as the context considered and the gender of the targets (Todorov et al., 2015). Social domains can play a role in determining a trait power to sway people's decisions. For example, although trustworthiness is the most important dimension of facial first impressions, within politics competence is the most relevant facial trait for candidates to exhibit: the more competent politicians' faces are, the more likely they are to win elections (Todorov et al., 2005; Olivola & Todorov, 2010). Thus, the tendency is to prefer one person over another on the basis of a perceived fit between traits perceived from their face and the most relevant trait in a specific domain.

Regarding the interplay between gender and facial inferences in swaying people's choices and behaviours, a few findings emerged from the leader appearance line of research (Fruhen et al., 2015; Poutvaara, Jordahl, & Berggren, 2009; Rule & Ambady, 2008b, 2009; Von Stockhausen, Koeser, & Sczesny, 2013). For example, while male CEOs' salaries are associated to facial competence (Graham et al., 2016), female CEOs' individual compensation is predicted by facial

dominance (Rule & Ambady, 2009). Notably, Graham and colleagues (2016) also found that more competent looking CEOs did not perform better than their peers perceived as less competent.

The idea of inferring individuals' personalities from their faces has fascinated people over the centuries from Aristotele, the author of the first document about physiognomy, to Francis Galton, a nineteenth century scientist who tried to make physiognomy a science (Todorov, 2017). Echoing Cesare Lombroso's claim, some studies have recently confirmed that people are accurate in detecting criminality from facial photographs (e.g., Porter, England, Juodis, Ten Brinke, & Wilson, 2008), as well as political (e.g., Rule & Ambady, 2010) and sexual orientations (Rule & Ambady, 2008a; Rule, Ambady, & Hallett, 2009). However, more than one limitation[3] in these studies has been pointed out, such as the quality of the photos (Todorov et al., 2015). For instance, in sexual orientation research (Rule & Ambady, 2008a; Rule et al., 2009), facial photos are usually retrieved from online dating websites. Thus, the fact that sexual orientation can be detectable in those photos may just mean that users chose and posted effective profile pictures to convey their purposes (Todorov et al., 2015). A similar argument can be made on research about facial criminality, since they provided participants with America's Most Wanted photos and Nobel Peace Prize winners photos (Porter et al., 2008), for instance. Hence, conclusions about accuracy can be considered solid when it is possible to control the material (e.g., background, expression, clothes), in order to control possible confounding variables (Todorov et al., 2015).

Although facial first impression and the fundamental contents of social judgments have been treated as two different research fields, few points in common emerged.

**Convergences between social and facial models**

Oosterhof and Todorov (2008) in their article suggested that trustworthiness and dominance could correspond to warmth and competence, respectively (Fiske et al., 2007). For instance, both trustworthiness and warmth concern perceived intentions to help or harm, whereas both dominance and competence refer to the perceived ability to carry out those intentions. Moreover, as Abele and

[3] Although it is important, the accuracy issue is beyond the scope of the current dissertation. The interested reader is referred to Todorov (2017) and Todorov et al. (2015).

Wojciszke (2007) found on communion/warmth dimension compared to agency/competence, in both Oosterhof and Todorov (2008) and Sutherland and colleagues (2013) 's works, the first factor extracted and identified as trustworthiness explained a larger amount of variance than the other factor representing dominance. Furthermore, various affinities exist between facial trustworthiness and morality, one of the two sub-components of communion/warmth dimension. Firstly, 'trustworthy' is a trait that falls into morality. Secondly, both facial trustworthiness and morality prompt approach-avoidance behaviour and are linked with threat feelings. Indeed, when a social target is perceived as immoral (i.e., dishonest and untrustworthy), he or she is seen as a threat to the integrity and stability of the community (Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Brambilla et al., 2013; Brambilla, Biella, & Freeman, 2018). Similarly, the process of inferencing trustworthiness from faces implies the activation of a subcortical area of the brain called amygdala, which is involved in the detection of potentially threats (Engell, Haxby, & Todorov, 2007; Todorov, Mende-Siedlecki, & Dotsch, 2013; Todorov, Said, Oosterhof, & Engell, 2011). Thirdly, facial trustworthiness has a prominent role in first impression processes, alike morality has (e.g., Brambilla & Leach, 2014; Todorov, 2008). For instance, trustworthiness ratings are formulated faster than any other trait, from 20ms to 33ms, and it provides a good approximation of the overall judgment of faces because 'trustworthy' is highly correlated with many other facial traits (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006).

Starting from this evidence, Sutherland and colleagues have experimentally tested a possible correspondence between the two fundamental dimensions of social cognition and the two dimensions underlying face perception (Sutherland, Oldmeadow, & Young, 2016). Across four studies and employing different methodologies, they found that trustworthy face averages correspond to warm face averages. Dominance and competence, instead, appeared to be less related especially in female face averages, suggesting that they may represent two different dimensions.

While the distinction between morality and sociability is established in social cognition research, the fact that competence and dominance could not be conceptually equivalent had been

suggested (Abele & Wojciszke, 2014; Carrier, Louvet, Chauvin, & Rohmer, 2014) and only recently demonstrated by Abele and her colleagues (2016). Employing a factorial analysis method, they found that the classic two-factor model showed a good fit. In line with the investigators' expectations though, a better fit was provided by a four-factors model, which are the so-called 'facets' of the two fundamental content dimensions: competence and dominance as facets of the agency/competence dimension; morality and sociability as facets of communion/warmth dimension[4].

In a more recent 'cross-over' research, Oliveira, Garcia-Marques, Garcia-Marques, and Dotsch, (2019) studied the relationship between The Big Two associated traits and valence (i.e., positive, negative) using facial photos. They basically replicated Sutherland and colleagues' (2016) findings, supporting the idea that it is maybe more reasonable to consider The Big Two as composed of different 'branches' (i.e., facets), and suggested that each facet could be classified not only on the basis of the relationship between their related traits but also on their (positive or negative) valence.

It is worth noting that the two lines of research converge on the distinction between competence and dominance, whilst findings diverge on the distinction between morality and sociability. As mentioned above, morality and sociability are widely recognized as sub-components of communion/warmth dimension in social cognition research (e.g., Leach et al., 2007). On the other hand, this distinction fails to be detected in facial first impressions research as Sutherland and colleagues (2016) specifically reported they had not found morality and sociability as two different facets of the trustworthiness dimension (see p. 265). However, one could argue that the distinction did not come out because of a lack in the methodology: indeed, the photos had not been rated on traits related to sociability. In other words, if the list of social traits used to evaluate photos during

[4] These facets labels were chosen in order to favour readers' comprehension and avoid a long discussion on different name labels, only briefly mentioned above. However, Abele and colleagues (2016) used assertiveness and warmth instead of dominance and sociability, respectively.

the experiments had also included traits such as ‘kindness’ or ‘friendliness’, maybe the distinction between facial morality and facial sociability would have been emerged from the results.

Finally, social and facial models have one more point in common: gender characterization. On one hand, as reported above, trustworthiness and dominance are associated with a more feminine and a more masculine face appearance respectively (e.g., Sutherland et al., 2013) and the trustworthiness dimension better explains first impressions inferred from female faces than male faces (Oh, Dotsch, Porter, & Todorov, 2019). A masculinity-femininity appearance was also attributed to computer generated faces that vary on competence (Olivola & Todorov, 2010). Specifically, as the perceived competence increased, the artificial faces showed: a less round shape, a reduced distance between the eyebrows and the eyes, higher cheekbones and less angular jaws; facial features more typical of male adults (McArthur & Apatow, 1984). Indeed, participants rated highly competent faces as more masculine and mature (Olivola & Todorov, 2010).

On the other hand, agency and communion have been adopted by gender literature, given the two dimensions related to masculinity and femininity (e.g., Abele, 2003) and have been linked to gender stereotypes (e.g., Eagly & Steffen, 1984). Competence and warmth, as mentioned above, have been used to explain many different stereotypes (i.e., the Stereotype Content Model) and also gender stereotypes.

**Gender stereotypes**

Stereotypes are expected traits or attributes applied to social groups and, consequently, to all the members of those groups (e.g., Agars, 2004; Ellemers, 2018). As a pervasive cue of social categories, gender easily, rapidly and automatically triggers stereotypical thinking (e.g., Banaji & Hardin, 1996; Banaji, Hardin, & Rothman, 1993; Blair & Banaji, 1996).

Psychological research on gender stereotypes began in the 1950s (e.g., McKee & Sherriffs, 1957), and intensified during the 1970s. Thanks to Broverman, Rosenkrantz, and their colleagues’ studies (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968) and Spence and Helmreich’s work (1972), it was possible to

reveal how the majority of the stereotyped beliefs about men and women was properly represented by the two aforementioned fundamental content dimensions: agency/competence and communion/warmth. Indeed, women are usually associated with the concern for others and described as being helpful, kind, nurturing, likeable, etc…, traits that fall into the communal/warmth dimension (e.g., Eagly & Karau, 2002, Cuddy et al., 2007). On the other hand, men are more commonly associated with being assertive, ambitious, efficient, capable, etc…, traits that fall into the agency/competence dimension (e.g., Eagly & Karau, 2002, Cuddy et al., 2007).

Compared to men, women are evaluated higher on communion/warmth and lower on agency/competence; whereas men are evaluated lower on communion/warmth and higher on agency/competence (e.g., Fiske, 1998; Heilman, Block, & Martell, 1995). Moreover, people tend to evaluate women more positively than men (e.g., Eagly & Mladinic, 1989, 1994; Williams & Best, 1990), a phenomenon called "women are wonderful effect" which has been found ubiquitous across cultures (Glick et al., 2000, 2004). This phenomenon is also in line with benevolent and hostile sexism research, which conceptualization basically reflects the well-known virgin-whore dichotomy (Glick & Fiske, 1997, 2001). Benevolent sexism, indeed, implies an idealization of women seen as pure and in need of protection and support. On the other hand, hostile sexism is a more negative view of women seen as in trying to control men with feminist ideology or sexuality. Note that, despite the opposite valence of the attitudes toward women, both kinds of sexism stem from the same assumptions (e.g., women are the weaker sex) and serve to justify and maintain traditional gender roles and status.

Regarding gender roles, the Social Role Theory is one of the oldest theories that explained sex differences and similarities in social behaviour (e.g., Eagly, 1987; Eagly & Wood, 2011; Wood & Eagly, 2012). Initially developed by Alice Eagly in the 1980s and inspired by psychological and sociological works, this biosocial theory differently emphasized communion versus agency in the explanation of the 'homemaker–provider' division of labour (Eagly & Steffen, 1984). In particular, the division of labour reflects the specialization of women and men "in activities for which they are

physically better suited" (Eagly & Wood, 2011, p. 465). For instance, women's reproductive abilities of pregnancy and lactation represent constrains that do not allow women to participate as fully as men in tasks such as ploughing or hunting. Instead, these tasks better fit men's qualities of greater size and strength. Therefore, social role beliefs derive from people's inferences about group members observed in those social roles in which they are overrepresented (Eagly & Wood, 2011). For example, given women are more often observed in domestic behaviours, such as cooking and childcare (Gawronski, 2003; Gilbert, 1998), perceivers' correspondent inferences are that women possess those communal traits which enable these behaviours, such as nurturance and warmth (Eagly & Wood, 2011). By contrast, men are usually engaged in paid roles that require assertive (i.e., agentic) behaviours. Interestingly, the members of a society tend to '*essentialize*' the division of labour, so that it is perceived as an inevitable and natural consequence of inherent differences between men and women.

The other widely mentioned theory that has tried to explain how gender stereotypes emerge is the Stereotypes Content Model (e.g., Fiske et al., 2007). While the Social Role Theory was primarily proposed as an explanation of gender stereotypes (but for an extension to other social groups, see Koenig & Eagly, 2014), the Stereotypes Content Model aims to explain any kind of stereotype. The Stereotypes Content Model and its extension called BIAS map (Behaviours from Intergroup Affect and Stereotypes) assume competence and warmth as contents of the stereotypes, that are judgment dimensions which underlie and differentiate the perception of other individuals or groups (Cuddy et al., 2007, 2008). From Fiske and colleagues (2007, 2008) point of view, the antecedents of warmth and competence judgments stem from two structural variables: interdependence/competition and status (i.e., economic success and prestigious job) respectively. Indeed, if people share a goal and cooperate, then their intent is perceived as warm (e.g., friendly and trustworthy). On the other hand, when people have conflicting goals and compete among each other, their intent is perceived as not warm (e.g., hostile and dishonest). Moreover, high competence judgments result from perceived high-status, as well as low-status drives incompetence judgments,

and the more people endorse hierarchical systems or believe in a just world, the more status and competence are correlated (Oldmeadow & Fiske, 2007).

The link between status and competence judgments was also argued by the Double Standards Theory (e.g., Foschi, 1992, 2000; Foschi & Foddy, 1988), as an advance of the broader Expectation States Theory (e.g., Berger, Fisek, Norman, & Zelditch Jr, 1977; Berger & Zelditch Jr, 1998; for an overview, see Correll & Ridgeway, 2003). The prediction of these sociological theories has been incorporated in the Shifting Standard Theory (e.g., (Biernat, 1995; Biernat, Manis, & Nelson, 1991), which explains how judgment standards can change according to stereotypes attributed to the social category in question.

***The Shifting Standards Model***

The Shifting Standard Model (SSM) posits that stereotypes of groups set standards of judgments against which members of the groups are compared to. Monica Biernat (1995) specifies what she means with stereotypes and standards. Stereotypes provide a cognitive representation of the characteristics attributed to groups, but they also imply explicit or implicit comparisons between groups (e.g., men are more competent than women). From this point of view, status is seen as a less variable quality of stereotypes: while each group have different stereotype contents (e.g., aggressiveness, height, etc…), any group is stereotyped as deficient in some characteristics relative to another group. Standards are criteria on which evaluations are based and derived from expectations (e.g., we expect for men to be tall), but also from a range of possible results or behaviours (e.g., a man should be from 1.75 m to 1.85 m tall) (Biernat, 1995).

Accordingly, the model makes two predictions: in judging group members considered deficient on an attribute, people can set low-minimum standards or stringent confirmatory standards (Biernat, 2012). In the first case, for instance, if people hold the belief that women are less competent than men, expectations of competence for women are lower than for men (Biernat & Kobrynowicz, 1997). Thus, judges will need less evidence of competence from a woman for evaluating her as competent (i.e., ‘she is very skilled, for a woman’). However, lower expectations

for negatively stereotyped groups can make judges set higher standards. In this case, for example, women need to provide more evidence of their ability than men in order to be considered as competent as men (Foschi, 2000). Applied in a work setting, more lenient standards for women can make them more likely than men to be shortlisted for a job; but they would be less likely hired than men for the same job (Biernat & Fuegen, 2001).

Concerning job applicants, the fact that gender stereotypes can have a role in the workplace has been recognized for some time. For instance, in 1983, Deaux and Lewis asked participants to rate how masculine and feminine linked traits (e.g., independent, emotional), role behaviours (e.g., financial providers, tends the house), physical characteristics (e.g., deep voice, good-looking), but also occupations (e.g., bank teller, secretary) applied to a man, a woman, or a person. Results showed that masculine and feminine components had a significantly stronger association with men and women, respectively, a pattern that was found persistent in data collected in 2014 (Haines, Deaux, & Lofaro, 2016). For example, social roles such as 'assumes financial obligations' and 'is a leader' were more associated with men than women, and jobs like 'secretary' or 'administrative assistant' were more associated with women than men.

**Gender stereotypes in the workplace**

Stereotypes are descriptive and describe how men and women are, but they can also be prescriptive and proscriptive of behaviours that men and women should and should not perform in their social roles (e.g., Eagly, 1987; Rudman, Moss-Racusin, Phelan, & Nauts, 2012). For example, men should be assertive, independent, possess business sense and leadership ability, but they should not be emotional, naïve, weak, or insecure (Rudman et al., 2012). Conversely, women should be warm, sensitive to others, friendly and supportive, but they should not be dominating, self-centred, ruthless, or rebellious.

While, according to the theory of ambivalent sexism, women are evaluated more positively when stereotyped into specific traits and roles (e.g., being helpful) (Glick & Fiske, 1997, 2011), they can face social costs when these rules are violated (e.g., Rudman & Glick, 2001; Rudman et al.,

2012; M. J. Williams & Tiedens, 2016). For example, women who display explicit counter-stereotypic behaviours (e.g., being assertive) are disliked and can be discriminated against in selection processes, compared to males and females who show stereotypic behaviours. Something like this so-called 'backlash effect' has been also found in face perception research by Sutherland, Young, Mootz, and Oldmeadow (2015), with dominant female faces (i.e., counter-stereotypic appearance) rated more negative than stereotypic female faces and their male counterparts. Similarly, Todorov and colleagues recently posited that women are negatively evaluated when their facial appearance does not conform to gender stereotypic expectations, as well as that competence, confidence, and masculinity cues in real-life images of female faces cause a 'backfiring effect' (Oh, Buck, & Todorov, 2019; Oh, Dotsch, et al., 2019).

More broadly, a 'gender bias' is activated when people use stereotyped beliefs to judge men and women, such as considering men for masculine jobs (e.g., engineer, carpenter) and women as more suitable for feminine jobs (e.g., nurse, telephone operator) (Davison & Burke, 2000; but on this latter point, see Koch, D'Mello, & Sackett, 2015).

***Gender bias: Lack of Fit Model and Role Congruity Theory***

The Lack of Fit Model suggests that the activation of both gender stereotypes and the gendered perception of the job itself affect the occurrence of gender bias (Heilman, 1983, 2001; Heilman et al., 1995). Specifically, expectations about the future successful or unsuccessful individual's performance in a specific job strongly influence personnel decisions. These expectations originated from a perceived fit between skills and abilities attributed to a person and skills and abilities required to succeed in a specific position.

Hence, if an individual's stereotyped attributes reflect job requirements, the perceived fit will be positive. For example, since women are described as warm, gentle, and friendly (Cuddy et al., 2007; Fiske 1998; Rudman et al., 2012), feminine traits that are considered necessary to perform well as nurse or secretary, they tend to be favoured for these occupations (Davison & Burke, 2000; Heilman, 1983). Conversely, if an individual's presumed attributes do not reflect job requirements,

the perceived fit will be negative. For example, given that it is considered essential for managers to be dominant and assertive (Gaucher, Friesen, & Kay, 2011), traits encompassed in male stereotype content (Cuddy et al., 2007; Fiske 1998; Rudman et al., 2012), men will be preferred over women for higher level occupations (Schein, 2001). In any case, whether positive or negative, these perceived fit-based performance expectations profoundly affect job evaluation processes.

Analogous conclusions have been proposed by the Role Congruity Theory (Eagly & Karau, 2002), an extension of the Social Role Theory previously described (e.g., Eagly, 1987). Role Congruity Theory considers the congruity between gender roles and other roles, especially in leadership. Specifically, prejudice toward female leaders can arise from the perceived discrepancy between the predominantly communal qualities associated to women and the predominantly agentic qualities that people believe necessary to succeed as a leader.

In other words, people hold similar stereotyped beliefs about the link between leadership roles and men and this so-called 'think manager – think male' phenomenon was originally uncovered by Schein (1973, 1975, 2001, 2007), more recently replicated (Jackson, Engstrom, & Emmers-Sommer, 2007; Ryan, Haslam, Hersby, & Bongiorno, 2011), who also found that it was particularly held by men in many different countries, such as the United States, the United Kingdom, Germany, and Japan (Lee & Hoon, 1993; Schein, 2001). These findings are also in line with literature showing that in comparison to women, men are more likely to hold traditional stereotyped beliefs about women (e.g., passive, timid; Massengill & Di Marco, 1979), to see gender egalitarianism less positively (e.g., Eagly & Mladinic, 1989; Spence & Hahn, 1997), and to attribute higher selection ratings to male applicants than female applicants (Davison & Burke, 2000).

In social and organizational psychology, there is abundant empirical evidence consistent with these claims about the influence of a perceived fit or congruency between job and gender stereotypes (Davison & Burke, 2000; Koch et al., 2015). For example, a meta-analysis of 49 studies revealed that women receive lower selection ratings and lower compensation offers than men for male type-jobs. The opposite emerged for female stereotyped jobs, with women favoured over men,

whether no gender bias emerged for gender neutral stereotyped jobs (but on these points, see Koch et al., 2015).

Similar findings have been found in appearance and face perception research. First of all, Heilman (1983) used the Lack Fit Model to explain the 'beauty is beastly effect', which is attractive female applicants evaluated as less hireable and as less likely to be promoted in male-typed or managerial jobs than unattractive women and male counterparts (Heilman & Saruwatari, 1979; Heilman & Stopeck, 1985). Although being attractive has its advantages, attractiveness in women can highlight their femininity, which in turn activates female stereotype contents, according to which women are not agentic (Heilman, 1983; Paustian-Underdahl, Walker, & Woehr, 2014). Thus, attractiveness creates negative expectations about women on their suitability for masculine kinds of jobs. The Lack of Fit Model has also been evoked in a study in which, only on the basis of their photos, gay men were considered as more suitable for feminine jobs than masculine jobs (Rule, Bjornsdottir, Tskhay, & Ambady, 2016). Indeed, since they are stereotyped as feminine (Kite & Deaux, 1987), compared to straight men, gay men were considered as more suitable for nurse and paediatrician jobs than engineering and management jobs. Finally, Imhoff and his colleagues (2013) created typical faces of nursery teachers (i.e., feminine job) and typical faces of managers (i.e., masculine job) collecting participants' judgments with a reverse correlation image classification task. Replicating Stereotype Content Model findings (e.g., Fiske et al., 2008), typical faces of nursery teachers resulted as more feminine, warmer and less competent (i.e., female stereotype content) than typical faces of managers. Similarly, Oldmeadow, Sutherland, and Young (2013) created face models for four job occupations, bankers, nurses, teachers, and drug dealers as representative of the four quadrants obtained from the combination of high versus low competence and warmth judgments. The resulted face averages of bankers and drug dealers occupations looked masculine, while face averages of teachers and nurses occupations looked feminine.

Beyond job characteristics, the context can have a role in raising or reducing the perceived levels of gender typicality of a job (Heilman, Manzi, & Braun, 2015): a managerial position at a

financial services firm is probably perceived as more masculine than a managerial position at a day-care centre. An indirect support of this claim can be the underrepresentation of women from lower to higher levels of hierarchy in more masculine-type organizational environments, which are science, technology, engineering, and mathematics (STEM) and financial firms (Adams & Kirchmaier, 2016). Indeed, according to statistics, there are on average 1.8% fewer women on corporate boards than non-STEM firms. A more direct proof has been reported by Lyness and Heilman (2006): in a large financial services company, women received less positive evaluations than men if employed in departments such as business management and operations management, but not when employed in departments such as human resources or external affairs.

So far, it was presented the established literature, but recent research has highlighted a few changes on gender stereotypes and gender discrimination at work.

**Recent evidence on gender stereotypes and gender bias**

Leach, Carraro, Garcia, and Kang (2017) examined women's gender stereotypes of violence, strength, competence (i.e., characteristics of the global dimension of agency/competence dimension), trustworthiness, and sociability (i.e., characteristics of the global dimension of communion/warmth). In line with previous evidence of women being seen as warmer than men (e.g., Williams & Best, 1990), they evaluated themselves as more sociable and trustworthy than men, with trustworthiness playing a central role in shaping implicit in-group positivity. Investigators also found that women perceived men as markedly superior only on violence whilst, contrary to the classical literature, women were implicitly portrayed as more competent than men.

In line with these findings, Hentschel and colleagues (2019) have employed different facets of agency and communion in order to investigate contemporary gender stereotype content. To this aim they asked more than 600 American female and male participants to rate men, themselves or women on: assertiveness, independence, instrumental competence, leadership competence (facets of agency/competence dimension), concern for others, sociability and emotional sensitivity (facets of communion/warmth dimension). Results showed that participants rated men and women evenly

high on instrumental competence. Moreover, male participants rated women as less agentic than men, while female participants rated women as less assertive but equally independent and leadership competent than men. Participants also rated themselves mostly in a more stereotypic way, that is, compared to female participants, male participants judged themselves reporting higher ratings on agentic traits and lower ratings on communal traits. However, there was no differences in female and male participants' self-ratings on instrumental competence, independence, and sociability.

Similar conclusions about the competence dimension can be made on the results obtained by Eagly and colleagues (2019). They run a meta-analysis on over 30,000 people's ratings retrieved from the integration of 16 different pools sampled from 1946 to 2018 and meant to be representative of the USA public opinion. In each included pool, there was at least one item that asked respondents whether each of the communion (e.g., affectionate, honest), and/or agency (e.g., ambitious, confident), and/or competence (e.g., creative, intelligent) characteristics is more true of women or men, or equally true of both. Exploring how gender stereotypes have changed over decades, compared to men, the ascription of communion to women has clearly increased, while no changes over time has been registered in the higher association of agency to men than women. However, women have increased in competence ascriptions, mostly among who note gender difference in competence: indeed, among these groups both female and male respondents evaluated women as the more competent sex.

Overall, these findings suggest that gender stereotype contents along the facets of the agency dimension have been changing. In considering different facets of the communion dimension, recent works has highlighted also novel mechanisms underlying gender bias in the workplace.

***The perfection bias***

Moscatelli and her colleagues (2020) have studied for the first time whether evaluations along competence, morality and sociability dimensions relate to gender bias in the workplace. The aim was to examine how information about male and female job candidates informed hiring and

retention decisions in real and laboratory contexts. In particular, real reports of professional personnel selector were content-analysed (Study 1). The relevance of the content dimensions was investigated in selecting male and female candidates for an organization (Study 2) or a work team, as well as the relative weight of competence and morality in predicting actual decisions to hire (Study 4) or retain (Study 5) male and female candidates with identical profiles. Across five studies, results showed that for male candidates, competence was the only predictor of employment decisions; evaluations on female candidates, instead, were based on multiple criteria. In other words, while men were only judged on their competence, women were evaluated on all the dimensions considered, especially on those dimensions on which they appeared to be relatively weak. As investigators stated, this evidence suggests that people adopt two different approaches depending on candidates' gender. When individuals evaluate male candidates, they search information that confirm or not the stereotypic belief about men being competent. When individuals evaluate female candidates, they are affected by an implicit assumption: women face more difficulties than men in the workplace, such as less easily being accepted and respected by their colleagues (van den Brink & Benschop, 2011). This assumption might lead evaluators to carry out a more systematic and comprehensive judgment process on female candidates, which is likely to make evaluators focusing on women's weaknesses and, as a consequence, to let them find a reasoning for preferring male candidates over female candidates.

This phenomenon labelled as 'perfection bias' has been found in another work which investigated the occurrence of competence, morality, and sociability content dimensions in real professional committees' evaluations of work performance (Prati et al., 2019). Consistent with Moscatelli and colleagues (2020), findings showed that male employees were mostly judged on their competence, whereas female employees were judged on all the three criteria considered.

Despite the novelty of these findings, in this research only written descriptions of applicants and workers have been used as stimuli, not faces. Moreover, competence, morality, and sociability were employed as judgment dimensions, but recent evidence has highlighted four facets of the

fundamental dimensions of social judgment: competence, morality, sociability and dominance (Abele et al., 2016).

## The present research

The key goal of the present research was to verify the presence of a perfection bias at face perception level, which is women being evaluated on more (facial) criteria than men. As reviewed above, traits inferences from faces can affect different outcomes in diverse domains (e.g., Todorov et al., 2015). However, it has not been tested yet whether women are evaluated in a hiring process on more facial traits than men.

To this end, four experiments were conducted. First, it was investigated whether facial attractiveness, competence and morality perceived from male and female applicants could differently affect Italian students' selection evaluations to a gender-neutral job occupation (Study 1). In the subsequent three experiments, Italian and British students (Study 2) as well as Prolific workers (Study 3, 4) were recruited to evaluate male and female applicants' hireability for leadership job positions. Compared to Study 1, these three experiments have three other novel aspects. Firstly, saying that women are evaluated on more criteria than man implies that all the criteria in question should be more important in evaluating female applicants than male applicants. Therefore, the relative importance of judgment criteria in affecting men and women's likelihood of being selected was directly investigated with two prestudies on the job descriptions employed in the main experiments and running specific analyses (e.g., Dominance analysis; Budescu, 1993). Secondly, along with facial attractiveness, facial competence and facial morality, also facial dominance and facial sociability were considered as judgment criteria. To the best of the author's knowledge, this represents the first empirical work on gender discrimination in a work context that investigated the predictive power of all the four facets of the fundamental content dimensions perceived from faces (Abele et al., 2016). Thirdly, a non-common methodology was employed: facets ratings (i.e., independent variables) and hireability ratings (i.e., dependent variable) of applicants' facial photos were collected in two different experiments. In particular, as a first step, Italian students were asked to rate facial photos on attractiveness, competence, dominance, morality, and sociability (Study 2, pre-test photos); these ratings were collected as measures of

general facial first impressions on the applicants. The same photos were subsequently rated on hireability for an Area Sales Manager position (Study 2, 3) and a Finance Manager position (Study 4). Then, the facets ratings (i.e., general facial first impressions ratings) were used to predict hireability ratings, in order to investigate the relative importance of the four facets perceived from female and male applicants' faces in affecting their hireability (for a similar methodology, see for example Oldmeadow et al., 2013). This methodology allowed a more implicit approach to the investigation of how 'general' (i.e., no context specific) facial first impressions affect selection evaluations about men and women.

Indeed, as reviewed above, the context can influence the relevance of trait dimensions and competence had been recognized as the most important trait in a work setting, by social cognition research (e.g., Brambilla et al., 2011), face perception research (e.g., Rule & Ambady, 2008b) and gender stereotype at work research (e.g., Moscatelli et al., 2020). Therefore, it was expected that competence would be the most important facial facet in predicting both male and female applicants' selection. However, according to Moscatelli and colleagues (2020), along with competence, all the other facial traits should be relevant in selecting women. Specifically, it was expected that attractiveness, competence, morality (in Study 1), but also dominance and sociability (in Study 2, 3, and 4) would be important in selection evaluations about female applicants, while only competence would be important in selection evaluations about male applicants (The Perfection Bias Hypothesis).

The hypothesis was based on previous work in which dominance had not been considered as a judgment dimension in the experiments (Moscatelli et al., 2020; Prati et al., 2019). Obviously, it was impossible that dominance, along with competence, would have emerged as an important predictor of male applicants' hireability. On the contrary, given the link among dominance, male stereotype contents and beliefs on leadership, it would be possible in this research (i.e., in Study 2, 3 and 4). Indeed, dominance is not only seen as an essential attribute for successful managers (e.g., Schein, 2001), but dominance traits (e.g., confident) also fall into the agency content dimension

(e.g., Abele et al., 2016), the same content dimension that represents the male stereotype (e.g., Eagly & Karau, 2002). Hence, the possibility that both dominance and competence would be relevant for male applicants' hireability to leaderships position in Study 2, 3 and 4 was not excluded.

# CHAPTER II

## Study 1

Study 1 tested the presence of a perfection bias considering attractiveness, competence and morality as judgment dimensions perceived from applicants' faces for a job position pre-tested to be perceived as suitable for both women and men. According to the perfection bias hypothesis, we expected that all the dimensions considered would be relevant in judging female applicants, while competence would be the only predictor of hiring decisions on male applicants. Although competence dimension is important in a work context, social cognition research (e.g., Brambilla & Leach, 2014) and facial first impression research (e.g., Oosterhoff & Todorov, 2008) showed that the most important dimension in first impression process is morality or trustworthiness, respectively. Moreover, morality is also a facet of the communion/warmth dimension, which represents the female stereotype content. Indeed, gender can also interact with facial first impression in predicting real outcomes (Todorov et al., 2015). For example, company rank and/or profits correlate with agentic traits perceived from male CEOs' faces, but they correlate with communal traits perceived from female CEOs' faces (Pillemer, Graham, & Burke, 2014). In trying to address these ambiguous findings, as additional hypothesis, Study 1 also tested the role of global impressions in explaining as mediator the influence of facial traits perceived from male and female faces in a work setting. In line with the perfection bias hypothesis, it was expected that global impressions on female applicants would be able to explain the influence of all the facial traits on hiring decisions. Conversely, for male applicants, global impressions would mediate only the influence of competence, given it was expected as the only significant predictor of hiring decisions.

### Method

#### *Pre-test: job description*

A pre-test was run recruiting 34 university students (14 men, 20 women; $M_{age}$ = 22.68, $SD_{age}$ = 2.57) to test whether the job description was equally suitable for men and women. The position was a temporary, part-time position in the administrative office and it was chosen as students are

likely to know it better than most non-academic jobs. The tasks of the job were data entry, making copies of teaching material, managing reservation of teaching and meeting rooms, administering the news section on the website. Results of one sample t-test supported that this position was perceived as suitable (1 = *not at all*; 7 = *very much*) for both women ($M$ = 4.98, $SD$ = 1.06) and men ($M$ = 4.97, $SD$ = 1.06), $t(33) = .10$, $p$ =.919, 95% CI [-.13, .14].

***Participants***

Participants were 221 students (88 men, 106 women, 7 participants did not report their gender; $M_{age}$ = 21.44, $SD_{age}$ = 3.14) of the University of Bologna, who completed the questionnaire at the end of classes. Half of them evaluated male applicants ($n$ = 111) and the other half female applicants ($n$ = 110). It is commonly recommended that SEM models incorporating latent variables require a sample size of at least 200 participants to be accurate or that the ratio for sample size to estimated parameters should be 5:1 (Kelloway, 2014). Given that the model has 36 parameters (64 for multigroup analyses), it was collected a sample size larger than 200 and in between the two ratios.

***Procedure and material***

Participants were asked to imagine they were a member of the Teaching Board of their Department, which, by statute, is composed of equal numbers of professors and students. The Board had to select a student ("the applicant") for the pre-tested temporary, part-time position in the administrative office. Participants were then asked to evaluate an applicant on the basis of the brief CV that s/he sent. The CV was the same for all applicants and reported basic information in Europass format: name, age 21, nationality, undergraduate student of the Department of Psychology, marks average 27/30, English level B1, good digital competence with Office package. A photo of the applicant was attached to the top-left corner of the CV. Faces with neutral expressions were retrieved from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998). The person depicted are Caucasian, wear a grey t-shirt with no jewellery, piercings, or other marks. The photos

were chosen on the basis of the levels of trustworthiness, intelligence (which fall into the morality and competence dimensions, respectively; Leach et al., 2007), and attractiveness attributed to the faces by Oosterhof and Todorov's (2008) study. In order to use faces with varying levels of trustworthiness, intelligence, and attractiveness, 16 photos were selected (8 males and 8 females) distributed along the quartiles of Oosterhof and Todorov's scores (Figure 1). After reading the CV, participants were asked to rate (1 = *not at all*; 7 = *very much*) to what extent the person portrayed in the photo looked: "honest", "moral" (facial morality; α = .77), "intelligent", "competent" (facial competence; α = .73); "good-looking" and "attractive" (facial attractiveness; α = .89).

**Figure 1.** Photos employed in the experiment and retrieved from the KDEF (Lundqvist et al., 1998).

Morality was assessed by means of the traits "honest" and "moral" because the meaning of "trustworthy" in Italian is closer to reliable and can be intended as strictly

related to competence, especially in the work domain. In contrast, "honest" is undoubtedly a trait of the moral domain (e.g., Brambilla et al., 2011). Then, it was measured participants' global impression on each applicant (1 = *very negative*; 7 = *very positive*) and the hiring decision by means of two items (α = .87): "In your opinion, how likely is it that the applicant would be selected for the job?", and "Would you select the applicant?" (1 = *very unlikely*; 7 = *very likely*). Finally, participants filled in the demographic information form.

## Results

### *Preliminary analyses*

Descriptive statistics and correlations among variables are reported in Table 1.

**Table 1.** Means, standard deviations, and correlations among all study variables.

| | M | SD | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **1. Morality** | 4.14 | 1.03 | .30*** | .31*** | .40*** | .26*** |
| **2. Competence** | 4.55 | 0.98 | - | .32*** | .60*** | .57*** |
| **3. Attractiveness** | 2.66 | 1.27 | | - | .48*** | .36*** |
| **4. Impression** | 4.16 | 1.10 | | | - | .71*** |
| **5. Hiring decision** | 4.07 | 1.51 | | | | - |

**Note.** *** $p < .001$

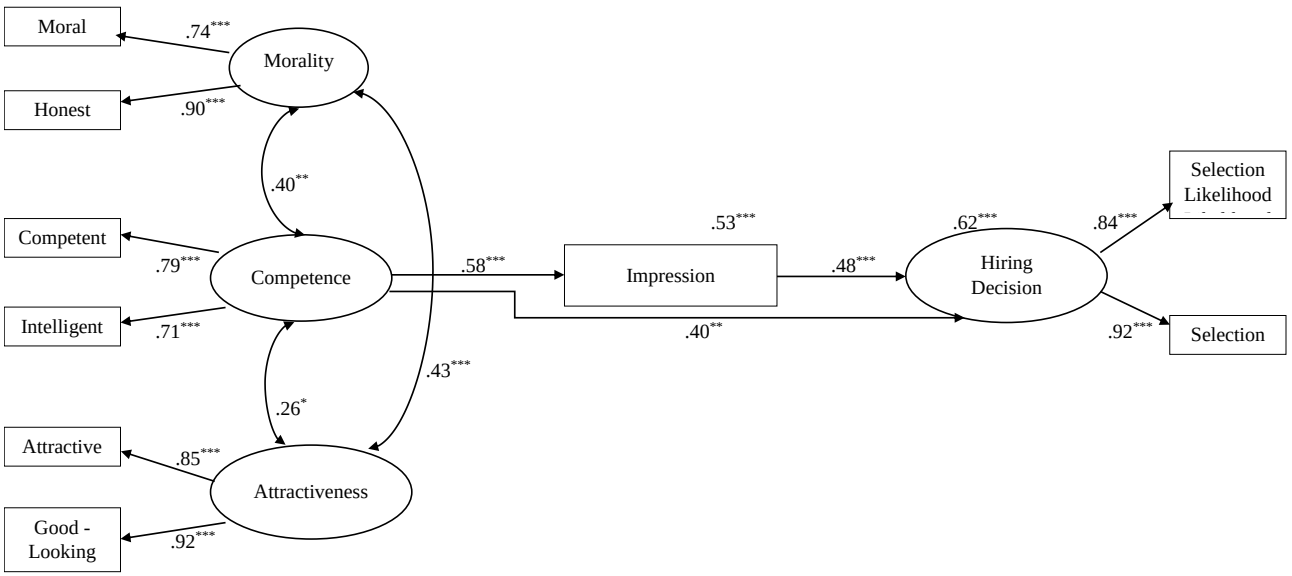
Preliminary analyses were conducted to test the correspondence between the scores of trustworthiness, intelligence, and attractiveness attributed to the selected faces in Oosterhof and Todorov's (2008) study, and the ratings of morality, competence, and attractiveness provided by participants of in the current study. A 2 (trustworthiness: high, low) × 2 (participant gender) × 2 (applicant gender) univariate ANOVA was conducted on morality traits inferred from applicants' faces. Results showed the main effect of trustworthiness, such that applicants with high trustworthy-looking faces were perceived as more moral ($M$ = 4.29, $SD$ = 1.01) than those with low trustworthy-looking faces ($M$ =

3.99, *SD* = 1.03), *F*(1, 205) = 4.75, *p* = .030, $\eta_p^2$ = .02. No other effects were significant, all *p*s > .108. A similar 2 (intelligence: high, low) × 2 × 2 univariate ANOVA on competence traits inferred from applicants' faces revealed a main effect of intelligence, *F*(1, 205) = 6.36, *p* = .012, $\eta_p^2$ = .03. Participants attributed more competence to applicants with high intelligent-looking faces (*M* = 4.74, *SD* = 0.88), than to applicants with low intelligent-looking faces (*M* = 4.35, *SD* = 1.03). The main effect of participant gender showed that male respondents (*M* = 4.34, *SD* = 0.97) attributed lower competence to applicants than female respondents did (*M* = 4.75, *SD* = 0.95), *F*(1, 205) = 7.09, *p* = .008, $\eta_p^2$ = .03. No other effects were significant, all *p*s > .252. The 2 (attractiveness: high, low) × 2 × 2 univariate ANOVA on inferences of applicants' attractiveness showed that applicants with more attractive faces were actually considered as more attractive (*M* = 4.93, *SD* = 1.02) than those with low attractive faces (*M* = 4.60, *SD* = 0.98), *F*(1, 206) = 5.02, *p* = .026, $\eta_p^2$ = .024. No other effects were significant, all *p*s > .053. Finally, to test whether hiring decisions differed as a function of the gender of participants and applicants, it was performed a 2 (participant gender) × 2 (applicant gender) univariate ANOVA on hiring decision. No significant effects were found, all *p*s > .083, indicating that the chances of being hired of male and female applicant were comparable.

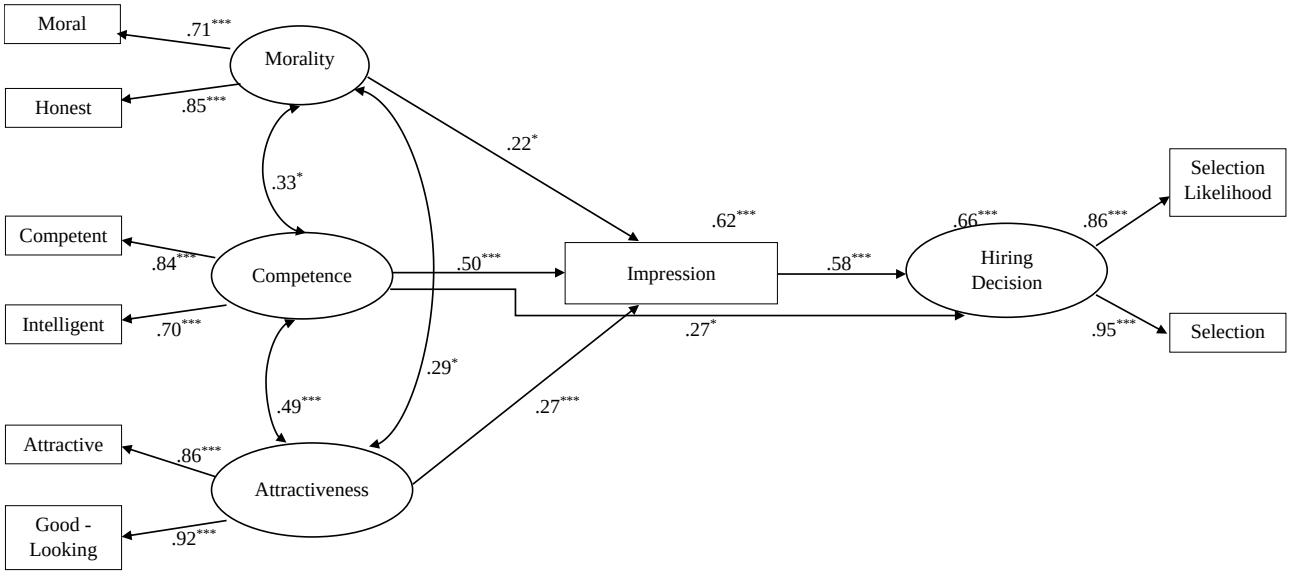
**Structural equation modelling analyses**

To test the hypotheses, it was conducted multi-group Structural Equation Modelling (SEM) analyses in M*plus* 8.1 (Muthén & Muthén, 2017), using the Maximum Likelihood Robust (MLR) estimator. It was tested a model whereby the dimensions (morality, competence, and attractiveness) underlying trait inferences from faces were represented by three latent variables (with two observed indicators each), and predicted hiring decisions, represented by one latent variable (with two observed indicators), both directly and indirectly, through the mediation of overall impression (represented by an observed variable). Correlations between the three dimensions inferred from faces were also

included in the model that was tested in two independent groups, defined on the basis of the gender of applicants. As a preliminary step, measurement invariance was tested to establish whether the measurement model with four latent variables was invariant across the two groups under investigation (i.e., participants who evaluated faces of males' candidates or females' candidates). To this end, the configural (baseline) model was compared with the metric model, in which factor loadings were constrained to be equal across groups.




*Panel a*. Male Applicants




*Panel b*. Female Applicants

**Figure 2**. Structural Equation Model for a) Male Applicants and b) Female applicants. Values reported above Impression and Hiring Decision indicate portions of explained variance.

**Note.** $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

To determine differences between models, at least two out of these three criteria had to be matched: $\Delta\chi SB2$ significant at $p < .05$ (Satorra & Bentler, 2001), $\Delta CFI \geq -.010$, and $\Delta RMSEA \geq .015$ (Chen, 2007). Model comparisons indicated that metric invariance could be clearly established ($\Delta\chi SB2 = 5.207$, $\Delta df = 4$, $p = .267$, $\Delta CFI = -.002$, $\Delta RMSEA = .004$).

Based on this, it could be possible to proceed with analyses aimed at unraveling associations among study variables across the two groups of interest. The model fit was evaluated by means of multiple indices: the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), with values higher than .90 indicative of an acceptable fit and values higher than .95 suggesting an excellent fit; and the Root Mean Square Error of Approximation (RMSEA), with values below .08 indicative of an acceptable fit and values less than .05 representing a very good fit (Byrne, 2012). In addition, the 90% confidence

interval of the RMSEA was inspected: when the upper bound of this confidence interval is ≤ .10, the model fit can be considered acceptable.

The results of the multi-group analyses indicated that the model tested in the two separate groups fitted the data very well, $\chi SB\ 2 = 50.417$, df = 44, $p$ = .235, CFI = .993, TLI = .988, RMSEA = .036 [.000, .076]. Standardized parameter estimates are reported in Figure 2a (for male applicants) and 2b (for female applicants). In line with the hypotheses, meaningful differences were found based on applicants' gender. Specifically, for *male applicants* (Figure 2a) only facial competence was significantly related to hiring decision, both directly and indirectly, by positively affecting overall impression, which in turn was strongly and positively related to hiring decisions, $\beta$ = .28, $p$ < .001, 95% CI = [.15, .42]. There were no direct nor indirect effects of facial morality, $\beta$ = .06, $p$ = .246, 95% CI = [-.04, .15], and attractiveness, $\beta$ = .10, $p$ < .107, 95% CI = [-.02, .21] on hiring decision.

In contrasts, for *female applicants* (Figure 2b) all indirect effects were statistically significant, facial morality, $\beta$ = .13, $p$ = .039, 95% CI = [.01, .25], facial competence, $\beta$ = .29, $p$ < .001, 95% CI = [.15, .43], and facial attractiveness $\beta$ = .16, $p$ = .007, 95% CI = [.04, .27]. This means that all traits inferred from female applicants' faces were significantly related to overall impression, which significantly mediated their effects on hiring decisions. In addition to these indirect effects, a direct effect of competence on hiring decisions was also detected. Notably, percentage of explained variance were high for both overall impression (53% and 62% for male and female applicants, respectively) and hiring decisions (62% and 66% for male and female applicants, respectively).

Overall, these findings showed that facial competence predicted the hiring decision on the applicants, both directly and indirectly through the global impression. However, while competence inferred from faces was the only predictor of impression and, in turn, hiring decision about male applicants, competence, morality, and attractiveness *all*

predicted the hiring decision about female applicants with the mediation of overall impression.

**Discussion**

Study 1 confirmed the perfection bias hypothesis, since all the facial judgment criteria significantly affected impression and in turn hiring decision on female applicants, while only competence affected hiring decision on male applicants, both directly and indirectly through global impression.

The results also supported previous findings on the role of competence in the workplace (e.g., Brambilla et al., 2011), according to which it is the most important judgment dimension. Indeed, facial competence was a significant predictor of hiring decision in both female and male applicants' models and, in the female applicants' model, beta coefficient of facial competence was the highest compared to the beta coefficients of attractiveness and trustworthiness. Moreover, in both models, facial competence had a direct as well as an indirect effect though global impression on hiring decision.

However, along with competence, multiple judgment dimensions affected global impression and, in turn, hiring decision on women, confirming the perfection bias hypothesis. It is interesting to note that there were not significant differences about hiring decisions on men and women; this is in line with evidence showing no gender bias in gender-neutral job types (Koch et al., 2015).

# CHAPTER III

This chapter presents Study 2, in which the perfection bias hypothesis was tested employing a leadership position. Study 2 also extended previous research considering all the four facets of the fundamental content dimensions of social judgments: competence, dominance, morality, and sociability (Abele et a., 2016). According to the perfection bias hypothesis, all the facial traits considered should be more important in evaluating female applicants than male applicants. In this vein, the importance ranking of the four facets in predicting men and women's hireability was directly gauged employing specific analyses. Moreover, a prestudy test was run to investigate the relative importance of the four facets in order to succeed as Area Sales Manager, the job description subsequently employed in Study 2.

## Prestudy 1: Area Sales Manager position

Prestudy 1 explored how important competence, dominance, morality, and sociability are for either a man, a person, or a woman in order to be successful in the leadership position subsequently used in the main experiment. The aim was to verify whether there would be a correspondence between the judgments on the relative importance of facets in order to succeed as Area Sales Manager and the results about predictor ranking of importance that would be measured in Study 2.

In this vein, while it was decided to select a non-STEM or finance managerial position, which would be employed later (i.e., Study 4), in order to gradually approach the leadership context, the Area Sales Manager position was specifically selected to exploratively examine other two aspects emerged in research: the influence of the context (e.g., Todorov et al., 2015; Brambilla et al., 2011) and job characteristics in the applicants' selection (e.g., Rule et al., 2016; Heilman, 2001). According to the 'Lack of Fit Model' and the 'Role Congruity theory', gender bias is due to a perceived fit or congruency between job characteristics and gender stereotypes. More broadly, in person perception research, Wojciszke (2005) highlighted that context and interaction goal affect the type of information that people consider useful to gather. Similarly, Todorov and colleagues

(2015) have stressed the role of social domains in determining facial traits power to sway people's decisions.

Therefore, on one hand, the leadership position should influence participants' judgments in considering competence and dominance as the two more important facets. On the other hand, since in common beliefs a seller should be 'sociable' in order to be effective with clients, not necessarily 'moral', the specific Area Sales Manager position should influence participants' judgments in considering sociability as more important than morality.

In sum, the position was chosen in order to examine the interplay between the context and the specific job description (as well as gender bias), and their influence in determining facial traits power (i.e., in Study 2). The aim, however, was also to try to manipulate sociability versus morality. Indeed, while in social cognition research they are recognized as two different components of communion/warmth (e.g., Leach et al., 2008), face perception research has failed to detect this distinction (i.e., Sutherland et al., 2016). Hence, if the prestudy would show a difference on the relevance of these two facets, in the main experiment (i.e., Study 2), facial sociability should emerge as more important than facial morality, telling that these two facets could be considered as separated even in face perception.

**Method**

University of Exeter ($N$ = 57; 27 female, 30 male) undergraduates volunteered to fulfill the questionnaire. Participants were asked to read the description of an Area sales manager job, adapted from real job descriptions retrieved from Randstad and Adecco British web sites (Figure 3).

Then, they were asked to indicate "To succeed in this role, to what extent would it be important for either a man, a person, or a woman in this job to be: competent, efficient, intelligent (competence; α = .683); determined, dominant, self-confident (dominance[5]; α = .696); honest, moral, sincere (morality; α = .828); and caring, kind, sociable (sociability α = .684). Responses

[5] Dominant trait was excluded due to unexpected negative covariances with the other two traits. Thus, the dominance index was created averaging determined and self-confident traits.

were given on a 7-point Likert scale (1 = *not at all*, 7 = *very much*). After the questionnaire, they filled in the demographic information form.

**Area Sales Manager**

* Base salary: £50,000 per annum

* Up to 20% bonus for top performance

* Benefits: Company Car, Laptop, and Phone

* Wide range of flexible benefits

***The Role***

A large international brand is looking for a person to manage a team of sales representatives. The successful candidate will have strong leadership skills and be responsible for developing an effective strategy to achieve a significant growth in sales within their area. This role is perfect for someone who is sales driven, passionate about customer management and is looking for the next step in their career.

***Responsibilities***

- To lead and supervise a team of 30 sales representatives
- To increase current sales levels
- To deal with competitors
- To resolve contractual and commercial problems

***Key skills***

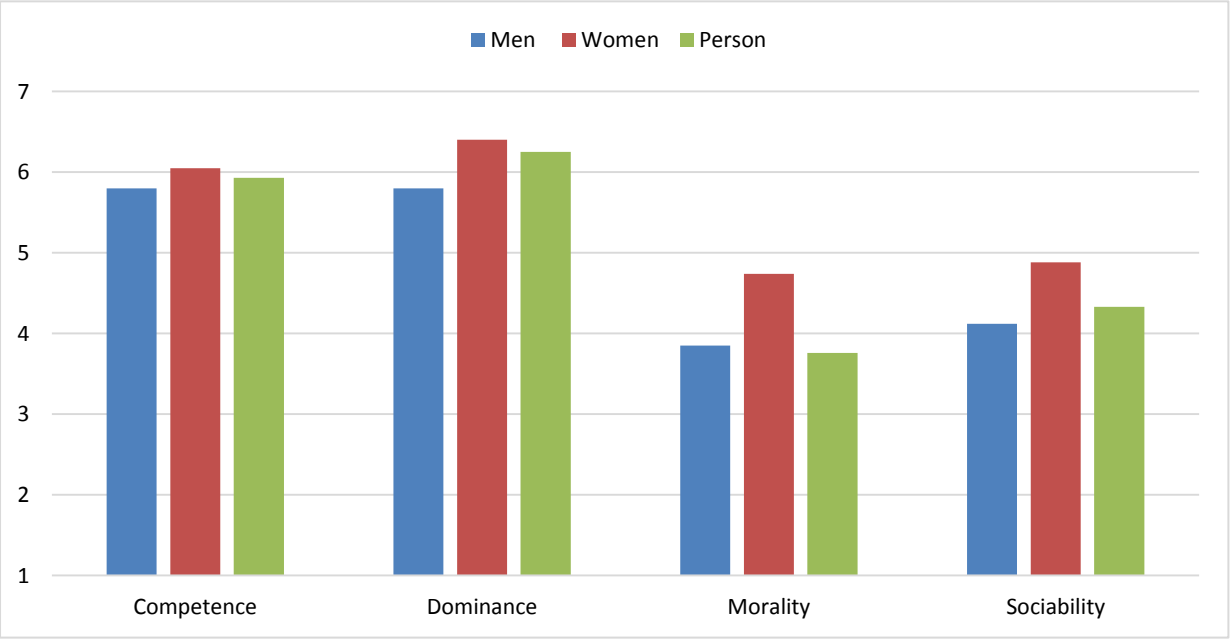
- Excellent leadership skills
- Ability to formulate strategies and concepts
- Ability to deliver results & meet customer expectations
- Solid organisational and communication skills

**Figure 3.** Job description of the Area Sales Manager.

## Results and discussion

A 4 x 3 x 2 mixed-model ANOVA was conducted with facet (competence, dominance, morality, sociability) as a within-subjects factor, and target (person, woman, man) and participant gender as between participants factors (for means, see Graphic 1). The raters' gender was also

inserted for the possibilities of emerging differences in judging women versus men, given literature suggests that some differences could emerge (e.g., Davison & Burke, 2000; Prati et al., 2019).




**Graphic 1.** Means for each facet by target.

Results revealed a main effect of facet[6], $F(2.245, 114.507) = 83.338$, $p < .001$, $\eta_p^2 = .620$. Pairwise comparisons revealed that competence was considered significantly more important to succeeding in the role than morality, $p < .001$, and sociability, $p < .001$. Dominance was also considered significantly more important to succeeding in the role than morality, $p < .001$, and sociability, $p < .001$. There were no significant differences between competence and dominance, or morality and sociability, all $p$s > .166. There was also a main effect of the target, $F(2, 51) = 4.520$, $p = .016$, $\eta_p^2 = .151$. Pairwise comparisons revealed that the mean scores attributed in woman condition were higher than mean scores attributed in man condition, $p = .016$. There were no significant differences between woman and person conditions, or man and person conditions, all $p$s > .168. There was no other significant effect, all $p$s > .053.

[6] Mauchly's test was significant, $\chi^2(5) = 27.313$, $p < .001$, meaning that the assumption of sphericity had not been met. Greenhouse- Geisser estimates of sphericity ($\varepsilon = .748$) were used as degrees of freedom correction.

In sum, as expected, dominance and competence were considered significantly more important (at the top of the scale), than sociability and morality (around the mid-point). Although not significant, sociability was rated as more important than morality. Overall mean ratings in woman condition were higher than overall mean ratings in man condition, suggesting that all the four facets were considered as more important qualities to possess for a woman than a man. Specifically, even if they were not significant, the differences between man and woman target condition appear more evident looking at the mean scores of the morality facet and the sociability facet. Thus, these findings suggested the perfection bias could work even for leadership position. Notably, competence and dominance had the same mean scores (i.e., 5.8) in man target condition, suggesting that, beside competence, also dominance was considered as an important quality to possess for men in order to succeed in this position.

## Study 2

Study 2 investigated for the first time the perfection bias at face perception level for a leadership position. As recent literature has shown, considering diverse facets instead of the two fundamental dimensions could offer a more complex and different picture of the gender bias at work (e.g., Moscatelli et al., 20). In this vein, the perfection bias hypothesis was tested along all the facets of the fundamental dimensions as judgment criteria in evaluating male and female candidates.

It was expected that all the predictors would be relevant for female applicants hireability, while competence were expected to be the most important predictor of male applicants' hireability. Along with competence, dominance was also expected to be as a relevant dimension in judging male applicants, this because of the results of Prestudy 1. As stated above, that result was not totally unforeseen, given the link among agency dimension content, male stereotype content and beliefs about leaders 'qualities (e.g., Ryan et al., 2011; Schein, 2001). Hence, it was expected that competence and dominance would be the two most important predictors of male applicants' hireability. Finally, since the results of Prestudy 1, it was also expected that facial sociability should be more important than facial morality in evaluating applicants.

Compared with Study 1, a different methodology was used. Indeed, facial ratings on the four facets collected during a pre-test (i.e., pre-test photos) and attractiveness ratings, retrieved from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), were used to examine how each impacted the perceived hireability of men versus women. Specifically, the four facets and attractiveness ratings were used as predictors of male and female applicants' hireability ratings, which were collected during the main experiment. The motivation underlying this choice was to collect ratings about 'general' facial first impressions which should be not anchored to a specific context. Moreover, specific metrics were computed in order to specifically investigate the relative importance of the four facets and attractiveness perceived from faces in affecting female and male applicants' hireability. The metrics and the specific reasoning underlying the choice to compute them are broadly described and explained in the 'plan of the analysis' paragraph.

**Method**

***Pre-test: photos***

Students from the University of Bologna ($N$ = 95; 50 men, 43 women; $M_{age}$ = 24.6, $SD_{age}$ = 4.13) rated 183 (90 female, 93 male) photos retrieved from the Chicago Face Database (Ma et al., 2015). All the individuals depicted in the photos are Caucasian, have a neutral expression and wear a grey t-shirt (Figure 4). Participants were asked to rate (1 = *not at all*; 7 = *very much*) to what extent the person portrayed in the photo looked either: competent, efficient, intelligent; determined, dominant, self-confident; honest, moral, sincere; or caring, kind, sociable. Thus, around 20 participants evaluated all the photos on one facet. One hundred photos were selected of 50 female models and 50 male models who appeared between 25 and 40 years old, aiming to make them more credible applicants for a leadership position.

**Figure 4.** Two examples of the pre-tested photos retrieved from the CFD (Ma et al., 2015).

Then, multiple Anovas were run to verify whether ratings of male and female photos in terms of competence (three items averaged: α = 94), dominance (three items averaged: α = 94), morality (three items averaged: α = 98), sociability (three items averaged: α = 94), attractiveness and perceived age were equivalent (the last two scores were retrieved from the CFD, Ma et al., 2015), all *p*s > .358.

***Power analysis***

In Study 1, the effect sizes (Cohen's $f^2$) obtained was 1.94 (female applicants' model) and 1.63 (male applicants' model). Anticipating an effect size 1.50, a power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) revealed that it was needed at least 20 participants to achieve 95% power for multiple linear regression analysis with five predictors and a 5% false-positive rate. However, such a small sample size would not have been much representative of the population. Moreover, due to the methodology employed and, consequently, the way in which the dataset would be created, the analyses would be based on the numbers of photos instead of the numbers of participants. For these reasons, it was decided to recruit a larger number of participants. In order to have a landmark, another power analysis was run entering a medium effect size (i.e., $f^2 = .15$), from which resulted a sample size of 138 participants. Thus, it was chosen to recruit around 140 participants.[7]

[7] Note that the sample sizes in Study 3 and 4 are based on this power analysis.

***Participants***

University of Exeter ($N$ = 144; 120 female, 18 male; $M_{age}$ = 17.7, $SD_{age}$ = 1.34) undergraduates volunteered to participate in the online (Qualtrics) experiment in fulfilment of course requirements. Participants were mostly British (72.5%) and more than a half reported having work experience (68.1%).

***Procedure***

Participants were informed that an international organization was looking for a new Area Sales Manager and the job description was provided. Then, they were randomly presented with the 100 (50 female and 50 male) applicants' photos. Assuming the role of recruiters, their task was to make 'quick' facial first impression judgements of each applicants and to report their hireability judgment for the leadership position with two items (adapted from Rudman & Glick, 1999): "Would you recommend this candidate be interviewed for the job?", "How likely is it that this candidate would be hired for the job?" (1 = *Very Unlikely*, 7 = *Very Likely*; α = .994). Finally, participants filled in the demographic information form.

**Plan of the analysis**

The perfection bias hypothesis was tested in three ways. First, two different regression analyses were conducted entering attractiveness ratings retrieved from the CFD (Ma et al., 2015) and competence, dominance, morality, and sociability ratings retrieved from the pre-test as predictors on male applicant hireability and female applicant hireability. In support of the perfection bias hypothesis, all the facets plus attractiveness would be significantly associated with female applicants hireability, whereas competence and dominance were expected to be the two only significant predictors of male applicants hireability.

Second, other two metrics were calculated in order to directly measure the relative importance of the four facets and attractiveness as predictors of female and male applicants hireability: Dominance analysis (Azen & Budescu, 2003; Budescu, 1993) and Relative Weights

(Johnson, 2000, 2004)[8]. Even if general dominance coefficients and relative importance weights coefficients have different computation methods, they usually produce consistent results (Lorenzo-Seva, Ferrando, & Chico, 2010). Nevertheless, to obtain more reliable results, it was chosen to report them both. These two metrics were also selected in order to deal with possible high correlations among predictors. In fact, since competence and dominance are two components of the agency dimension, as well as morality and sociability are two components of the communion dimension (Abele et al., 2016), high correlations among independent variables were reasonably expected. However, Dominance analysis and Relative Weights tend to de-emphasize redundant predictors in case of shared explained variance (Kraha, Turner, Nimon, Zientek, & Henson, 2012). These metrics were computed using R (R Development Core Team, 2013) 'yhat' package (Nimon, Oswald, & Roberts, 2013) that allows also to perform bootstrap (1000 resamples) and calculate 95% percentile confidence interval around the difference between predictors coefficients within each metrics (e.g., only among general dominance coefficients, only among relative importance weights coefficients). In sum, while the two metrics evaluate the relevance of the five predictors, the confidence intervals are basically a significance test used to establish support for the perfection bias hypothesis. Specifically, for male applicants, relative importance weights coefficients of competence and dominance were expected to be higher than relative importance weights coefficients of attractiveness, morality, sociability. And the difference between relative importance weights coefficient of competence and relative importance weights coefficients of attractiveness, morality, and sociability (i.e., a pair at a time) should be significant, as well as the difference between coefficient of dominance and coefficients of attractiveness, morality, and sociability (i.e., a pair at a time). The same pattern should emerge comparing general dominance coefficients of competence and dominance with general dominance coefficients of the other facets plus attractiveness. These results would mean that competence and dominance were the two most

[8] These two analyses inform about the contribution of each predictor to the r-squared effect (i.e., the sum of coefficients equals the $R^2$), so they directly gauge the importance of the predictors in the model and provide a sort of predictors ranking derived from what variables contribute the most to $R^2$.

important predictors of male applicants' hireability. For female applicants, all the differences between each pair of (general dominance, relative importance weights) coefficients for the five predictors were expected to be not significant, meaning that all the facial inferences should be equally important for female applicants' hireability.

With 'yhat' package, structure coefficients were also computed. Structure coefficients are bivariate correlations between an observed predictor variable and the predicted estimate of the outcome variable (Courville & Thompson, 2001; Kraha et al., 2012). Structure coefficients represent a measure of the relationship between the predictor variable and the predicted variable because they have universal statistical boundaries (–1 to +1), which indicate the direction of the relationship (positive or negative). They are specifically recommended in case of correlated predictors in multiple regression analysis. For example, a beta can be negative and non-significant, even if structure coefficient indicates that the direction of the relationship between the independent and the dependent variable is positive and the independent variable is the best one among predictors (for an example, see Courville & Thompson, 2001, p. 241). Moreover, beta weights evaluate how much the criterion variable increases when the predictor variable is increased by a standard deviation, holding constant other variables in the model. Thus, whereas structure coefficients can be interpreted as measuring relationships, betas cannot be. Structure coefficients were used to directly compare female and male applicants on the relative importance of each of the five facial traits for their hirebility. In particular, five different correlations comparisons (e.g., Field, 2009) were run to confront each single predictor (e.g., only attractiveness, only morality, etc.) in female applicants' model with the corresponding predictor in male applicants' model (e.g., female facial attractiveness Vs. male facial attractiveness). The tests should show that, compared to male applicants' model, the relationship between the five facial inferences and hireability (i.e., structure coefficients) was significantly stronger (i.e., higher) in female applicants' model. This result would mean that attractiveness and the four facets were more important for female applicants hireability than for male applicants hireability.

## Results

Table 2 reports means, standards deviations and Pearson correlation coefficients among all study variables differentiated by applicant gender[9].

As preliminary analysis, a one-way Anova was run with applicant gender as fixed factor on hireability ratings. There was not a significant difference between female applicants' hireability ($M$ = 3.84, $SD$ = .65) and male applicants' hireability ($M$ = 3.74, $SD$ = .67), $p$ = .473, meaning that men and women were judged as equally hireable.

**Table 2.** Means, standard deviations, and Pearson correlations among all study variables, by applicant gender.

| **Female applicants** | **M** | **SD** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|
| **1. Attractiveness** | 3.11 | .77 | .677** | .626** | .164 | .090 | .833** |
| **2. Competence** | 3.95 | .55 | - | .468** | .581** | .502** | .819** |
| **3. Dominance** | 3.68 | .69 | | - | -.231† | -.181† | .599** |
| **4. Morality** | 3.72 | .45 | | | - | .656** | .390** |
| **5. Sociability** | 3.62 | .57 | | | | - | .344* |
| **6. Hireability** | 3.84 | .65 | | | | | - |
| **Male applicants** | **M** | **SD** | **2** | **3** | **4** | **5** | **6** |
| **1. Attractiveness** | 2.97 | .62 | .614** | .362** | .477** | .219† | .777** |
| **2. Competence** | 3.92 | .48 | - | .470** | .564** | .522** | .784** |
| **3. Dominance** | 3.55 | .68 | | - | -.153† | .036† | .389** |
| **4. Morality** | 3.72 | .41 | | | - | .649** | .637** |
| **5. Sociability** | 3.55 | .59 | | | | - | .579** |
| **6. Hireability** | 3.74 | .67 | | | | | - |

**Note.** † $p$ > .11, * $p$ < .05, ** $p$ < .01.

Table 3 reports the results of the two regression analyses on female applicants hireability and male applicants hireability, in which facial attractiveness, competence, dominance, morality, and sociability were entered as predictors. Both models were highly significant explaining 82.4%

[9] All the analyses were re-run excluding foreigners, but they did not highlight a meaningfully different pattern.

and 81% of the variance of female applicants' hireability and male applicants' hireability respectively. Only attractiveness and dominance were significant for female applicant hireability[10], while in male applicants' model, only attractiveness and sociability were significantly associated with hireability.

**Table 3.** Standardize regression coefficients of attractiveness, dominance, morality, and sociability on **hireability**, for female and male applicants**.**

| | Female Applicants | | | Male Applicants | | |
|---|---|---|---|---|---|---|
| | **β (SE)** | **t** | **95% CI** | **β (SE)** | **t** | **95% CI** |
| **Attractiveness** | .502*** (.080) | 5.329 | .266, .588 | .478*** (.095) | 5.428 | .323, .704 |
| **Competence** | .191† (.161) | 1.406 | -.098, .550 | .219† (.150) | 2.014 | .000, .605 |
| **Dominance** | .259* (.098) | 2.518 | .049, .443 | .122† (.089) | 1.361 | -.058, .301 |
| **Morality** | .162† (.154) | 1.526 | -.075, .546 | .128† (.186) | 1.135 | -.164, .587 |
| **Sociability** | .144† (.098) | 1.673 | -.033, .360 | .272** (.101) | 3.065 | .106, .512 |
| | $R^2$= .842, $adjR^2$ = .824 $F(5, 44) = 46.956$, $p < .001$ | | | $R^2$= .830, $adjR^2$= .810, $F(5, 44) = 42.908$, $p < .001$ | | |

**Note.** † $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

In order to evaluate the relevance of the five predictors in affecting male and female applicants' hireability, general dominance weights (GDW) and relative importance weights (RIW) were computed (for results, see Table 4). Across these two metrics, the predictors ranking of importance of female applicants' hireability is consistent, that is in descending order of importance: facial attractiveness, facial competence, facial dominance, facial morality, and facial sociability. The bootstrap confidence intervals around the differences between coefficients showed that general

[10] The Variance Inflection Factor (VIF) of the competence variable was 5.146 suggesting collinearity issue. Since Tolerance value was higher than .01 (.194) and the Collinearity index was less than 30 (16.833), collinearity was not problematic (Barbaranelli & D'Olimpio, 2007) and competence was not excluded from the regression analysis.

dominance weights of facial attractiveness was significantly higher than general dominance weights of facial dominance, morality and sociability; the difference between general dominance weights coefficients of facial attractiveness and facial competence was not significant. General dominance weights coefficient of facial competence was significantly higher than general dominance weights coefficients of facial dominance, morality and sociability. General dominance weights coefficient of facial dominance was significantly higher than general dominance weights coefficients of facial morality and sociability. The difference between general dominance weights coefficients of facial morality and facial sociability was not significant. The same pattern emerged with the bootstrap confidence intervals around the differences among relative importance weights coefficients, except that the difference between relative importance weights coefficients of facial competence and facial dominance was not significant.

**Table 4.** General dominance weights (GDW), relative important weights (RIW), and structure coefficients ($r_s$) for female and male applicants.

| | Female Applicants | | | Male Applicants | | |
|---|---|---|---|---|---|---|
| | **GDW** | **RIW** | $r_s$ | **GDW** | **RIW** | $r_s$ |
| **Attractiveness** | .317 a | .325 a | .908 | .277 a | .287 a | .853 |
| **Competence** | .255 a | .211 ab | .892 | .218 a | .193 b | .861 |
| **Dominance** | .159 b | .175 b | .653 | .067 b | .076 c | .427 |
| **Morality** | .064 c | .074 c | .425 | .142 c | .139 b | .699 |
| **Sociability** | .048 c | .057 c | .375 | .125 bc | .136 bc | .636 |

**Note.** Different letters in column mean significant difference among coefficients.

For male applicants' hireability, results of the general dominance weights and relative importance weights were again consistent in delineating the predictors ranking, that is in descending order: facial attractiveness, facial competence, facial morality, facial sociability, and facial

dominance. The bootstrap confidence intervals around the differences between coefficients showed that general dominance weights coefficient of facial attractiveness was significantly higher than general dominance weights coefficient of facial dominance, morality and sociability; the difference between general dominance weights coefficients of facial attractiveness and facial competence was not significant. General dominance weights coefficient of facial competence was significantly higher than general dominance weights coefficients of facial dominance, morality and sociability. General dominance weights coefficient of facial dominance was significantly lower than general dominance weights coefficients of facial morality, while the difference between general dominance weights coefficients of facial dominance and facial sociability was not significant. The difference between general dominance weights coefficients of facial morality and facial sociability was not significant. The same pattern emerged with the bootstrap confidence intervals around the differences between relative importance weights coefficients. The only exception is that the difference between relative importance weights coefficient of facial attractiveness was significantly higher than relative importance weights coefficient of facial competence.

Finally, female and male applicants' models were directly compared on each facial traits structure coefficient. Correlations comparisons showed that morality structure coefficient for male applicants was significantly higher than morality structure coefficient for female applicants, $Z = -2$, $p = .045$, $n = 50$, meaning that morality perceived from male applicants' faces was more important than morality perceived from female applicants' faces in predicting their hierability. All the other comparisons were not significant, all *p*s > .084.

**Discussion**

According to the results obtained from the regression analyses, Dominance analyses, Relative Important Weights analyses, and correlations comparisons of structure coefficients, the perfection bias hypothesis was not supported. Indeed, in regression analyses, only dominance and attractiveness instead of all the predictors were significantly associated to female applicants' hireability, while competence and dominance were not significant predictors of male applicants'

hireability. Dominance analyses and Relative Important Weights analyses showed that attractiveness, competence, and dominance were more important than morality and sociability for female applicants' hireability, instead of showing that all the predictors were equally important. Moreover, competence emerged as an important predictor of male applicants' hireability but it was not the best one (i.e., attractiveness was), and dominance was the least important among predictors. Finally, according to correlations comparisons, structure coefficients of female applicants were not higher than structure coefficients of male applicants. Conversely, results showed only one significant difference in the opposite direction: facial morality structure coefficients for female applicants was significantly lower than facial morality structure coefficients for male applicants. Concerning the relative importance of sociability compared to morality, in male applicants' model, sociability beta was significant while morality beta was not. However, across all the other metrics, coefficients of morality were higher than coefficients of sociability, meaning that morality was more important in affecting participants' judgments.

Although they are different to the expectations, results showed an interesting pattern. First, facial attractiveness was the most important judgment criterion in evaluating applicants, indeed, attractiveness betas were significant in both female and male applicants' regression models. Moreover, general dominance and relative weights coefficients of attractiveness for both female and male applicants were significantly higher than general dominance and relative weights coefficients of dominance, morality and sociability. Above significance, across all the metrics considered, coefficients of facial attractiveness were always higher than coefficients of the other facial traits. Despite the results obtained in Study 1, the fact that facial attractiveness emerged as the most important among predictors was not particularly surprising, since literature had broadly highlighted the relevance of attractiveness in affecting job related outcomes for both men and women (Hosoda et al., 2003). Second, facial competence was perceived as the second most important predictor of applicants' hireability. Indeed, general dominance and relative weights coefficients of competence for female applicants were significantly higher than general dominance and relative weights

coefficients of dominance, morality and sociability. Above significance, across all the metrics considered, coefficients of facial competence were always higher than coefficients of dominance, morality and sociability in both female and male applicants' models. Given classical literature on gender stereotypes and the perfection bias hypothesis, this finding was unexpected for female applicants. However, this finding can be comprehensible referring to the literature that recognized competence as the most important facet in a work context (e.g., Brambilla et al., 2011).

Difficult to explain was instead the findings concerning the relevance of facial dominance, facial morality and facial sociability. Indeed, besides attractiveness, dominance was the only other significant predictor in female applicants' regression model and, according to the bootstrap confidence intervals, it was more important than morality and sociability in affecting female applicants' hireability. Thus, facial dominance was considered as more important than facial morality and facial sociability in evaluating women. Regarding male applicants, beside attractiveness, sociability was the only other significant predictor in the regression model and, according to the bootstrap confidence intervals, facial dominance was significantly less important than facial morality. Most strikingly, according to correlations comparisons, facial morality was perceived as more important in evaluating male applicants than female applicants. Thus, facial morality and facial sociability was considered as more important than facial dominance in evaluating men. Since morality and sociability are facets of the communion/warmth dimension (Abele et al., 2016) which represents the female content stereotypes (e.g., Abele..), while dominance is a facet of the agency/competence dimension (Abele et al., 2016) which represents the male content stereotypes (e.g., Abele..), these findings suggested that participants had relied on counter-stereotypic judgment criteria in evaluating men and women.

In conclusion, it is interesting to note that women were not judged as less hireable than men, as in Study 1. Although it had not been a surprise in the previous experiment in which the job description was gender neutral, such a result was unexpected given the leadership position employed in this study.

# CHAPTER IV

## Study 3

Given the original results of the previous experiment, it was decided to replicate it. Hence, in Study 3, the perfection bias hypothesis was tested with the same material (i.e., photos, job description), procedure and facial traits scores (as predictors of applicants hireability) employed in Study 2. However, since in Study 2 participants had been mostly female students, in Study 3 it was decided to recruit a more gender balanced sample of working adults.

### Method

#### *Participants*

One hundred and twenty-nine (64 female, 65 males; $M_{age}$ = 37.14, $SD_{age}$ = 9.82) working adults residing in the United Kingdom were recruited through Prolific (www.prolific.co). The majority (i.e., 90.7%) reported to be British citizens.

#### *Procedure and material*

As for the previous experiment, participants were provided with an Area Sales Manager job description and 100 facial photos (50 men and 50 women). For each applicant, participants answered two questions: "Would you recommend this candidate be interviewed for the job?", "How likely is it that this candidate would be hired for the job?" (1 = Very Unlikely, 7 = Very Likely; α = .998). At the end, participants filled in the demographic information form.

### Results

Table 5 reports means, standards deviations and Pearson correlation coefficients among all study variables differentiated by applicant gender[11].

As a preliminary analysis, a 2 (participant gender: men, women) x 2 (applicant gender) ANOVA was run on hireability judgments, with the last factor within participants. Results showed a main effect of applicant gender, $F(1, 127) = 17.328$, $p < .001$, $\eta_p^2 = .120$, so that female applicants

[11] All the analyses were re-run by foreigners and participants' gender, but they did not highlight a meaningfully different pattern.

(*M* = 3.86, *SD* = .75) were considered as more hireable than male applicants (*M* = 3.78, *SD* = .71). No other effect was significant, all *p*s > .116.

**Table 5. Means, standard deviations, and correlations among all study variables, by applicant gender.**

| Female applicants | M | SD | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **1. Attractiveness** | 3.11 | .77 | .677** | .626** | .164† | .090† | .873** |
| **2. Competence** | 3.95 | .55 | - | .468** | .581** | .502** | .803** |
| **3. Dominance** | 3.68 | .69 | | - | -.231† | -.181† | .594** |
| **4. Morality** | 3.72 | .45 | | | - | .656** | .337* |
| **5. Sociability** | 3.62 | .57 | | | | - | .272† |
| **6. Hireability** | 3.86 | .73 | | | | | - |
| **Male applicants** | **M** | **SD** | **2** | **3** | **4** | **5** | **6** |
| **1. Attractiveness** | 2.97 | .62 | .614** | .362** | .477** | .219† | .715** |
| **2. Competence** | 3.92 | .48 | - | .470** | .564** | .522** | .781** |
| **3. Dominance** | 3.55 | .68 | | - | -.153† | .036† | .418** |
| **4. Morality** | 3.72 | .41 | | | - | .649** | .601** |
| **5. Sociability** | 3.55 | .59 | | | | - | .572** |
| **6. Hireability** | 3.71 | .72 | | | | | - |

**Note.** † $p > .06$, * $p < .05$, ** $p < .01$.

Then, two regression analyses were run entering facial attractiveness, competence, dominance, morality, and sociability as predictors of female applicants' hireability and male applicants' hireability (for results, see Table 6). Both female and male applicants' models were significant and explained 83.3% and 74.5% of the variance respectively. Results showed that attractiveness and competence were significantly associated to female applicants' hreability, while attractiveness, competence and sociability were significant predictors of male applicants' hireability.

**Table 6.** Standardize regression coefficients of attractiveness, dominance, morality, and sociability on **hireability**, for female and male applicants**.**

| | Female Applicants | | | Male Applicants | | |
|---|---|---|---|---|---|---|
| | **β (SE)** | **t** | **95% CI** | **β (SE)** | **t** | **95% CI** |
| **Attractiveness** | .594*** (.088) | 6.482 | .391, .744 | .371*** (.118) | 3.637 | .191, .665 |
| **Competence** | .276* (.176) | 2.090 | .013, .721 | .259* (.187) | 2.061 | .008, .761 |
| **Dominance** | .119† (.107) | 1.190 | -.088, .343 | .173† (.111) | 1.655 | -.040, .406 |
| **Morality** | .070† (.168) | .681 | -.225, .454 | .135† (.232) | 1.036 | -.227, .707 |
| **Sociability** | .055† (.107) | .661 | -.145, .286 | .261* (.125) | 2.532 | .065, .570 |
| | $R^2$= .850, $adjR^2$ = .833 $F(5, 44) = 50.015$, $p < .001$ | | | $R^2$= .771, $adjR^2$ = .745, $F(5, 44) = 29.656$, $p < .001$ | | |

**Note.** † $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

In order to evaluate the relative importance of the five predictors in the two models, Dominance analyses and Relative weight analyses were conducted (Table 7). For female applicants, general dominance weight and relative importance weight coefficients delineated a consistent pattern of importance, that is in descending order: facial attractiveness, facial competence, facial dominance, facial morality, and facial sociability. The bootstrap confidence intervals around the differences showed that general dominance weights coefficient of facial attractiveness was significantly higher than general dominance weights coefficients of facial competence, dominance, morality and sociability. General dominance weights coefficient of facial competence was significantly higher than general dominance weights coefficients of facial dominance, morality and sociability. General dominance weights coefficient of facial dominance was significantly higher than general dominance weights coefficient of facial sociability, but the difference between general dominance weights coefficients of facial dominance and facial morality was not significant. The difference between general dominance weights coefficients of facial morality and facial sociability

was not significant. The same pattern emerged with the bootstrap confidence intervals around the differences between relative importance weights coefficients, except that the difference between relative importance weights coefficients of facial competence and facial dominance was not significant.

**Table 7.** General dominance weights (GDW), relative important weights (RIW), and structure coefficients ($r_s$) for female and male applicants.

| | Female Applicants | | | Male Applicants | | |
|---|---|---|---|---|---|---|
| | **GDW** | **RIW** | $r_s$ | **GDW** | **RIW** | $r_s$ |
| **Attractiveness** | .375 a | .385 a | .947 | .215 ab | .222 a | .814 |
| **Competence** | .256 b | .222 b | .871 | .225 a | .198 a | .889 |
| **Dominance** | .143 c | .153 bc | .644 | .081 b | .091 a | .476 |
| **Morality** | .047 cd | .055 cd | .365 | .126 ab | .126 a | .684 |
| **Sociability** | .029 d | .035 d | .295 | .124 ab | .134 a | .651 |

**Note.** Different letters in column mean significant difference among coefficients.

For male applicants hireability, results of the general dominance weights showed the following predictors ranking: facial competence, facial attractiveness, facial morality, facial sociability, and facial dominance. According to the bootstrap confidence intervals around the differences, general dominance weights coefficient of facial competence was significantly higher than general dominance weights coefficient of facial dominance. No other difference was significant. The relative importance weights outlined a slightly different predictors ranking, that is in descending order: facial attractiveness, facial competence, facial sociability, facial morality, and facial dominance. The bootstrap confidence intervals did not show any significant difference.

Female and male applicants' models were then directly compared on each facial trait structure coefficient. Correlations comparisons showed that attractiveness structure coefficient of

female applicants was significantly higher than attractiveness structure coefficient of male applicants, $Z$ = 3.21, $p$ = .001, $n$ = 50. Morality structure coefficient of female applicants was significantly lower than morality structure coefficient of male applicants, $Z$ = -2.2, $p$ = .028, $n$ = 50. Sociability structure coefficient of female applicants was significantly lower than sociability structure coefficient of male applicants, $Z$ = -2.29, $p$ = .022, $n$ = 50. All the other comparisons were not significant, all $p$s > .231.

**Discussion**

The results did not support the perfection bias hypothesis. Indeed, in regression analyses, only facial attractiveness and facial competence instead of all the predictors were significantly associated to female applicants' hireability, while facial attractiveness, facial competence, and facial sociability were significant predictors of male applicants' hireability. Moreover, according to Dominance analyses and Relative Important Weights analyses, the judgment criteria were not equally important in affecting female applicants' hireability, given facial attractiveness and facial competence were perceived as more important than facial dominance. On the other hand, for male applicants' almost no differences emerged in the relevance of facial traits. Moreover, correlations comparisons showed that only facial attractiveness was higher for female applicants than male applicants, instead of all the facial traits structure coefficients. Finally, beta and relative weights coefficients of sociability were higher than beta and relative weights coefficients of morality. However, across all the other metrics in both models, coefficients of sociability were lower than coefficients of morality, meaning that morality was more important than sociability in affecting participants' judgments.

In fact, the general pattern was quite consistent with findings of Study 2. Indeed, facial attractiveness and facial competence emerged as the two most important predictors for both female and male applicants' hireability. This because attractiveness and competence betas were significant in both female and male applicants' models and, besides the significance test of the bootstrap

confidence intervals, general dominance and relative weights coefficients, as well as structure coefficient, of attractiveness and competence were higher than coefficients of the other facial traits.

Regarding the other facial traits, dominance was more important than morality and sociability in affecting female applicants' hireability. Note that only the difference between dominance and sociability coefficients were significant according to the bootstrap confidence intervals around general dominance and relative weight coefficients. However, looking at betas, general dominance, relative weights and structure coefficients of female applicants, facial dominance was consistently more important than facial morality. Conversely, facial morality and facial sociability were more important than facial dominance in affecting male applicants' hireability. In fact, sociability beta coefficient was significant and, even if bootstrap confidence intervals did not show significant differences around coefficients, general dominance and relative weights coefficients of morality and sociability were consistently higher than dominance. Surprisingly, according to correlation comparisons, both facial morality and facial sociability were perceived as more important in evaluating men than women. Hence, Study 3 confirmed that, besides attractiveness and competence, participants put more attention on counter stereotypic facets in evaluating men and women. In sum, Study 2 and Study 3 consistently showed that men and women were evaluated on attractiveness, competence, and their counter-stereotypic traits.

To conclude, women were rated as more hireable than men and, since the leadership context, that represented a curious result. Maybe it could be explained under the framework of the benevolent sexism or the 'women are wonderful effect' (e.g., Glick et al., 2000) as a general positive attitude toward women held by participants.

# CHAPTER V

In this chapter it is reported Study 4, which represented the final test of the perfection bias hypothesis for leadership positions, employing a different job description: Finance Manager. As for the previous one, a prestudy test was run in order to evaluate the importance of the four facets and attractiveness for this specific position.

## Prestudy 2: Finance Manager position

The Prestudy 2 tested the relative importance of competence, dominance, morality, and sociability plus attractiveness for being a Finance Manager. The choice of this position has twofold motivations: firstly, compared to the Area Sales Manager, this is a more salient position in gender discrimination literature. Indeed, the financial sector is a particularly male dominated field and usually mentioned along with the STEM work sectors, since the fraction of women on the financial activities boards is only 1.6 percent (Adams & Kirchmaier, 2016). Thus, the Finance Manager job description was selected in order to have not only a general leadership position, but a leadership position for which women are more strongly discriminated. Secondly, since what happened in the last decade, the financial world is under discussion for its (im)morality (e.g., see the online prof. Yeo Min Yoon's letter published by the Financial Time). Hence, while an Area Sales Manager is probably stereotyped as sociable, it is probable that the morality facet would be more important than the sociability facet in evaluating applicants for a Finance manager position. Hence, as for the previous job description, this position was chosen also to exploratively investigate the relevance of the context and job characteristics in affecting traits relevance, to verify a possible correspondence with main study results (i.e., Study 4), and to manipulate morality versus sociability relevance.

### Method

One hundred and fifteen (53 female, 54 male, 8 missing values; $M_{age}$ = 37.35, $SD_{age}$ = 9.59) participants, recruited via Prolific (www.prolific.co), were asked to read the description of a Finance Manager position that had been adapted from actual job advertisements retrieved from Randstad and Adecco UK web sites (Figure 5).

**Finance Manager**

* Base salary: £50,000 per annum

* Up to 20% bonus for top performance

* Benefits: Company Car, Laptop, and Phone

* Wide range of flexible benefits

***The Role***

A large international company is looking to appoint a Finance Manager to lead a team of management accountants. The successful candidate will have to display strong leadership. They will be responsible for partnering operational and product leaders across multiple businesses to provide financial support in all their business decisions and future financial outlook. This role offers a breadth of responsibilities in a multi-disciplinary and multi-national finance operation and is perfect for someone who is looking for the next step in their career.

***Responsibilities***

- To lead and supervise a team of 30 management accountants
- To identify and mitigate financial risks in the business
- To have oversight over the preparation of financial analysis for management and other stakeholders
- To have oversight over the preparation and review of monthly balance sheet account reconciliations

***Key skills***

- Excellent leadership skills
- Ability to formulate strategies and manage key stakeholders
- Ability to deliver results and meet shareholder expectations
- Solid organisational skills and confident communicator

**Figure 5.** Job description of the Finance Manager.

Then, participants read the following instructions: "Based on the Finance Manager job description, in order to be selected, how important would it be for a man/person/woman who applied for this job to be: attractive; competent, efficient, intelligent (competence; $\alpha$ = .751); determined, dominant, self-confident (dominant; $\alpha$ = .494); honest, moral, sincere (morality; $\alpha$ =

.776); caring, kind, sociable" (sociability; α = .776) (1 = *not at all*, 7 = *very much*). After the questionnaire, they filled in the demographic information form.

**Results and discussion**

A 4 x 3 x 2 mixed-model Anova was run with facet (competence, dominance, morality, sociability, and attractiveness) as a within-participants factor, and target (men, person, women) and participant gender (male, female) as between participants factors (for means, see Table 8).

**Table 8.** Means (standard deviations) for each facets and attractiveness by target and participant gender.

| | | **Competence** | **Dominance** | **Morality** | **Sociability** | **Attractiveness** |
|---|---|---|---|---|---|---|
| Target | Participant Gender | **M (SD)** | **M (SD)** | **M (SD)** | **M (SD)** | **M (SD)** |
| **Person** | **Female** | 6.72 (.32) | 6.26 (.45) | 5.14 (1.10) | 3.79 (1.05) | 2.26 (1.33) |
| | **Male** | 6.52 (.53) | 6.04 (.82) | 4.93 (1.28) | 4.24 (1.21) | 2.33 (1.45) |
| | **Total** | 6.62 (.44) | 6.15 (.65) | 5.04 (1.18) | 4.01 (1.14) | 2.30 (1.37) |
| **Woman** | **Female** | 6.52 (.65) | 6.09 (.56) | 5.35 (1.29) | 4.41 (1.16) | 2.50 (2.18) |
| | **Male** | 6.49 (.47) | 5.76 (.73) | 5.26 (1.04) | 4.39 (1.34) | 1.88 (1.73) |
| | **Total** | 6.50 (.56) | 5.93 (.66) | 5.31 (1.15) | 4.40 (1.23) | 2.20 (1.97) |
| **Man** | **Female** | 6.90 (.26) | 6.54 (.50) | 5.73 (1.43) | 4.40 (1.15) | 2.38 (1.82) |
| | **Male** | 6.32 (.56) | 5.68 (.64) | 5.03 (.1.19) | 4.21 (1.40) | 2.16 (1.46) |
| | **Total** | 6.58 (.53) | 6.08 (.72) | 5.35 (1.33) | 4.29 (1.28) | 2.26 (1.61) |
| **Total** | **Female** | 6.70 (.46) | 6.29 (.53) | 5.39 (1.26) | 4.18 (1.14) | 2.38 (1.77) |
| | **Male** | 6.44 (.52) | 5.83 (.73) | 5.07 (1.16) | 4.28 (1.30) | 2.13 (1.53) |
| | **Total** | 6.57 (.51) | 6.06 (.68) | 5.23 (1.22) | 4.23 (1.22) | 2.25 (1.65) |

Results showed a main effect of facet[12], $F(2.23, 224.75) = 264.35$, $p < .001$, $\eta_p^2 = .724$. All pairwise comparisons among facets were significant, all *p*s < .001. Based on the means, facets were

[12] Mauchly's test was significant, $\chi^2(9) = 173.81$, $p < .001$, meaning that the assumption of sphericity had not been met. Greenhouse- Geisser estimates of sphericity (ε =.56) were used as degrees of freedom correction.

significantly ranked in the following order of importance: competence, dominance, morality, sociability, and attractiveness. There was also a main effect of participant gender, $F(1, 101) = 4.83$, $p = .030$, $\eta_p^2 = .046$, with female participant's reporting higher ratings than male participants. No other effect was significant, all *p*s > .220. In sum, since no differences emerged between man and woman conditions, the perfection bias hypothesis was not supported. However, as expected, competence and dominance were considered more important than the other facets, and morality was rated as more important than sociability. Curiously, despite attractiveness had been resulted as one of the most important predictor of applicants' hireability in previous studies, especially for women, it received the lowest ratings in this prestudy. This would mean that people are not aware of its power in affecting their judgments.

## Study 4

Study 4 finally examined the relative importance of all the four facets of the fundamental content dimensions (Abele et al., 2016) in judging female and male applicants for a Finance Manager position. Study 2 and Study 3 had consistently showed that attractiveness, competence were the two most important dimensions in predicting hiring judgments on applicants. Moreover, dominance was more important than morality and sociability in affecting female applicants hireability, conversely morality and sociability were more important in affecting male applicants hireability. Hence, the perfection bias seemed to not work at face perception level for general leadership positions. However, it perhaps could emerge for a leadership position in which women are particularly discriminated against (Adams & Kirchmaier, 2016). In this case, competence and dominance should be more important predictors than the other facial traits of male applicants' hireability, while all the facial traits should be important in affecting female applicants' hireability. Since the relevance of the morality issue in the financial world, it was also expected that facial morality would be more relevant than facial sociability in affecting applicants' hireability to the Finance Manager position.

## Method

### *Participants*

One hundred and fifty-one (77 female, 74 male; $M_{age}$ = 37.67, $SD_{age}$ = 8.82) working adults residing in the United Kingdom were recruited via Prolific (www.prolific.co). The majority (i.e., 90.1%) reported to be British citizens.

### *Procedure and material*

Participants were provided with the Finance Manager job description and the same100 facial photos (50 men and 50 women) used in the previous two experiments. For each applicant, participants answered the same two questions employed in the last two studies, subsequently averaged to create the hireability index ($\alpha$ = .998). At the end, participants filled in the demographic information form.

## Results

Table 9 reported means, standard deviations, and correlations among all study variables, by applicant gender[13].

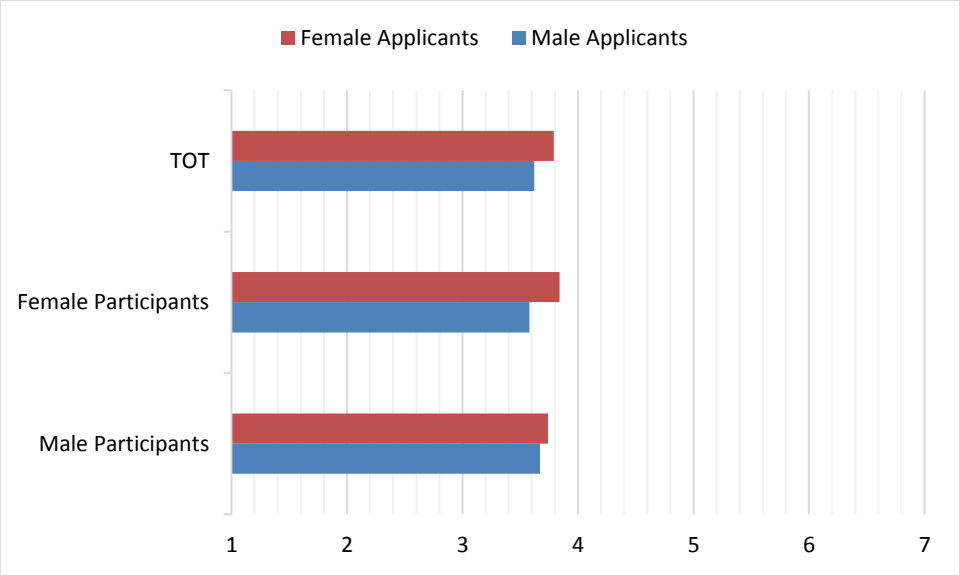
A 2 x 2 mixed-model ANOVA was run with participant gender as between-subjects variable and applicant gender as a within-subjects variable on hireability (for means, see Graphic 2). Results revealed a significant effect of applicant gender, $F(1, 149) = 18.880$, $p < .001$, $\eta_p2 = .112$, so that female applicants were considered as more hireable than male applicants. There was also a significant interaction effect, $F(1, 149) = 5.818$, $p = .017$, $\eta_p2 = .038$. Pairwise comparisons showed that female participants attributed higher ratings to female applicants than male applicants, $p < .001$. No other comparisons were significant, all *p*s > .178. The effect of participant gender was not significant, $p = .944$.

[13] All the analyses were re-run by foreigners and participants' gender, but they did not highlight a meaningfully different pattern.

**Table 9. Means, standard deviations, and correlations among all study variables, by applicant gender.**

| Female applicants | M | M | SD | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| 1. Attractiveness | 3.11 | .77 | .677** | .626** | .164† | .090† | .877** |
| 2. Competence | 3.95 | .55 | - | .468** | .581** | .502** | .779** |
| 3. Dominance | 3.68 | .69 | | - | -.231† | -.181† | .586** |
| 4. Morality | 3.72 | .45 | | | - | .656** | .305* |
| 5. Sociability | 3.62 | .57 | | | | - | .233† |
| 6. Hireability | 3.86 | .73 | | | | | - |
| **Male applicants** | **M** | **M** | **SD** | **2** | **3** | **4** | **5** |
| 1. Attractiveness | 2.97 | .62 | .614** | .362** | .477** | .219† | .676** |
| 2. Competence | 3.92 | .48 | - | .470** | .564** | .522** | .753** |
| 3. Dominance | 3.55 | .68 | | - | -.153† | .036† | .394** |
| 4. Morality | 3.72 | .41 | | | - | .649** | .575** |
| 5. Sociability | 3.55 | .59 | | | | - | .530** |
| 6. Hireability | 3.71 | .72 | | | | | - |

**Note.** † $p > .06$, * $p < .05$, ** $p < .01$, *** $p < .001$.




**Graphic 2.** Mean by applicants' gender and participants' gender.

Two regression models were run entering all the five facial traits on female and male applicants' hireability (Table 11). For female applicants, attractiveness and competence significantly affected their hireability, while only attractiveness was significantly related to male applicants' hireability.

**Table 11.** Standardize regression coefficients of attractiveness, dominance, morality, and sociability on **hireability**, for female and male applicants**.**

| | **Female Applicants** | | | **Male Applicants** | | |
|---|---|---|---|---|---|---|
| | **β (SE)** | **t** | **95% CI** | **β (SE)** | **t** | **95% CI** |
| **Attractiveness** | .632*** (.091) | 6.557 | .413, .779 | .329** (.135) | 2.783 | .104, .649 |
| **Competence** | .286* (.182) | 2.052 | .007, .741 | .288† (.215) | 1.977 | -.008, .858 |
| **Dominance** | .070† (.111) | .659 | -.151, .297 | .154† (.127) | 1.278 | -.094, .419 |
| **Morality** | .038† (.175) | .351 | -.291, .414 | .145† (.267) | .956 | -.282, .792 |
| **Sociability** | .020† (.111) | .228 | -.198, .249 | .208† (.144) | 1.742 | -.040, .542 |
| | $R^2$= .834, $adjR^2$ = .816 $F(5, 44) = 44.329$, $p < .001$ | | | $R^2$= .693, $adjR^2$ = .659, $F(5, 44) = 19.902$, $p < .001$ | | |

**Note.** † $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

In order to test the relative importance of the four facets and attractiveness in predicting applicants' hireability, general dominance weights and relative important weights coefficients were computed (Table 12). For female applicants, general dominance weight and relative importance weight coefficients delineated a consistent ranking of importance, that is in descending order: facial attractiveness, facial competence, facial dominance, facial morality, and facial sociability. The bootstrap confidence intervals around the differences showed that general dominance weights coefficient of facial attractiveness was significantly higher than general dominance weights

coefficients of facial competence, dominance, morality and sociability. General dominance weights coefficient of facial competence was significantly higher than general dominance weights coefficients of facial dominance, morality and sociability. General dominance weights coefficient of facial dominance was significantly higher than general dominance weights coefficient of facial sociability. The difference between general dominance weights coefficients of facial dominance and facial morality was not significant, as well as the difference between general dominance weights coefficients of facial morality and facial sociability. The same pattern emerged with the bootstrap confidence intervals around the differences between relative importance weights coefficients, except that the difference between relative importance weights coefficients of facial competence and facial dominance was not significant.

**Table 12.** General dominance weights (GDW), relative important weights (RIW), and structure coefficients ($r_s$) for female and male applicants.

| | Female Applicants | | | Male Applicants | | |
|---|---|---|---|---|---|---|
| | **GDW** | **RIW** | $r_s$ | **GDW** | **RIW** | $r_s$ |
| **Attractiveness** | .391 a | .401 a | .960 | .188 ab | .194 ab | .812 |
| **Competence** | .245 b | .216 b | .853 | .216 a | .192 a | .904 |
| **Dominance** | .136 c | .143 bc | .642 | .071 b | .081 b | .473 |
| **Morality** | .040 cd | .047 cd | .334 | .117 ab | .117 ab | .690 |
| **Sociability** | .022 d | .027 d | .255 | .101 ab | .110 ab | .636 |

**Note.** Different letters in column mean significant differences among coefficients.

For male applicants' hireability, results of the general dominance weights showed the following predictors ranking: facial competence, facial attractiveness, facial morality, facial sociability, and facial dominance. According to the bootstrap confidence intervals around the differences, general dominance weights coefficient of facial competence was significantly higher

than general dominance weights coefficient of facial dominance. No other differences were significant. The relative importance weights outlined a slightly different predictors ranking, that is in descending order: facial attractiveness, facial competence, facial sociability, facial morality, and facial dominance. The bootstrap confidence intervals showed that relative importance weights coefficient of facial competence was significantly higher than relative importance weights coefficient of facial dominance.

Finally, structure coefficients of female applicants were compared to structure coefficients of male applicants. Correlations comparisons showed that attractiveness structure coefficient of female applicants was significantly higher than attractiveness structure coefficient of male applicants, $Z = 3.94$, $p < .001$, $n = 50$. Morality structure coefficient of female applicants was significantly lower than morality structure coefficient of male applicants, $Z = -2.43$, $p = .015$, $n = 50$. Sociability structure coefficient of female applicants was significantly lower than sociability structure coefficient of male applicants, $Z = -2.38$, $p = .017$, $n = 50$. All the other comparisons were not significant, all $p$s > .230.

**Discussion**

As in Study 2 and Study 3, the perfection bias hypothesis was not supported. Indeed, only attractiveness and competence betas instead of all the predictors were significant in the female applicants' regression model. In the male applicants' regression model, instead of competence and dominance, only attractiveness was a significant predictor. According to General Dominance and Relative Weights analyses, dominance was the least important predictor of male applicants' hireability, while differences among coefficients of female applicants emerged, meaning that facial traits were not considered equally important. Finally, correlations comparisons showed that only facial attractiveness was more important in evaluating women than men, not all the facial traits.

However, the pattern of the previous two experiments was confirmed. Indeed, attractiveness and competence were the two most important facial traits in affecting applicants hireability. This because both attractiveness and competence were significant predictors of female applicants'

hireability, and attractiveness was significantly associated to male applicants' hireability. General dominance as well as relative weights coefficients of attractiveness and competence were significantly higher than general dominance and relative weights coefficients of the other facial traits in the female applicants' model. Finally, beside significant tests, across all the analyses, attractiveness and competence had higher coefficients compared to the other facial traits.

Regarding the other three facial traits, morality and sociability were confirmed as more important predictors of male applicants' hireability than female applicants' hireability. According to correlations comparisons, facial morality and facial sociability structure coefficients for male applicants were significantly higher than facial morality and facial sociability structure coefficients for female applicants. Moreover, even if not significant, the predictors ranking outlined by general dominance and relative weights coefficients showed that facial morality and facial sociability were more important than dominance in affecting male applicants' hireability. Conversely, in female applicants' model, general dominance and relative weights coefficients of dominance was significantly higher than general dominance and relative weights coefficients of sociability and, besides significance, dominance coefficients were higher than both morality and sociability coefficients across all the metrics. Hence, facial dominance was considered more important in evaluating female applicants. In sum, besides attractiveness and competence, applicants' evaluations had been made on the basis of counter stereotypic facial traits. In sum, consistently with the last two studies, men and women were evaluated on facial attractiveness, facial competence and those facial traits on which they are stereotypically believed as weak.

To conclude, as in in the previous study, women were rated as more hireable than men and this could be explained under the framework of benevolent sexism or as a 'women are wonderful effect' (e.g., Glick et al., 2000).

# CHAPTER VI

## Mini meta-analysis

Study 2, Study 3 and Study 4 showed results different from expectations, but still the pattern was consistent. In order to gain more confidence about the estimated effects (Cumming, 2012, 2014), a mini meta-analysis was performed. Specifically, the aims were to examine across the last three studies: a) the relative importance of attractiveness, competence, dominance, morality and sociability in affecting female candidate hireability and male candidate hireability separately; b) the difference between female applicants and male applicants along each facial trait (i.e., attractiveness for female candidate vs attractiveness for male candidate; competence for female candidate vs competence for male candidate; so on…).

Using ProMeta 3.0 (idostatistics.com), numerous meta-analyses were performed in order to compare each facial trait with all the other facial trait (a) and to compare female applicants and male applicants on each facial trait (b)[14]. When a small (five or less) number of studies are included, it is recommended to conduct the analysis with both fixed model and random model because the estimation precision of the variance between studies decreases (Borenstein, Hedges, Higgins, & Rothstein, 2009). However, since sample size was equal across studies, the fixed effects model results and the random effects model results converged. Pearson's correlations between facial traits and hireability were entered as effect sizes, which were transformed in Fisher's z scores for analyses and then converted back to correlations for presentation.

Results are summarized in Table 13. Notably, all the total correlations (means of the correlations from Study 2, Study 3 and Study 4) of facial traits were highly significant.

[14] Heterogeneity tests resulted mostly not significant in the numerous meta-analyses performed here, meaning normal sampling variation across the three studies Nevertheless, note that the heterogeneity test has low power when few studies are included (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006).

**Table 13. Results of the mini meta-analysis.**

| | Attractiveness (95% CI) | Competence (95% CI) | Dominance (95% CI) | Morality (95% CI) | Sociability (95% CI) |
|---|---|---|---|---|---|
| **Female Applicants** | .86**a*** (.82, .90) | .80*a (.74, .86) | .59***b*** (.49, .70) | .34***c*** (.20, .49) | .28***c*** (.13, .44) |
| **Male Applicants** | .73***ab*** (.65, .80) | .77*a (.71, .84) | .40***c*** (.26, .54) | .60***bd*** (.50, .71) | .56***cd*** (.45, .67) |

**Note.** Total effect sizes (Pearson's correlations averaged from Study 2, Study 3 and Study 4) and 95% confidence intervals per each facial trait differentiated per applicants' gender; * $p < .001$. Different letters in row indicate significant differences ($p$s < .02) among facial traits for female and male applicants at the Q test. Within each column, values in bold and italics indicate significant differences ($p$s < .03) between female and male applicants at the Q test.

Examining the relative importance of facial traits (a) for female applicants, Q-tests highlighted that total effect size of facial attractiveness was significantly higher than total effect sizes of dominance, morality and sociability, all $Q$s > 27.07 and $p$s < .001, whereas no significant difference emerged with total effect size of competence, $Q(1) = 2.84$, $p = .092$. Total effect size of competence was significantly higher than total effect sizes of dominance, morality and sociability, all $Q$s > 12.37 and $p$s < .001. Total effect size of dominance was significantly higher than total effect sizes of morality and sociability, all $Q$s > 7.36 and $p$s < .007. No significant difference emerged between total effect sizes of morality and sociability, $Q(1) = .32$, $p = .571$.

For male applicants, total effect size of facial attractiveness was significantly higher than total effect sizes of dominance and sociability, all $Q$s > 5.73 and $p$s < .017, while no significant difference emerged with total effect size of competence and morality, all $Q$s < 3.35 and $p$s > .067. Total effect size of competence was significantly higher than total effect sizes of dominance, morality and sociability, all $Q$s > 7.53 and $p$s < .006. Total effect size of dominance was significantly higher than total effect size of morality, $Q(1) = 5.40$, $p = .02$, while no difference emerged with total effect size of sociability, $Q(1) = 3.10$, $p = .078$. Finally, no significant difference emerged between total effect sizes of morality and sociability, $Q(1) = .32$, $p = .573$.

Comparing female and male applicants on each facial trait (b), results of the Q-tests showed that total effect size of attractiveness was higher for female applicants than male applicants, $Q(1) =$

10.35, $p = .001$, as well as dominance, $Q(1) = 4.70$, $p = .03$. Total effect size of morality was higher for male applicants than female applicants, $Q(1) = 8.23$, $p = .004$, as well as sociability, $Q(1) = 8.26$, $p = .004$. Finally, no significant difference emerged between total effect sizes of competence for female and male applicants, $Q(1) = 0.38$, $p = .538$.

In sum, the mini meta-analysis confirmed results of the single studies. Hence, attractiveness and competence were the two most important facial traits in affecting both female and male applicants hireability. Moreover, dominance was more important than morality and sociability for female applicants hireability; vice versa morality and sociability were more important than dominance in affecting male applicants hireability.

# CHAPTER VII

## General discussion

The principal aim of the dissertation was to investigate the presence of a perfection bias at face perception level. Indeed, the present research extended previous findings by Moscatelli and colleagues (2020) and Prati and colleagues (2019) testing the perfection bias hypothesis with faces as stimuli, from which judgment criteria had been inferred. Moreover, Study 2, Study 3 and Study 4 not only tested for the first time the perfection bias hypothesis in leadership contexts, but to the best of the author's knowledge, they represent the first experiments in which all the four facets of the fundamental dimensions of social judgments perceived from faces were employed in the investigation of gender bias in the workplace.

The perfection bias hypothesis was confirmed by Study 1, which highlighted that women were evaluated on multiple criteria, since male applicants were judged only on their facial competence, while female applicants were judged on all the other facial traits (i.e., competence, morality and attractiveness). Hence, these findings suggested that while for men looking competent was enough, women needed to look competent, moral, and attractive in order to have the same chances to be hired. Accordingly, the Prestudy 1 test on the Area Sales Manager job description also showed that all the judgment criteria were considered as more important qualities to possess for a woman than a man in order to succeed in the role.

Nevertheless, in the subsequent studies in which the relative importance of the facial traits was directly examined, results consistently delineated a different pattern. Indeed, attractiveness and competence perceived from female and male applicants' faces were the two most important facial traits in affecting participants' hireability judgments. More surprisingly, compared to facial morality and sociability, facial dominance was more important in affecting women's likelihood of being hired. Conversely, facial morality and facial sociability were more important than facial dominance in affecting men's likelihood of being hired. These results were also corroborated by a mini meta-analysis. According to the classical literature on gender bias, these latter results are quite surprising.

**The 'deficit bias'**

As reviewed in Chapter I, women are seen as less agentic/competent and more communion/warm than men (e.g., Cuddy et al., 2007). Consequently, since leadership requires agentic traits, women face a 'Catch 22' situation where they are discriminated against for leadership positions because they are lacking in agentic traits, but they are also discriminated against in trying to behave more agentic (e.g., Rudman et al., 2012; Schein, 2001). However, the main findings of the present research highlighted that competence and dominance (i.e., facets of agency) were more important than morality and sociability (i.e., facets of communion) in evaluating female applicants.

The emergence of facial competence as one of the more important judgment criteria for female applicants' hireability could be explained with recent research on gender stereotypes, according to which the female stereotype has acquired more competence over the years (Eagly et al., 2019; Hentschel et al., 2019). Facial competence, however, was an important criterion for evaluating both men and women's hireability. Thus, the relevance of competence was probably due to the prominence of this facet in the work context (e.g., Brambilla et al., 2011). More confusing was the prominence of facial dominance (i.e., as facet of agency) for female applicants' hireability, since women are still perceived as less agentic than men (Eagly et al., 2019; Hentschel et al., 2019) and should not exhibit agentic traits (e.g., Rudman et al., 2012). In the same vein, given men are still considered as less communal/warm than women (Eagly et al., 2019; Hentschel et al., 2019), the importance of morality and sociability (i.e., as facet of communion/warmth) in evaluating male applicants was surprising. In other words, besides attractiveness and competence, the main findings suggested a sort of 'deficit bias', so that men and women were evaluated on those judgment dimensions on which they are stereotypically perceived as weak.

At this point one could argue that, while agency/competence traits are traditionally associated with leadership, communion/warmth traits are usually not. Therefore, given that morality and sociability are not considered relevant for leaders, the fact that these facets were more important than dominance in affecting male applicants' hireability appeared inconsistent within the

framework of classical literature. However, in the aforementioned Ryan and colleagues' (2011) work, it was found that while male managers of successful companies were characterized by being dominant, female managers of successful companies were characterized by being tactful and understanding. Hence, somehow communion/warmth traits can be relevant for leadership. Moreover, recent literature on leadership styles seems to suggest that leaders should possess qualities which echo communion traits. Indeed, among those styles that have been receiving increasing attention in the last two decades (Anderson & Sun, 2017), there is for instance the Authentic leader, who should self-regulate her/him selves in accordance to internalized moral standards and values (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008). Another example is the Servant leadership style (Dennis & Bocarnea, 2005; Russell & Gregory Stone, 2002; Spears, 2002), which requires moral and unconditional love to others who should be treated as individuals and not as means to an end. Similarly, the Ethical leader is defined as a moral person, that is a fair, honest, and trustworthy decision-maker, as well as a moral role model, a person who behaves accordingly with what he/she says (Brown, Treviño, & Harrison, 2005; Mayer, Kuenzi, Greenbaum, Bardes, & Salvador, 2009). Finally, even the well-known Transformational leadership style (Bass, 1985), which is a conceptualization of the charismatic and dominant leadership (Judge, Woolf, Hurst, & Livingston, 2008), implies 'individualized consideration', that is "the degree to which the leader attends to each follower's needs, acts as a mentor or coach to the follower, and listens to the follower's concerns and needs" (Judge & Piccolo, 2004, p. 755). Besides leadership, evidence of an increasing relevance of emotional and social skills in the work context has been registered. For example, the Organization for Economic Co-operation and Development (OECD, 2017) reports how social and emotional skills influence diverse job outcomes, such as occupational status and salaries, and are equally important as cognitive skills in determining future employments. In sum, since morality and sociability facets of communion/warmth dimension are becoming more relevant in the work context as well as for leaders and given that men are still seen as lacking in

communion/warmth traits, it makes sense that these facets were perceived as more relevant than dominance in evaluating male applicants' hireability.

Hence, participants seemed to rely on both gender and job beliefs in evaluating candidates. Indeed, since leaders are required to be competent and dominant as well as moral and sociable, participants unconsciously evaluated both female and male candidates along all the facial traits considered. Then, probably referring to gender stereotypes, they mostly grounded their judgments on those traits believed as weak in candidates, namely: dominance in women, morality and sociability in men. This is essentially the idea underling the Lack of Fit Model (e.g., Heilman, 2001) and Role Congruity Theory (Eagly & Karau, 2002), according to which evaluators need to perceive a fit/congruence between job characteristics and candidate's 'presumed' skills. However, these theorizations were more focused on women being evaluated as less competent than men and consequently lacking the fit/congruence with job requirements. In the present research though, also male candidates were scrutinized on those traits which would possibly not fit the job requirements.

The mechanism of focusing on those qualities considered as relevant for task accomplishment and on which individuals are expected to be weak is also not unknown. Indeed, the Double Standards Theory (Foschi, 1992, 2000) posits that gender is a diffuse socially significant characteristic that generates different performance expectations (i.e., women will perform worse than men) and consequent status attributions (women are lower status group than men), which in turn produce different standards for assessing ability (i.e., women need to provide more evidence of their ability). The theory, though, considers gender as only one of the different characteristics that trigger performance expectations and it does not take into account gender stereotype contents (e.g., dominance, sociability). Gender stereotypes, instead, have a key role in the Shifting Standards Model (Biernat, 2012). In this model, stereotypes set standards of judgments against which members of the groups are compared to and, since women are perceived as less competent than men, they need to provide less evidence to be judged as competent. Hence, only competence is taken into account, not different facets of the fundamental content dimensions of social judgments.

Moreover, both models focus on women's low abilities and not on men's lacking abilities, whereas the deficit bias in the present research affected both female and male candidates.

**Theoretical and practical implications**

From a theoretical point of view, these findings offer new insights on the effect of gender stereotypes in the workplace. Indeed, in considering different facets of the agency dimension, they suggest that the source of gender discrimination for women could be related to the dominance facets, not to the competence facets. This is in line with recent evidence showing few changes in the gender stereotypes content and, specifically, with the increase in the levels of competence stereotypically attributed to women (Eagly et al., 2019; Hentschel et al., 2019). The present research went further in showing how these changes could affect evaluations on women in the workplace.

These findings also added to facial first impressions and their consequences line of research. Indeed, while the prominence of facial competence into the work context is recognized, this research showed how a specific job description can sway facial traits relevance (i.e., morality versus sociability). Moreover, along with the specific domain, the diverse importance of facial dominance, facial morality, and facial sociability for female and male applicants' hireability was probably due to the influence of gender stereotypes. Thus, this research offers new evidence on the interplay among facial traits inferences, a specific domain and gender in determining traits power to affect people' s judgments, as auspicated by Todorov and colleagues (2015).

From a more practical point of view, the present work has confirmed previous findings on the influence of facial first impressions in the workplace (e.g., Rule & Ambady, 2008b, 2009) and more specifically in personnel selection processes. In particular, these findings suggest the possibility that HR professionals can make biased hiring decisions, especially since they are massively exposed to applicants' photos. During real recruitment and selection processes, indeed, CVs are analysed in several stages (Costa & Gianecchini, 2009) and they usually contain the applicants' photos. Even if attaching photos on CVs is forbidden in some Countries (e.g., in the UK), HR professionals frequently search applicants' social network accounts (Hoffman &

Casnocha, 2012; Napolitano, 2010; Soumitra, 2010), such as on Facebook, Linkedin or Twitter, where photos are normally posted. They do that in order to infer applicants' personality characteristics, such as extraversion and maturity (Caers & Castelyns, 2011). However, as briefly mentioned in Chapter I, traits unconsciously inferred from faces can be misleading relative to the actual individuals' personality (Todorov et al., 2015). Thus, the risk is to hire a person that might not actually fit job requirements. Especially because, although HR professionals are aware of the biasing effect of attractiveness and appearance (e.g., clothes), they probably have no knowledge of the biasing influence of personal and social traits inferred from faces.

**Limitations and future directions**

While the first experiment of this research confirmed a perfection bias at face perception level, the other three experiments showed a deficit bias. These diverging results can be due to the different materials and methodologies used. Firstly, compared to Study 1, leadership job descriptions were employed in the last three studies. Moreover, as repeatedly stated above, Moscatelli and colleagues (2020) as well as Prati and colleagues (2019) have found the perfection bias considering non managerial roles. These findings seem to suggest that the perfection bias works in evaluations of applicants for lower level jobs than managerial positions. However, further research is needed to investigate this possibility.

Secondly, the experimental methodologies employed were quite different: in Study 1, facial traits ratings were collected during the main experiment (i.e., a work context) and only on attractiveness, competence and morality. In Study 2, 3, and 4, ratings on all the four facets were instead collected with a pre-test (i.e., neutral context). Given that different domains can affect the predictive power of facial traits (Todorov et al., 2015), the different results on the relevance of facets in the studies can be due to the different contexts in which the facial first impression ratings were collected. Moreover, the fact that facial traits are highly correlated, specifically in female faces (Oh, Dotsch, et al., 2019), it is well known and this could represent a problem in interpreting regression analyses results (Kraha et al., 2012). Even if no severe multicollinear problems were

detected in this research, it is possible that high correlations among the four facets have played a role in differentiating the results obtained in Study 1 compared to the rest of the studies. Therefore, more studies are needed to test the perfection bias as well as the deficit bias, perhaps employing different stimuli (e.g., written descriptions, instead of faces) and varying the experimental methodology. For example, in trying to avoid high correlation issues, it would be interesting to test the four facets relevance using averaging - morphing techniques, as used by Sutherland and colleagues (2013, 2016), or more sophisticated computer software, as used by Todorov and colleagues (2008, 2010) or Walker and Vetter (2009, 2016), in order to create face models for each facet.

Concerning methodology, the consistent results obtained about the deficit bias could be due to the fact that the same ratings collected in an Italian sample have been used as predictors in all the three last studies on ratings collected in British samples. Given that people judgments tend to converge about facial first impressions, a fact that has been established since Secord's work (1958), and since not so many cultural differences have emerged in literature (Todorov et al., 2015), this methodological choice looked appropriate as well as convenient. As stated above, the reason for collecting facial first impression ratings in a separate session was to collect 'pure' facial first impression ratings, not biased by the context. Additionally, this allowed to significantly reduce the duration of the experiments and, consequently, to avoid the 'respondent fatigue' issue, which is basically a low quality of the responses collected due to participants' tiredness (Lavrakas, 2008). Indeed, rating 100 photos on three items, instead of 14 (three traits per each of the four facets and two hireability items), makes a considerable difference in terms of experimental duration. Nevertheless, cultural differences can affect facial first impressions and need to be further explored (Todorov et al., 2015). For example, future studies could verify whether collecting facial first impressions data from diverse samples, cross-culturally heterogeneous and homogenous compared to the hireability ratings samples, would show a similar pattern of results.

Another limitation can be the 50 female photos and the 50 male photos that did not significantly differ on attractiveness, competence, dominance, morality and sociability perceived from faces and the ratings were quite low (i.e., means around or above the middle point). This methodological choice was made in order to avoid findings due to the different levels of traits perceived from faces and not to differences between female and male applicants. In the real world though, men are perceived as more agentic and less warm than women (e.g., Eagly et al., 2019). This is perhaps the reason why no actual discrimination against women emerged in this research, with female applicants even rated as more hireable than male applicants in Study 3 and 4. Therefore, future research should investigate whether employing a more realistic and representative sample of female and male photos, in terms of stereotypical traits levels perceived from faces, the same results would emerge. Moreover, low levels of dominance perceived from female faces may have hampered a possible backlash effect (e.g. Rudman et al., 2012), which has been similarly found in face perception research (e.g., Sutherland et al., 2015). Thus, future results could verify whether women exhibiting high levels of facial dominance would not only be less liked but also discriminated against in hiring processes.

Speaking of ecological validity, HR professionals are usually provided with more pieces of information about job applicants, such as a longer and more detailed CV and a motivation letter. According to literature, while more pieces of information should mitigate facial bias, the effect persists (Todorov et al., 2015). For example, in economic/strategic games, participants prefer to invest in players with trustworthy faces rather than players with untrustworthy faces even when they are informed about players' past behaviour that contradicts their reliability (Rezlescu et al., 2012). This was the reasoning underlying the choice to provide participants with only applicants' facial photos in Study 2, 3 and 4. However, research should experimentally prove this claim in the work context. Going further, it would be interesting to investigate how applicants' facial traits, varying in terms of perceived levels (high vs low), and information, varying in terms of quantity (more vs less) as well as quality (high vs low), could interact in determining hiring decisions. For example, future

studies could examine whether high levels of facial competence could be more relevant than CVs and motivation letters, which instead convey low levels of competence, in affecting job applicants' hireability.

To conclude, a difficulty in developing the present research was the labelling choice for the four facets (e.g., communion or warmth?) and which traits select in order to represent them (e.g., self-confident, ambitious, independent, and so on as traits representing dominance). Indeed, it is worth noting that the three lines of research considered in this project still diverge, despite the numerous overlapping aspects, and this creates confusion in the operationalisation of the dimensions and their facets. Clearly, these differences reflect the enormous complexity of the issues addressed, but sometimes they just look like as a 'jungle fallacy' (Kelley, 1927; in Block, 1995) which is the tendency to use different labels to identify the same phenomenon. For example, it has been proven that facial trustworthiness corresponds to communion/warmth dimension of social judgments (e.g., Sutherland et al., 2016). This would suggest that no matter the labels or stimuli (i.e., written descriptions or faces), first impressions are formed by the same underlying dimensions. The labels issue is much more evident in the literature about the two fundamental content dimensions of social judgments where several efforts have been made in order to provide a more comprehensive conceptualization (e.g., Abele et al., 2016). However, despite the intention to consider different facets, recent research on gender stereotypes has failed to consider all the four facets highlighted in the literature of the fundamental dimensions. Indeed, both Eagly and colleagues (2019) as well as Hentschel and colleagues (2019) did not distinguish between morality and sociability as facets of communion/warmth dimension (e.g., Leach et al., 2008). In a similar vein, Oh, Buck, and Todorov (2019) concluded that facial attractiveness, facial confidence and facial masculinity are components of competence, which is a stimulating finding in uncovering interaction mechanisms among facial traits as well as confusing from a theoretical point of view. Indeed, confidence and masculinity are two traits that fall into dominance as the dimension underlying facial first impressions (e.g., Oostherhof & Todorov, 2008), and confidence and

dominant are two traits that fall into the agency/competence dimension which represents the content of male stereotypes (e.g., Schein, 2001).

Obviously, these issues would deserve more considered discussion, and this lies beyond the scope of the present dissertation. The point is to stimulate more effort in trying to find a comprehensive and integrated conceptualization about the two fundamental content dimensions and their facets. For example, Sutherland and colleagues (2016) did not find a convergence between dominance (i.e., as facial dimension) and competence (i.e., as fundamental dimension of social judgment). This could be due to the fact that they actually represented two different facets of the agency dimension, as conceptualized by Abele and colleagues (2016). On the other hand, this does not mean that research should neglect differences between facial and social models. For instance, Sutherland and colleagues (2016) recognized attractiveness as a fundamental trait in shaping facial first impressions. However, it is difficult to imagine an experiment in which attractiveness would be manipulated with written descriptions. Furthermore, the Prestudy 2 of the present research highlighted how people underestimate the relevance of attractiveness in affecting their judgments, despite its prominent influence as a facial trait emerged in all the studies reported. In sum, future research could aim to gain a more consistent conceptualization about the dimensions and their facets underlying social judgments and first impressions, in order to have a solid operationalisation applicable to different lines of research, such as facial bias, gender stereotypes and so on.

## Conclusion

This dissertation begins emphasizing the persistence of gender inequality, especially for leadership positions. Hentschel and colleagues (2019) have recently argued for the importance of subdividing the two fundamental dimensions of social judgments in different facets in order to improve and specify our understanding on gender stereotypes and gender bias. The present research purposely considered all the four facets (i.e., competence, dominance, morality, and sociability) highlighted by Abele and colleagues (2016). In doing this, it was possible to uncover an original deficit bias, that is men and women being evaluated on those stereotypic traits in which they are

believed weak. Hence, even if the deficit bias needs more evidence to be confirmed, in considering the four facets, the present research has demonstrated that Hentschel and colleagues' (2019) argument was right.

## References


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.

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology*, *47*(5), 935–948.

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*(7), 1063–1065.

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the Fundamental Content Dimensions: Agency with Competence and Assertiveness—Communion with Warmth and Morality. *Frontiers in Psychology*, *7*, 1810.

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*(7), 1202–1217.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751.

Abele, A. E., & Wojciszke, B. (2014). Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A Dual Perspective Model. *Advances in Experimental Social Psychology*, *50*, 195–255.

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology*, *42*(4), 619.

Adams, R. B., & Kirchmaier, T. (2016). Women on boards in finance and STEM industries. *American Economic Review*, *106*(5), 277–81.

Agars, M. D. (2004). Reconsidering the impact of gender stereotypes on the advancement of women in organizations. *Psychology of Women Quarterly*, *28*(2), 103–111.

Alexander, M. G., Brewer, M. B., & Hermann, R. K. (1999). Images and affect: A functional analysis of out-group stereotypes. *Journal of Personality and Social Psychology*, *77*(1), 78.

Anderson, M. H., & Sun, P. Y. (2017). Reviewing leadership styles: Overlaps and the need for a new 'full-range'theory. *International Journal of Management Reviews*, *19*(1), 76–96.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science*, *323*(5918), 1183–1183.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*(2), 129.

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.*

Bales, R. F., & Parsons, T. (1955). *Family: Socialization and interaction process*. Glencoe, IL: Free Press.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948–17953.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*(2), 272.

Barbaranelli, C., & D'Olimpio, F. (2007). *Analisi dei dati con SPSS*. LED.

Barreto, M. E., Ryan, M. K., & Schmitt, M. T. (2009). *The glass ceiling in the 21st century: Understanding barriers to gender equality*. American Psychological Association.

Bass, B.M. (1985). *Leadership and Performance Beyond Expectations*. New York, NY: Free Press.

Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch Jr, M. (1977). Status characteristics and expectation states: A graph-theoretical formulation. *Status Characteristics and Social Interaction: An Expectations States Approach*, 91–134.

Berger, J., & Zelditch Jr, M. (1998). *Status, power, and legitimacy: Strategies and theories*. Transaction Publishers.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*(1), 3.

Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In *Stereotype accuracy: Toward appreciating group differences* (pp. 87–114). Washington, DC: American Psychological Association.

Biernat, M. (2012). Stereotypes and shifting standards: Forming, communicating, and translating person impressions. In *Advances in experimental social psychology* (Vol. 45, pp. 1–59). Elsevier.

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues*, *57*(4), 707–724.

Biernat, M., & Kobrynowicz, D. (1997). Gender-and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*(3), 544.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*(4), 485.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*(6), 1142.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological bulletin*, *117*(2), 187.

Borenstein, M.,Hedges, L. V.,Higgins, J. P. T.,&Rothstein,H. R. (2009). *Introduction to Meta-analysis*. UK:Wiley.

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34-42.

Brambilla, M., & Leach, C. W. (2014). On the Importance of Being Moral: The Distinctive Role of Morality in Social Judgment. *Social Cognition*, *32*(4), 397–408.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135–143.

Brambilla, M., Sacchi, S., Menegatti, M., & Moscatelli, S. (2016). Honesty and dishonesty don't move together: Trait content information influences behavioral synchrony. *Journal of Nonverbal Behavior*, *40*(3), 171-186.

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*(5), 811-821.

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British journal of social psychology*, *51*(1), 149-166.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-Role Stereotypes: A Current Appraisal 1. *Journal of Social Issues*, *28*(2), 59–78.

Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, *97*(2), 117–134.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*(3), 542.

Buss, D. M. (2005). *The handbook of evolutionary psychology*. Hoboken, NJ: Wiley Online Library.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge/Taylor & Francis Group.

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*.

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.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Chiao, J. Y., Bowman, N. E., & Gill, H. (2008). The political gender gap: Gender bias in facial inferences that predict voting behavior. *PloS One*, *3*(10), e3666.

Correll, S. J., & Ridgeway, C. L. (2003). Handbook of social psychology. *New York: Kluwer Academic Publishers*, 29–51.

Costa, M., & Corazza, L. (2006). *Psicologia della bellezza*. Firenze-Milano: Giunti Editore.

Costa, G., & Gianecchini, M. (2009). *Risorse Umane. Persone, relazioni e valore*. Milano: McGraw-Hill.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*(2), 229–248.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, *40*, 61-149.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J. P., ... Sleebos, E. (2009). Is the stereotype content model culture-bound? A cross-cultural comparison reveals systemiatic similarities and differences. *British Journal of Social Psychology, 48*(1), 1-33.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta–analysis*. New York, NY: Routledge.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*(2), 225–248.

De Bruin, E. N., & Van Lange, P. A. (2000). What people look for in others: Influences of the perceiver and the perceived on information selection. *Personality and Social Psychology Bulletin*, *26*(2), 206–219.

Deaux, K., & Lewis, L. L. (1983). Components of gender stereotypes. *Psychological Documents*, *13*(2), 25–50.

Dennis, R. S., & Bocarnea, M. (2005). Development of the servant leadership assessment instrument. *Leadership & Organization Development Journal*, *26*(8), 600–615.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285.

Dubois, N., & Beauvois, J.-L. (2005). Normativeness and individualism. *European Journal of Social Psychology*, *35*(1), 123–146.

Eagly, A. H. (1987). *John M. MacEachran memorial lecture series; 1985. Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology*, *5*(1), 1–35.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American Psychologist*.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735.

Eagly, A. H., & Wood, W. (2011). Feminism and the evolution of sex differences and similarities. *Sex Roles*, *64*(9–10), 758–767.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *Journal of cognitive neuroscience*, *19*(9), 1508-1519.

European Commission, (2018). *2018 Report on equality between women and men in the EU*. Retrieved from: https://ec.europa.eu

European Institute for Gender Equality, (2017). *Gender Equality Index 2017 – Main findings*. Retrieved from: http://eige.europa.eu/rdc/eige-publications/gender-equality-index-2017-measuring-gender-equality-european-union-2005-2015-main-findings

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. *The Handbook of Social Psychology*, *2*(4), 357–411.

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.

Foschi, M. (1992). Gender and double standards for competence. In *Gender, interaction, and inequality* (pp. 181–207). Springer.

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, *26*(1), 21–42.

Foschi, M., & Foddy, M. (1988). Standards, performances, and the formation of self-other expectations. In M. Webster, Jr. & M. Foschi (Eds.), *Status generalization: New theory and research* (pp. 248-260). Palo Alto, CA: Stanford University Press.

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management 1. *Journal of Applied Social Psychology*, *21*(13), 1039–1057.

Fruhen, L. S., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *The Leadership Quarterly*, *26*(6), 1005–1016.

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, *101*(1), 109.

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of Women Quarterly*, *21*(1), 119–135.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*(2), 109.

Glick, P., & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, *35*(3), 530–535.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., … Alao, A. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*(5), 763.

Glick, P., Lameiras, M., Fiske, S. T., Eckes, T., Masser, B., Volpato, C., … Sakalli-Uğurlu, N. (2004). Bad but bold: Ambivalent attitudes toward men predict gender inequality in 16 nations. *Journal of Personality and Social Psychology, 86*(5), 713.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535-549.

Graham, J. R., Harvey, C. R., & Puri, M. (2016). A corporate beauty contest. *Management Science*, *63*(9), 3044–3056.

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 40*(3), 353–363.

Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (2009). Shallow cues with deep effects: Trait judgments from faces and voting decisions. *The Political Psychology of Democratic Citizenship, 1*, 583–605.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*(4), 657–674.

Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality, 10*(4), 237.

Heilman, M. E., Manzi, F., & Braun, S. (2015). Presumed incompetent: Perceived lack of fit and gender bias in recruitment and selection. *Handbook of Gendered Careers in Management: Getting in, Getting on, Getting Out*, *90*, 90–104.

Heilman, M. E., & Saruwatari, L. R. (1979). When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and nonmanagerial jobs. *Organizational Behavior and Human Performance*, *23*(3), 360–372.

Heilman, M. E., & Stopeck, M. H. (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. *Organizational Behavior and Human Decision Processes*, *35*(2), 202–215.

Henderson, J. J., & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior*, *24*(5), 351–356.

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves. *Frontiers in Psychology*, *10*.

Hoffman, R., & Casnocha, B. (2012). *Teniamoci in contatto. La vita come impresa*. Milano: Egea.

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology*, *56*(2), 431–462.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index?. *Psychological Methods*, *11*, 193–206.

Iachini, T., Pagliaro, S., & Ruggiero, G. (2015). Near or far? It depends on my impression: Moral information and spatial behavior in virtual interactions. Acta psychologica, 161, 131-136.Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, *25*(14), R621–R634.

Jackson, D., Engstrom, E., & Emmers-Sommer, T. (2007). Think leader, think male and female: Sex vs. seating arrangement as leadership cues. *Sex Roles*, *57*(9–10), 713–723.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 1–19.

Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods*, *7*(3), 283–299.

Judge, T.A. and Piccolo, R.F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *Journal of Applied Psychology*, 89, pp. 755–768.

Judge, T.A.,Woolf, E.F., Hurst, C. and Livingston, B. (2008). Leadership. In Barling, J. and Cooper, C.L. (eds), *The SAGE Handbook of Organizational Behavior: Volume 1, Micro Approaches*. Los Angeles, CA: Sage, pp. 334–352.

Kelley, E. L. (1927). *Interpretation of educational measurements*. Yonkers, NY: World.

Kelloway, E. K. (2014). *Using Mplus for structural equation modeling: A researcher's guide*. Thousand Oaks, CA, US: Sage Publications.

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, *40*(3), 414.

Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, *11*(1), 83–96.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371.

Kraha, A., Turner, H., Nimon, K., Zientek, L., & Henson, R. (2012). Tools to Support Interpreting Multiple Regression in the Face of Multicollinearity. *Frontiers in Psychology*, *3*, 44.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390.

Langlois, J. H., Ritter, J. M., Casey, R. J., & Sawin, D. B. (1995). Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology*, *31*(3), 464.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publications. Retrievd from: https://methods.sagepub.com

Leach, C. W., Carraro, L., Garcia, R. L., & Kang, J. J. (2017). Morality stereotyping as a basis of women's in-group favouritism: An implicit approach. *Group Processes & Intergroup Relations, 20*, 153-172.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234–249.

Lee, J., & Hoon, T. H. (1993). Business Students' Perceptions of Women in Management-the Case in Singapore. *Management Education and Development*, *24*(4), 415–429.

Little, A. C. (2014). Facial attractiveness. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 621–634.

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.

Lorenzo-Seva, U., Ferrando, P. J., & Chico, E. (2010). Two SPSS programs for interpreting multiple regression results. *Behavior Research Methods*, *42*(1), 29–35.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet*, *91*, 630.

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, *91*(4), 777.

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*(4), 1122–1135.

Massengill, D., & Di Marco, N. (1979). Sex-role stereotypes and requisite management characteristics: A current replication. *Sex Roles*, *5*(5), 561–570.

Mattes, K., Spezio, M., Kim, H., Todorov, A., Adolphs, R., & Alvarez, R. M. (2010). Predicting election outcomes from positive and negative trait assessments of candidate images. *Political Psychology*, *31*(1), 41–58.

Mayer, D. M., Kuenzi, M., Greenbaum, R., Bardes, M., & Salvador, R. B. (2009). How low does ethical leadership flow? Test of a trickle-down model. *Organizational Behavior and Human Decision Processes*, *108*(1), 1–13.

Mazur, A., Mazur, J., & Keating, C. (1984). Military rank attainment of a West Point class: Effects of cadets' physical features. *American Journal of Sociology*, *90*(1), 125–150.

McArthur, Leslie Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215.

McArthur, Leslie Zebrowitz, & Apatow, K. (1984). Impressions of baby-faced adults. *Social Cognition*, *2*(4), 315–342.

McKee, J. P., & Sherriffs, A. C. (1957). The differential evaluation of males and females. *Journal of Personality*, *25*(3), 356–371.

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. In *Advances in experimental social psychology* (Vol. 30, pp. 93–161). Elsevier.

Morrow, P. C., McElroy, J. C., Stamper, B. G., & Wilson, M. A. (1990). The effects of physical attractiveness and other demographic characteristics on promotion decisions. *Journal of Management*, *16*(4), 723–736.

Moscatelli, S., Menegatti, M., Ellemers, N., Mariani, M. G., & Rubini, M. (2020). Men Should Be Competent, Women Should Have it All: Multiple Criteria in the Evaluation of Female Job Candidates. *Sex Roles*, 1-20.

Muller, U., & Mazur, A. (1997). Facial dominance in Homo sapiens as honest signaling of male quality. *Behavioral Ecology*, *8*(5), 569–579.

Muthén, L. K., & Muthén, B. O. (2017). *1998–2017. Mplus User's Guide.* Los Angeles, CA: Muthén & Muthén.

Napolitano, A. (2011). *Linkedin. La Rete per trovare il lavoro dei sogni*. Milano: Apogeo.

Nimon, K., Oswald, F., & Roberts, J. K. (2013). yhat: Interpreting Regression Effects. R package version 2.0-0. *Yhat: Interpreting Regression Effects. R Package Version 2.0-0.*

OECD, (2017). UPDATED Social and Emotional Skills - Well-being, connectedness and success. Retrieved from: https://www.oecd.org

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science*, *30*(1), 65–79.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2019). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*.

Oldmeadow, J., & Fiske, S. T. (2007). System-justifying ideologies moderate status= competence stereotypes: Roles for belief in a just world and social dominance orientation. *European Journal of Social Psychology*, *37*(6), 1135–1148.

Oldmeadow, J. A., Sutherland, C. A., & Young, A. W. (2013). Facial stereotype visualization through image averaging. *Social Psychological and Personality Science*, *4*(5), 615-623.

Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2019). Good to Bad or Bad to Bad? What is the relationship between valence and the trait content of the Big Two? *European Journal of Social Psychology*.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110.

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

Pagliaro, S., Brambilla, M., Sacchi, S., D'Angelo, M., & Ellemers, N. (2013). Initial impressions determine behaviours: Morality predicts the willingness to help newcomers. *Journal of Business Ethics*, *117*(1), 37–44.

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, *99*(6), 1129.

Peeters, G. (1992). Evaluative meanings of adjectives invitro and in context-some theoretical implications and practical consequences of positive-negative asymmetry and behavioral-adaptive concepts of evaluation. *Psychologica Belgica*, *32*(2), 211–231.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., … Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*(6696), 884.

Pflüger, L. S., Oberzaucher, E., Katina, S., Holzleitner, I. J., & Grammer, K. (2012). Cues to fertility: Perceived attractiveness and facial shape predict reproductive success. *Evolution and Human Behavior*, *33*(6), 708–714.

Phalet, K., & Poppe, E. (1997). Competence and morality dimensions of national and ethnic stereotypes: A study in six eastern-European countries. *European Journal of Social Psychology*, *27*(6), 703–723.

Pillemer, J., Graham, E. R., & Burke, D. M. (2014). The face says it all: CEOs, gender, and predicting corporate performance. *The Leadership Quarterly*, *25*(5), 855–864.

Poppe, E., & Linssen, H. (1999). In-group favouritism and the reflection of realistic dimensions of difference between national states in Central and Eastern European nationality stereotypes. *British Journal of Social Psychology*, *38*(1), 85–102.

Porter, S., Brinke, L. ten, & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477–491.

Porter, S., England, L., Juodis, M., Ten Brinke, L., & Wilson, K. (2008). Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *40*(3), 171.

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology*, *45*(5), 1132–1135.

Prati, F., Menegatti, M., Moscatelli, S., Kana Kenfack, C. S., Pireddu, S., Crocetti, E., … Rubini, M. (2019). Are Mixed-Gender Committees Less Biased Toward Female and Male Candidates? An Investigation of Competence-, Morality-, and Sociability-Related Terms in Performance Appraisal. *Journal of Language and Social Psychology*.

Quereshi, M. Y., & Kay, J. P. (1986). Physical attractiveness, age, and sex as determinants of reactions to resumes. *Social Behavior and Personality: An International Journal*, *14*(1), 103–112.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS One*, *7*(3), e34293.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology*, *91*(1), 125–140.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283.

Rosenkrantz, P., Vogel, S., Bee, H., Broverman, I., & Broverman, D. M. (1968). Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology*, *32*(3), 287.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105.

Rule, N. O., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, *19*(2), 109–111.

Rule, N. O., & Ambady, N. (2009). She's got the look: Inferences from female chief executive officers' faces predict their success. *Sex Roles*, *61*(9–10), 644–652.

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PloS One*, *5*(1), e8733.

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. O., Bjornsdottir, R. T., Tskhay, K. O., & Ambady, N. (2016). Subtle perceptions of male sexual orientation influence occupational opportunities. *Journal of Applied Psychology*, *101*(12), 1687.

Russell, R. F., & Gregory Stone, A. (2002). A review of servant leadership attributes: Developing a practical model. *Leadership & Organization Development Journal*, *23*(3), 145–157.

Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. (2011). Think crisis–think female: The glass cliff and contextual variation in the think manager–think male stereotype. *Journal of Applied Psychology*, *96*(3), 470.

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*, *22*(9), 1183–1190.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, *57*(2), 95.

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology*, *60*(3), 340.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, *57*(4), 675–688.

Schein, V. E. (2007). Women in management: Reflections and projections. *Women in Management Review*, *22*(1), 6–18.

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. *Person Perception and Interpersonal Behavior*, 300–315.

Secord, P. F., Bevan, W., & Dukes, W. F. (1953). Occupational and physiognomic stereotypes in the perception of photographs. *Journal of Social Psychology*, *37*(2), 261.

Secord, P. F., Bevan, W., & Katz, B. (1956). The Negro stereotype and perceptual accentuation. *The Journal of Abnormal and Social Psychology*, *53*(1), 78.

Secord, P. F., Dukes, W. F., & Bevan, W. (1954). Personalities in faces: I. An experiment in social perceiving. *Genetic Psychology Monographs*.

Shackelford, T. K., Schmitt, D. P., & Buss, D. M. (2005). Universal dimensions of human mate preferences. *Personality and Individual Differences*, *39*(2), 447–458.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of personality and social psychology*, *52*(4), 689.

Soumitra, A. (2010). What's your personal social media strategy. *Harvard business review*, *88*(11), 127-130.

Spears, L.C. (2002). Introduction: tracing the past, present and future of servant-leadership. In L.C. Spears, (Ed.), *Focus on Leadership* (pp. 1–18.) New York, NY: John Wiley.

Spence, J. T., & Hahn, E. D. (1997). The attitudes toward women scale and attitude change in college students. *Psychology of Women Quarterly*, *21*(1), 17–34.

Spence, J. T., & Helmreich, R. L. (1972). *The Attitudes Toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society*. American Psychological Association.

Spezio, M. L., Rangel, A., Alvarez, R. M., O'Doherty, J. P., Mattes, K., Todorov, A., … Adolphs, R. (2008). A neural basis for the effect of candidate appearance on election outcomes. *Social Cognitive and Affective Neuroscience*, *3*(4), 344–352.

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

Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, *157*, 257–267.

Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PloS One*, *11*(2), e0148284.

Team, R. D. C. (2013). R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *URL Http://Www. R-Project. Org. R Foundation for Statistical Computing, Vienna, Austria*.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*(1), 208–224.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton, NJ: Princeton University Press.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). *Social judgments from faces. Current opinion in neurobiology*, *23*(3), 373-380.

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*(1), 519–545.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, *27*(6), 813–833.

Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (2011). Task-invariant brain responses to the social value of faces. *Journal of Cognitive Neuroscience*, *23*(10), 2766-2781.

Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. Personality and Social Psychology Bulletin, 25(6), 687-697.van den Brink, M., & Benschop, Y. (2011). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, *19*, 507–524.

Von Stockhausen, L., Koeser, S., & Sczesny, S. (2013). The gender typicality of faces and its impact on visual processing and on hiring decisions. *Experimental Psychology*.

Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, *9*(11), 12–12.

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, *110*(4), 609.

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management*, *34*(1), 89–126.

Watkins, L. M., & Johnston, L. (2000). Screening job applicants: The impact of physical attractiveness and application quality. *International Journal of Selection and Assessment*, *8*(2), 76–84.

Wiederman, M. W., & Allgeier, E. R. (1992). Gender differences in mate selection criteria: Sociobiological or socioeconomic explanation? *Ethology and Sociobiology*, *13*(2), 115–124.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*(3), 395.

Williams, J. E., & Best, D. L. (1990). Measuring sex stereotypes: A multination study (revised edition). *Cross-Cultural Research and Methodology Series*, *6*.

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, *142*(2), 165.

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

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology*, *16*(1), 155–188.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263.

Wojciszke, B., & Klusek, B. (1996). Moral and competence-related traits in political perception. *Polish Psychological Bulletin*.

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In *Advances in experimental social psychology* (Vol. 46, pp. 55–123). Elsevier.

Ybarra, O., Chan, E., Park, H., Burnstein, E., Monin, B., & Stanik, C. (2008). Life's recurring challenges and the fundamental dimensions: An integration and its implications for cultural differences and similarities. *European Journal of Social Psychology*, *38*(7), 1083–1092.

Yoon, Y. (2018). The big flaw in economics is the absence of morality. *Financial Times* Retrieved from: https://www.ft.com/content/bfc77aec-fd11-11e7-9b32-d7d59aace167

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