ORIGINAL ARTICLE

Men Should Be Competent, Women Should Have it All: Multiple Criteria in the Evaluation of Female Job Candidates

Silvia Moscatelli[1] • Michela Menegatti[1] • Naomi Ellemers[2] • Marco Giovanni Mariani[1] • Monica Rubini[1]

Abstract

The present research investigated whether evaluations of female and male job candidates rely on different dimensions. Going beyond previous studies on the role of gender stereotypes, we examined the relative importance of competence, morality, and sociability in employment decisions. In Study 1, we content-analyzed 68 archival reports of professionals to explore the extent to which they spontaneously referred to the three dimensions in evaluations of women and men. In Study 2, 259 Italian student participants rated the importance of different traits in hiring a female or male candidate for a job position. Additionally, we tested the relative influence of specific information about candidate competence and morality in predicting hiring (Study 3; $n = 123$ Italian students) and contract renewal (Study 4; $n = 108$ Italian students) decisions. Findings consistently showed that competence was the most important dimension in evaluations and decisions concerning male candidates, whereas all dimensions were important for female candidates. Moreover, decisions concerning women were influenced by the dimension on which they appeared to be relatively weak. Overall, findings suggest that women are evaluated against multiple criteria and might therefore be asked to meet more requirements than men to be selected and make a career. These findings can help evaluators and decision makers adopt assessment strategies that prevent more critical evaluations of women, such as establishing specific evaluation criteria before the disclosure of a candidate's gender.

Keywords Morality · Competence · Sex discrimination · Employment discrimination · Diversity in the workplace

The last decades have seen substantial progress toward gender equality in education and employment, at least in Western countries (European Commission 2018; Geiger and Parker 2018). Nevertheless, the gender gap in workforce participation and pay is still considerable. In 2017, the employment rate of women in Europe was 66.6%, whereas it was 78.1% for men (European Commission 2018). The gender gap in employment was even higher in some countries; for instance, it was about 18% in Italy (Fulton and Sechi 2018). Moreover, access to high status positions for women remains unequal in many occupations and countries, and women are largely under-represented in

public undertakings and corporate boards of large public companies (European Commission 2018).

The aforementioned disparities persist in spite of the wide recognition of the advantages of gender diversity in teams and organizations. In fact, women are thought to provide different qualities and skills that enhance collective performance (Ellemers and Rink 2016), and a higher female representation in leadership positions is associated with companies' better financial performance and higher innovation (Deszö and Ross 2012). Thus, a major issue for social scientists is understanding why women are still so underrepresented in the labor force.

Most studies have explained the difficulties women encounter when entering the workplace and making careers in organizations in terms of gender stereotypes (Davison and Burke 2000; Ryan et al. 2016) and "lack-of-fit" between the characteristics that are typically ascribed to women and many workplace positions (Heilman 2012). In fact, stereotypes depict women as high in warmth but not so well-equipped in competence, which is attributed to men to a higher extent. Competence is, however, the most required characteristic,

✉ Silvia Moscatelli
  silvia.moscatelli@unibo.it

1   Department of Psychology, University of Bologna, viale Berti Pichat 5, 40127 Bologna, Italy

2   Department of Psychology, Utrecht University, Utrecht, The Netherlands

especially in high-level positions which are perceived to be male gender-typed (Heilman et al. 2015).

The present research aimed to explain gender bias in employment evaluations and decisions by adopting a different approach. Toward this aim, we built on a recent model of social judgment, according to which impressions of others are basically organized around three (rather than formerly two) dimensions: competence, morality, and sociability (Brambilla et al. 2012; Leach et al. 2007). In four studies, we examined whether competence, morality, and sociability have a different importance in evaluations and decisions concerning female and male candidates, even for positions that are considered equally suitable for women and men. Specifically, we proposed that competence is the key dimension in evaluations of male candidates whereas multiple criteria are important in the evaluations of female candidates.

Gender Bias in Personnel Evaluation and Selection

A wide corpus of research has documented a variety of subtle, yet very real, barriers that affect women's chances of being hired and reaching high status positions (Barreto et al. 2009; Bruckmüller et al. 2013; Ellemers 2018; Menegatti et al. 2017; Menegatti and Rubini 2017; Rubini and Menegatti 2014). As we mentioned, most research has explained this issue by focusing on the key role of gender stereotypes, which depict men as higher in agency and competence, whereas women are seen as warmer and therefore more suitable for caring activities or domestic roles (Glick et al. 1988). Accordingly, there is a perceived lack of fit between the requirements of specific organizational positions—which focus on competence-related characteristics such as assertion, competitiveness, and ambitiousness—and the characteristics attributed to women, who therefore have fewer chances of being hired than men (Heilman 2012). Notably, men tend to be considered as more competent and hirable than women even when men and women display similar performances or qualifications (Moss-Racusin et al. 2012). The impact of stereotypic expectations on selection decisions is even amplified when information about the job candidates is incomplete, ambiguous or limited and rules or criteria for evaluation are not clear (Heilman et al. 2015).

Gender stereotypes also lead people to use different standards in judging male and female job candidates. As proposed by the *shifting standards* model (Biernat and Fuegen 2001), based on stereotypic expectations that women would not be as competent as men in employment settings, individuals initially tend to base judgments on within-gender standards, setting lower minimum standards of performance for women (e.g., "She is very skilled *for a woman*"). However, precisely because of these lower standards, women are required to provide

even more evidence of competence at later stages of decision-making to achieve the same outcome as male candidates (Levin et al. 2005).

Besides describing what women and men are like, stereotypes also designate how women and men should be. This prescriptive aspect of gender stereotypes (Heilman 2012) gives rise to normative expectations for women's and men's behavior and leads to social disapproval and penalties against women who violate gender norms. In particular, women who behave counter-stereotypically by presenting themselves as ambitious, competitive, and capable can be the target of backlash (Rudman and Glick 2001) and are likely to be viewed as weak in social skills. In fact, when evaluating agentic women, their presumed lack of social skills becomes more important than evidence of competence and can negatively impact their selection and career advancement (Heilman et al. 2004; Phelan et al. 2008). In any case, even when highly qualified women are evaluated as being as competent as men, they are often liked less and are less likely to be hired (Rudman et al. 2012). Although there are penalties even for men who violate male stereotypes and succeed in female gender-typed jobs (Heilman et al. 2015), men are still relatively advantaged in those jobs (for instance, they receive greater organizational rewards and are promoted more quickly) compared to women who have similar qualifications (Budig 2002; Williams 1992).

Thus, overall research has highlighted that women are disadvantaged in employment decisions because they are evaluated as relatively low in competence. However, even when women are highly qualified, they are penalized for being weak on typically "feminine" attributes. This evidence suggests that more dimensions beyond just competence might be important when evaluating women.

Competence, Morality, and Sociability in Evaluations

Research on gender bias has relied on the distinction between competence and warmth as key components of gender stereotypes. However, recent theorization has claimed that warmth includes two different dimensions: *sociability*, that is, the ability or intention to have good relationships with others, and *morality*, conceived as perceived correctness of social behavior, honesty, and trustworthiness (Leach et al. 2007). Moreover, Abele et al. (2016) have argued that in most studies, warmth has been mainly conceptualized in terms of being friendly, sociable, and benevolent to others in ways that allow the maintenance of affectionate social relationships with them, thus being equated to sociability.

Consistent evidence has been collected on the unique contributions of morality and sociability to social judgment, pointing out that morality is primary in forming impressions of individuals and groups (Brambilla and Leach 2014;

Ellemers 2017; Moscatelli et al. 2019; Pagliaro et al. 2016). Moral information is quickly inferred from faces (Willis and Todorov 2006), dominates the overall impression people form (Brambilla et al. 2012; Goodwin et al. 2014), and even affects their capacity to act in coordination with unknown others (Brambilla et al. 2016; Menegatti et al. 2019). Moreover, individuals show more prosocial behaviors toward moral others (Prati et al. 2018) and are more willing to cooperate with moral than with competent leaders (Pagliaro et al. 2013).

Despite the primacy of morality in impression formation, there is little evidence of its role in employment decision making. In the work context, competence is obviously primary. Indeed, Brambilla et al. (2011) found that student participants, in a trait-selection task, preferred to gain information about a target's competence (compared to morality or sociability) when their goal was selecting the most suitable candidate for a position. However, identifying candidates' moral character in organization is important because it predicts employees' harmful and helpful behaviors (Cohen et al. 2014). Nonetheless, the examination of morality in personnel selection practices has been mostly relegated to "integrity tests," which some companies use in the finale stage of decision making after candidates have been found suitable in terms of their competence (Sackett and Schmitt 2012).

So far, few studies have examined the role of morality in gender stereotypes and gender bias. As we mentioned, research has mostly focused on stereotypic attributions of competence and sociability to men and women, or it has at most considered differences between women and men in justice-oriented and care-oriented moral reasoning or in empathy (Baez et al. 2017; Gilligan 1982). However, the question about whether women are attributed higher morality than men, in terms of honesty or trustworthiness, remains almost unexplored (Goodwin et al. 2014). In this respect, Leach et al. (2017) reported that women endorse the stereotype of women as more trustworthy than men to a lower extent than they endorse the stereotype of women as higher in sociability. Sheppard and Johnson (2019) showed that attractive female leaders are perceived as less trustworthy than their less attractive counterparts; in turn, perceived untruthfulness explains why attractive female leaders are penalized in both feminine and masculine work contexts. Other studies have focused on actual moral behavior of women and men, providing mixed evidence. On the one hand, women tend to give more weight to others' interests in their decisions, score higher on personality traits such as sincerity and fairness, and show higher vocational interests in taking care of others (Franke et al. 1997; Lippa 1998; see also Crocetti et al. 2019). On the other hand, women appear to behave more ethically in groups, but not in individual decision making (Muehlheusser et al. 2015).

Of interest for the purposes of the present research, Prati et al. (2019) recently analyzed the extent to which managers in an Italian organization used competence-, morality- and sociability-related terms in written performance appraisals of employees. Their findings revealed that male employees were mainly described with competence-related terms, whereas appraisals of female employees were more likely to refer to morality and sociability qualities in addition to their competence. In other terms, managers referred to more dimensions in describing performance and qualities of female than male employees. (For similar results in the domain of face-perception and managers' evaluation, see Fruhen et al. 2015.) The present study therefore underlines the importance of fully understanding whether multiple dimensions are considered important and influence employment decisions when evaluations concern female in contrast to male candidates.

Overview of the Present Research

The key goal of the current research was to investigate the relative importance of competence, morality, and sociability—the main judgment dimensions according to a recent model (Brambilla and Leach 2014)—in evaluation of and employment decisions on male and female candidates. To this end, we conducted four studies. First, we content-analyzed archival reports of real-life hiring evaluations in order to examine the extent to which professional selectors spontaneously referred to the three dimensions in describing female and male candidates and how this related to the hiring or rejection decision (Study 1). We then investigated the perceived importance of competence, morality, and sociability in hiring female and male candidates for a job position (Study 2). Finally, we tested the relative influence of individuating information about candidate competence and morality in predicting hiring (Study 3) and contract renewal (Study 4) decisions. In all the experimental studies (Studies 2–4) we focused on jobs that are not considered as typically masculine or feminine. Thus, we aimed to examine the relative importance of different evaluation dimensions while going beyond the mere effect of the perceived lack of fit between type of job and candidate gender (Heilman 2012).

In general terms, we expected that competence would be the most important dimension in employment decision making (Brambilla et al. 2011). However, because women are stereotypically depicted as high in warmth (Heilman 2012) and morality is more important in performance appraisals of women than men (Prati et al. 2019), we advanced that morality and sociability would be more frequently mentioned in justifying employment decisions involving female candidates (Study 1) and would be considered relatively more important for female than for male candidates (Study 2). As a result, and in line with studies on the backlash effect (Phelan et al. 2008; see also Prati et al. 2019), we expected that when specific information on multiple dimensions were provided, evaluators would base their decision concerning female candidates on all

the judgment dimensions. Conversely, we predicted that decisions about male candidates would be influenced mainly by their competence (Studies 3–4).

In regard to possible differences due to the evaluators' gender, previous evidence is mixed. Some studies have suggested that female evaluators may be even more critical of female than male candidates (Faniko et al. 2017), whereas others have reported some ingroup bias by female evaluators (Gorman 2005). Thus, we did not advance specific hypotheses on this aspect. All studies were conducted in Italian.

Study 1

In Study 1 we content-analyzed the reports written by a hiring committee of professional selectors who had evaluated candidates for vacant positions in the customers service team of a large Italian banking organization. We should note that the organization we considered here has adopted an internal code against gender discrimination and a series of procedures to verify its implementation, in line with Italian Law 125/91 and Decree Law 187/03. Accordingly, there were no differences in the actual number of women and men selected or rejected for the job. We were, however, interested in possible differences in how hiring and rejection decisions on male and female candidates were motivated.

To this aim, we examined how often selectors spontaneously employed positively or negatively connoted terms relating to competence, morality, and sociability in motivating hiring and rejection decisions concerning male and female candidates (see Prati et al. 2019, for a similar procedure). Given the specific context and our goal, we expected that professional selectors would mainly refer to competence. However, we expected that selectors would use relatively more competence-related terms in reports of male candidates and would rely relatively more on social abilities and morality-related qualities in evaluating women.

Method

Units of Analysis

The units of analysis were the evaluation reports of 68 candidates (39 for women) who had applied for a position in the customer service team, with responsibility on credit management, in an Italian bank group in 2010. Reports were made available by the organization for research purposes after removal of committees' and employees' personal information. Thirty-one reports referred to candidates who had been hired (18 women and 13 men) and 37 referred to candidates who had been rejected (21 women and 16 men). Every candidate had been evaluated individually by a team of three professional selectors of an external HR counseling firm. No additional

information about the three selectors (e.g., gender) was available. In accordance with the Italian privacy law (Law Decree DL-196/2003), candidates signed a statement consenting to the processing of data.

The selection procedure consisted of a pre-evaluation of the fit between the resumé and the required profile, a cognitive test, a group test on relational abilities, and an individual interview. At the end of the procedure, selectors reported their collective judgment about each candidate. These final reports, which had been submitted on an electronic form, are kept in the company files.

Dependent Variables

Two independent judges (a man and a woman) coded the number of terms (nouns, adjectives or verbs) used by selectors to refer to candidates' competence, sociability, and morality. They also coded terms' positive or negative valence (for a similar coding see, Madera et al. 2009). Judges were unaware of our research purposes and hypotheses. They were aware, however, of candidates' gender because Italian nouns are often either feminine or masculine, and the related words (such as determiners, pronouns, and adjectives) change their form accordingly. Examples of coded terms are: efficacy, pragmatic, skilled (competence: positive); incompetent, disorganized, unskilled (competence: negative); open, friendly, sociable (sociability: positive); introverted, rigid, closed (sociability: negative); respectful, correct, reliable (morality: positive); disrespectful, unreliable, unfair (morality: negative). The agreement between the coders was high (showing convergence for 86% of coded terms). Disagreements were solved by discussion.

Results

Table 1 shows the untransformed mean frequencies of positive and negative terms referring to competence, sociability, and morality in reports of female and male candidates as a function of selection outcomes. The proportions of positive and negative terms related to the three dimensions were computed and subjected to arcsin transformation to normalize the data and remove the intercell dependencies (see Table 1). The arcsin-transformed proportions of positive and negative terms were submitted to separate 2 (candidate gender) × 2 (selection outcome: hired, rejected) × 3 (evaluative dimension: competence, sociability, morality) ANOVAs with repeated measures on the last variable.

Positive Terms

The analysis on the proportions of positive terms showed a main effect of the evaluative dimension, $F(2, 63) = 364.13$,

Table 1 Mean frequencies of positive and negative terms related to competence, morality, and sociability as a function of selection outcome and candidate gender, study 1

	Hired Candidate		Rejected Candidates	
	Women	Men	Women	Men
Evaluations	M (SD) [M (SD)]	M (SD) [M (SD)]	M (SD) [M (SD)]	M (SD) [M (SD)]
(a) Frequencies of Positive Terms				
Competence	13.22 (6.22) [.48$_a$ (.20)]	18.46 (4.31) [.90$_b$ (.17)]	4.24 (2.78) [1.03$_a$ (.53)]	5.63 (3.10) [.98$_a$ (.05)]
Sociability	11.50 (4.20) [.43$_a$ (.14)]	4.15 (2.04) [.16$_c$ (.07)]	1.24 (1.30) [.27$_b$ (.38)]	1.88 (2.09) [.22$_b$ (.26)]
Morality	3.67 (1.82) [.14$_b$ (.09)]	1.69 (1.25) [.07$_c$ (.05)]	.10 (.30) [.01$_c$ (.05)]	.00 (.00) [.00$_c$ (.00)]
(b) Frequencies of Negative Terms				
Competence	2.00 (1.68) [.97$_a$ (.58)]	2.38 (2.66) [.65$_a$ (.61)]	3.29 (2.35) [.30$_a$ (.20)]	6.50 (2.85) [.87$_b$ (.37)]
Sociability	1.06 (1.21) [.25$_b$ (.29)]	1.23 (1.48) [.40$_{ab}$ (.57)]	4.57 (2.98) [.43$_a$ (.21)]	1.31 (3.01) [.13$_c$ (.17)]
Morality	.00 (.00) [.00$_{bc}$ (.00)]	.23 (.83) [.05$_c$ (.18)]	3.05 (2.31) [.31$_a$ (.23)]	1.25 (.93) [.16$_c$ (.12)]

The left entries for each cell are the descriptive statistics for the raw, untransformed frequencies; the bracketed entries are the descriptive statistics for the arcsin-transformed proportions. Means with different subscripts differ significantly within the row and the column for each of the two panels (*p*s < .05)

$p < .001$, $\eta_p^2 = .920$. As hypothesized, competence-related terms were overall more frequent ($M = .85$, $SD = .46$) than sociability-related terms ($M = .28$, $SD = .27$; $p < .001$, $d = .08$), which in turn were used more frequently than morality-related terms ($M = .05$, $SD = .08$; $p < .001$, $d = 1.58$). The analysis showed a significant dimension × selection outcome interaction, $F(2, 63) = 77.82$, $p < .001$, $\eta_p^2 = .361$, which was qualified by the significant three-way interaction, $F(2, 63) = 3.27$, $p = .045$, $\eta_p^2 = .094$. To test our hypotheses, pairwise comparisons (based on Bonferroni test) were conducted to break down the three-way interaction. Results are presented considering hired and rejected candidates separately.

With respect to hired candidates, women were described with a lower proportion of positive morality- than competence- ($p < .001$, $d = 1.28$) or sociability-related terms ($p < .001$, $d = 1.87$) (see Table 1a). The proportions of competence- and sociability-related terms did not significantly differ ($p = .506$). Male candidates were described using a greater proportion of positive competence- than sociability- ($p < .001$, $d = 3.16$) or morality-related ($p < .001$, $d = 3.92$) terms, whereas the proportions of positive morality- and sociability-related terms did not differ ($p = .538$). Comparisons between reports of female and male hired candidates revealed that women were described with a smaller proportion of competence-related terms ($p < .001$, $d = 2.26$) and with larger proportions of sociability- ($p < .001$, $d = 2.44$) and morality-related ($p = .014$, $d = .96$) terms.

Although the use of positive terms is less informative with respect to rejected candidates, results showed that women were described with larger proportions of positive terms related to competence than to sociability ($p < .001$, $d = .86$) or morality ($p < .001$, $d = 1.86$) (see Table 1a). Sociability-related terms were more represented than morality-related terms ($p = .004$, $d = .68$). In a similar vein, men were described with

larger proportions of positive terms related to competence than to sociability ($p = .002$, $d = 2.65$) or morality ($p < .001$, $d = 19.10$), and with a larger proportion of sociability- than morality-related terms ($p = .37$, $d = .81$) There were no differences in the proportions of positive terms referring to competence ($p = .776$), sociability ($p = .671$), or morality ($p = .217$) used in reports of rejected female and male candidates.

Negative Terms

The ANOVA on the proportions of negative terms showed a main effect of dimension, $F(2, 63) = 51.23$, $p < .001$, $\eta_p^2 = .619$. Competence-related terms were more frequently used ($M = .68$, $SD = .52$) than sociability- ($M = .31$, $SD = .34$; $p <. 001$, $d = .45$) and morality-related terms ($M = .14$, $SD = .20$; $p <. 001$, $d = .84$), which in turn were less frequent than sociability-related terms ($p <. 001$, $d = .48$). There was also a significant dimension × selection outcome interaction, $F(1, 63) = 10.63$, $p < .001$, $\eta_p^2 = .252$, which was qualified by the significant three-way interaction, $F(2, 63) = 9.48$, $p < .001$, $\eta_p^2 = .231$. Pairwise comparisons were conducted to break down the three-way interaction. Hired and rejected candidates were considered separately.

With respect to hired candidates, women were described with a larger proportion of negative competence- than sociability- ($p =. 007$, $d = .87$) or morality-related terms ($p <. 001$, $d = 1.67$) (see Table 1b). Proportions of morality- and sociability-related terms did not differ ($p = .078$). Men were described with a larger proportion of terms that referred to competence than to morality ($p = .004$, $d = .92$), which in turn were less frequent than terms referring to sociability ($p = .029$, $d = .55$). The proportions of competence and sociability terms did not significantly differ ($p = .630$). When comparing reports of female and male rejected candidates, no differences emerged in the proportions of negative terms referring to

competence ($p = .139$), sociability ($p = .348$), or morality ($p = .246$).

The results concerning negative terms are more informative to understand rejection decisions. As for rejected female candidates, proportions of competence-related terms did not differ from those of sociability ($p = .474$) or morality-related terms ($p = .878$) (see Table 1b). Proportions of sociability- and morality-related terms did not differ ($p = .271$). In contrast, rejected male candidates were described with a larger proportion of negative terms referring to competence than to sociability ($p < .001$, $d = 1.45$) or morality ($p < .001$, $d = 1.54$). The proportions of sociability- and morality-related terms did not significantly differ ($p = .655$). Comparisons between reports of female and male rejected candidates showed that women were described with a smaller proportion of competence-related terms than men were ($p < .001$, $d = 1.92$). Conversely, women were described with larger proportions of sociability- ($p < .001$, $d = 1.57$) and morality-related terms ($p = .020$, $d = .82$) compared to men.

Discussion

As we expected, the analysis of hiring reports showed that competence—framed in either positive or negative terms—was the primary dimension that professional selectors used to motivate employment decisions. However, there were interesting differences in reports concerning female and male candidates. Hiring decisions concerning male candidates referred primarily to their competence-related qualities, and rejection decisions were primarily justified on the basis of their negative competence-related attributes. Conversely, hiring decisions concerning female candidates were more likely to be motivated through the use of positive terms referred to their sociability and morality in addition to their competence. Moreover, rejection of female candidates was mainly justified by referring to their perceived weakness in morality and sociability. These patterns are in line with those reported by Prati et al. (2019) with respect to performance appraisals. Furthermore, they provide support for our hypothesis that employment decisions concerning women are justified by relying on multiple evaluation criteria, whereas decisions concerning men appear to be more focused on their strengths or weaknesses in competence.

Study 2

In Study 2 participants were explicitly requested to rate the perceived importance of competence, morality, and sociability in selecting a female or a male candidate. Overall, we expected competence to be judged the most important dimension. However, given that female candidates have to conform to the prescriptive stereotype of women as high in warmth (Heilman

2012), we expected that sociability would be judged as more important for female than for male candidates. Based on previous evidence that evaluators refer to morality to a higher extent in performance appraisals of women than men (Prati et al. 2019), and in line with Study 1, we also expected that morality would be considered more important to select women. Conversely, competence should be the primary criterion in selecting male candidates.

In Study 2, we also explored the relations among the perceived importance of the three evaluative dimensions and participants' endorsement of hostile and benevolent sexism as measured with the Ambivalent Sexism Inventory (Glick and Fiske 1996). Even though hostile and benevolent sexisms derive from the same ideal of women as conforming to traditional gender roles, hostile sexism expresses more negative attitudes toward women who do not conform to these roles (e.g., women who make careers). Benevolent sexism is instead more positive in tone and includes the reverence of women in traditional roles of wives, mothers, and child caretakers as well as emphasizes their sexual purity and their need of protection (Glick and Fiske 1996). Individuals generally perceive ambivalent sexism as communicating a view of women as less competent than men. Benevolent sexism, in particular, is seen to portray a view of women as warmer than men (Ramos et al. 2018). In general terms, one might therefore expect that both forms of sexism would be positively associated to the attribution of higher importance to the masculine-stereotyped dimension of competence in selecting male candidates. Basing on Ramos et al. (2018), it also seems plausible that benevolent sexism would be positively associated with the perceived importance of the stereotypically feminine dimension of sociability in selecting women. Basing on findings of Prati et al. (2019), benevolent sexism should also be positively associated with the perceived importance of morality for women.

Method

Pre-Tests

Study 2 tested the importance of traits related to the three dimensions of social judgment in selecting a female or a male candidate for a vacant position in the Teaching Board of participants' academic department. A pre-test was run to assure that this position was considered suitable for both women and men. Thirty Italian university students were given the following instructions:

> There is a vacant position for a student in the Teaching Board of the Department, which is composed by academic staff and students. The main functions of the Teaching Board include monitoring students' career and helping them successfully complete the exams to graduate in due course.

Students were then asked two separate items: "Please rate the extent to which this position is suitable for a woman [for a man]." The response scale ranged from 1 (*not at all*) to 7 (*very much*). Findings supported that this position did not differ in perceived suitability for women ($M = 4.90$, $SD = 1.12$) and men ($M = 4.83$, $SD = 1.18$), $t(29) = .63$, $p = .536$.

A further pre-test was run to select the traits to be listed in the rating task. Twenty-eight Italian university students rated a pool of 15 traits for their competence-, morality-, and sociability- relatedness (for a similar procedure, see Moscatelli et al. 2019). The instructions for this task were: "Please rate the extent to which each of the following characteristic is related to: (a) competence, (b) morality, (c) sociability (1 = *not at all*; 7 = *very much*)". The order of presentation of traits was randomized among respondents. For competence, we selected three traits (competent, capable, efficient) that scored higher on the competence-relatedness scale ($M = 6.48$, $SD = 1.18$) than on the other scales ($M_{morality} = 2.26$, $SD = 1.23$; $M_{sociability} = 2.51$, $SD = 1.44$; $ts > 12.16$, $ps < .001$, $ds > 2.30$). The three traits selected for morality (honest, trustworthy, sincere) scored higher on the morality-relatedness scale ($M = 6.33$, $SD = .65$) than on the other scales ($M_{competence} = 3.69$, $SD = 1.61$; $M_{sociability} = 4.73$, $SD = 1.49$; $ts > 5.62$, $p < .001$, $ds > 1.06$). The traits selected for sociability (friendly, sociable, kind) scored higher on the sociability-relatedness scale ($M = 6.36$, $SD = .50$) than on the other scales ($M_{morality} = 3.35$, $SD = 1.12$; $M_{competence} = 3.01$, $SD = 1.26$; $ts < 13.82$, $ps < .001$, $ds > 2.60$).

Participants and Design

First-year undergraduate students enrolled at a large Italian university were asked to fill in a paper-and-pencil questionnaire at the end of a class and on a voluntary basis. This recruitment resulted in a convenient sample of 259 participants (142, 54.4% women; $M_{age} = 22.56$, $SD = 3.76$, range = 19–51 years-old). This number meets the criterion of 30 participants per cell in sample size determination (Wilson Van Voourhis and Morgan 2007), considering a 2 (candidate gender) × 2 (evaluator gender) × 3 (dimension; within-participants) design. A sensitivity analysis conducted with G*Power (Faul et al. 2007) showed that our sample was sufficient to detect small effects of $f^2 = .04$ (equivalent to $\eta_p^2 = .04$) assuming an alpha of .05 and power of .95 for a mixed-model ANOVA.

Procedure and Measures

The study was approved by the Bioethical Committee of the first author's university in November 2013. Before filling in the questionnaire, participants provided informed consent on a separate form. Participants were presented with a questionnaire using the same instructions as

those used in the pre-test. In the female candidate condition, the instructions added:

> As you presumably know, in order to promote equal opportunities, the Teaching Board is composed by equal numbers of women and men. The actual vacant position on the Board is for a woman. Please imagine that you have to select a female student for the vacant position in the Board.

In the male candidate condition, the instructions were identical except that they referred to a vacant position for a man, and participants were asked to imagine that they had to select a male student.

Perceived Importance of Competence, Morality, and Sociability The rating task was introduced by the following instructions

> You will be now presented with a list of characteristics that you might consider in selecting a female student [a male student] to the vacant position in the Teaching Board. Please rate the extent to which the following characteristics would be important for your decision.

The response scale ranged from 1 (*not at all*) to 7 (*very much*). Participants were then presented, in random order, with the nine traits (three for each dimension) that had been selected in the pre-test. For each dimension, the ratings of the three traits were averaged to obtain measures of the importance of competence ($\alpha = .71$), morality ($\alpha = .71$), and sociability ($\alpha = .71$). Alphas were similar to those ranging from of .70 to .82 recently reported by Moscatelli et al. (2019) in group evaluations provided by a sample of Italian students. Correlations among the measures ranged from .16 to .42 ($ps < .008$), supporting that these measured referred to related but distinct dimensions (see Moscatelli et al. 2019, for similar findings).

Ambivalent Sexism Inventory In order to assess participants' endorsement of sexist views about women, we used the Italian validated version of the Ambivalent Sexism Inventory (ASI; Manganelli Rattazzi et al. 2008), consisting of two 11-item self-report subscales that tap into benevolent (sample item: "Women should be cherished and protected by men") and hostile (sample item: "Most women interpret innocent remarks or acts as being sexist") sexism. Participants responded using a 6-point scale from 1 (*strongly disagree*) to 6 (*strongly agree*). Some items were removed because of low scale reliability (for a similar procedure, see Zaikman and Marks 2014). After reverse scoring appropriate items, subscale items were averaged together to create a benevolent sexism score (6 items, $\alpha = .84$) and a hostile sexism score (6 items, $\alpha = .84$). Higher scores indicate more sexism. Alphas were similar to

those of .80 and .87 reported by Manganelli Rattazzi et al. (2008) from an Italian sample of students (see also Zaikman and Marks 2014). Research has found that the ASI is a valid measure across cultures and reliably measures valid constructs (Glick and Fiske 1996; Glick et al. 2000).

Manipulation Checks Participants were asked, using two separate items, "Please rate the extent to which this position is suitable for a woman [a man]," with each rated from 1 (*not at all*) to 7 (*very much*). In order to assure that they thought to a female/a male candidate as required by instructions, participants were also asked: "Please indicate if you have been asked to select (a) a female student or (b) a male student."

Results

Manipulation Checks

All participants correctly reported the gender of the candidate they had been required to consider. A 2 (evaluator gender) × 2 (candidate gender) × 2 (gender-suitability of the job; within-participants) ANOVA confirmed that the job was perceived as equally suitable for women ($M = 5.07$, $SD = 1.14$) and men ($M = 5.07$, $SD = 1.07$), $F(1, 257) = .01$, $p = .955$. No other effects were significant (Fs < 2.91, ps > .090).

Perceived Importance of Competence, Morality, and Sociability

A 2 (candidate gender) × 2 (evaluator gender) × 3 (dimension; within-participants) ANOVA on the perceived importance of competence, morality, and sociability revealed a main effect of dimension, $F(2, 254) = 163.19$, $p < .001$, $\eta_p^2 = .562$. As predicted, competence traits were considered more important ($M = 6.21$, $SD = .87$) than morality ($M = 5.72$, $SD = 1.19$; $p < .001$, $d = .37$) or sociability traits ($M = 4.68$, $SD = 1.22$; $p < .001$, $d = 1.11$). Morality traits were rated as more important than sociability traits ($p < .001$, $d = .67$).

The analysis also revealed a main effect of candidate gender, $F(1, 255) = 24.35$, $p < .001$, $\eta_p^2 = .082$, whereby scores were overall higher for female ($M = 5.74$, $SD = .71$) than male ($M = 5.26$, $SD = .82$) candidates. This effect was qualified by a candidate gender × dimension interaction, $F(2, 254) = 3.50$, $p = .011$, $\eta_p^2 = .035$. Evaluators considered competence more important in hiring decisions concerning female ($M = 6.33$, $SD = .87$) than male ($M = 6.04$, $SD = .86$) candidates ($p = .018$, $d = .32$). They also considered morality more important for women ($M = 6.04$, $SD = .99$) than for men ($M = 5.29$, $SD = 1.30$; $p = .005$, $d = .65$). Similarly, sociability was deemed more important for women ($M = 4.86$, $SD = 1.18$) than for men ($M = 4.43$, $SD = 1.24$; $p = .018$, $d = .36$). Although competence was perceived to be more important than morality for both female candidates ($p = .018$) and male

candidates ($p < .001$), the size of the difference was smaller for female ($d = .24$) than for male candidates ($d = .52$). For female candidates, sociability was considered less important than competence ($p = .018$, $d = 1.07$) or morality ($p < .001$, $d = .95$). A similar pattern was found for male candidates, with ratings of sociability being lower than ratings of competence ($p < .001$, $d = 1.15$) or morality ($p < .001$, $d = .63$).

There was also an unexpected significant evaluator gender × dimension interaction, $F(2, 254) = 4.63$, $p = .011$, $\eta_p^2 = .035$. Male evaluators considered competence ($M = 6.01$, $SD = .94$) and morality ($M = 5.86$, $SD = 1.09$) as more important than sociability ($M = 4.83$, $SD = 1.21$; $p < .001$, $d = .82$); they did not differentiate between competence and morality ($p = .114$). Female evaluators rated competence as more important ($M = 6.37$, $SD = .78$) than both morality ($M = 5.61$, $SD = 1.25$; $p < .001$, $d = .58$) and sociability ($M = 4.55$, $SD = 1.22$; $p < .001$, $d = 1.45$), and rated morality as more important than sociability ($p < .001$, $d = .85$). Moreover, female evaluators considered competence as more important ($p = .002$, $d = .42$), and morality less important ($p = .049$, $d = .21$), compared to male evaluators. Female evaluators also considered sociability as less important than male evaluators did ($p = .040$, $d = .23$). No other effects were significant (Fs < 3.03, ps >. 083).

Endorsement of Sexist Views

A 2 (evaluator gender) × 2 (candidate gender) × 2 (type of sexism: hostile, benevolent; within-participants) ANOVA indicated less endorsement of hostile sexism ($M = 3.00$, $SD = 1.08$) than benevolent sexism ($M = 3.23$, $SD = 1.17$), $F(1, 255) = 9.65$, $p = .002$, $\eta_p^2 = .036$. Male evaluators, on average, were inclined to endorse more sexist views ($M = 3.49$, $SD = .82$) than female evaluators ($M = 2.81$, $SD = .90$), $F(1, 255) = 38.11$, $p < .001$, $\eta_p^2 = .130$. There was also a main effect of candidate gender, $F(1, 255) = 7.48$, $p = .007$, $\eta_p^2 = .028$, which revealed higher agreement with sexist views when the candidate was a woman ($M = 3.23$, $SD = .94$) rather than a man ($M = 2.97$, $SD = .89$). No interactions were significant (Fs < 2.98, ps > .087).

Correlation Analyses

Correlation analyses revealed no significant correlations between either hostile or benevolent sexism with the perceived importance of different dimensions in evaluating male (rs < .11, ps > .240) or female candidates (rs < .11, ps > .190).

Discussion

Findings showed that competence was the primary dimension used to evaluate job candidates and that more importance was assigned to morality than sociability traits. Although we deliberately considered a job that was in principle (and was

perceived as) equally suitable for women and men, the importance of the three judgment dimensions depended on candidates' gender. As in Study 1, in the case of male candidates the evaluative goal set by the job selection context seemed to lead participants to acknowledge more clearly the importance of competence over the other dimensions. More generally, all dimensions—including competence—were considered more important for female than for male candidates, suggesting that decisions concerning women might rely on more criteria than decisions concerning men.

There were also unexpected yet interesting differences in the ratings provided by female and male evaluators, with female evaluators giving relatively more importance to competence than to the other dimensions. The emphasis on the more masculine-stereotyped dimension of competence (Heilman 2012) might reflect women's attempt to adhere to a stereotypic view of job requirements while keeping from behaving as "typical women." This tentative explanation would, however, require more investigation.

With respect to ambivalent sexism, male evaluators endorsed sexist views more strongly than female evaluators did, a finding in line with previous results (Barreto and Ellemers 2005). Moreover, the request to select a female candidate led respondents to agree with sexist views to a higher extent. It is possible that thinking about female candidates rendered stereotypic views of women and gender roles more salient. Future research might investigate this issue.

Participants' sexist attitudes did not relate to the perceived importance of dimensions in selecting women or men. This non-finding might be due to the different nature of the two measures. Indeed, individuals might more easily control their self-reported attitudes on an explicit measure, such as that of ambivalent sexism, for reasons of social desirability and impression management (Barreto and Ellemers 2005; de Oliveira et al. 2015). Rating the importance of different evaluative criteria can instead be intended as a more indirect measure of attitudes toward men and women because participants are unlikely to be aware of the implications of their responses. Moreover, given that in the present study candidate gender was manipulated as a between-participants factor, we can be reasonably confident that respondents neither figured out the aim of the study nor varied their responses accordingly.

In sum, Study 2 highlighted a general tendency to take more dimensions into consideration in evaluations of women, whereas evaluations of men were more clearly centered on competence. Thus, our findings suggest that in personnel selection, female candidates might be held to reach higher standards than men across all dimensions. Accordingly, one might advance that employment decisions concerning female candidates are likely to be influenced by multiple evaluation criteria, whereas decisions about male candidates should be mainly influenced by their level of competence. This hypothesis was tested in Study 3.

Study 3

Study 3 examined selection decisions based on concrete information about male and female candidates. We reasoned that giving specific individuating information about each candidate would, to some extent, counteract the effect of stereotypical expectations concerning women and men (see Bosak and Sczesny 2011; Heilman et al. 2015). To this aim, we focused on the effects of information on competence and morality, which were the two most important criteria of evaluation in Study 2. A further, more theoretical-driven reason for not considering sociability is that studies on gender stereotypes have mainly focused on warmth (generally equating it with sociability; Abele et al. 2016) as a typically feminine characteristic whereas less is known about the role played by women's and men's perceived morality in employment decisions. Moreover, competence is the most important dimension used in personnel evaluations and selection (Brambilla et al. 2011; Prati et al. 2019) whereas morality represents the most important dimension in overall social judgment (Brambilla and Leach 2014).

Italian participants were presented with the qualifications of two male or two female candidates who had supposedly applied for a job. One candidate within each gender had been previously assessed as relatively high in competence and low in morality, whereas the other presented the opposite evaluation profile (i.e., relatively low in competence and high in morality). Participants were asked to rate the extent to which each candidate would behave in a competent and moral manner in his/her work and to indicate his/her likelihood of being selected. In general terms, we expected that candidates with a high competence and low morality profile would be more likely to be hired than candidates with a low competence and high morality profile. However, basing on studies on shifting standards and backlash (Biernat and Fuegen 2001; Phelan et al. 2008), we argued that qualifications being equal, judgments toward female candidates would be less favorable than those toward male candidates. Moreover, given our finding that women are evaluated against multiple criteria, we expected that perceptions of both morality and competence would be related to selection decisions concerning female candidates. Conversely, perceived competence should be the primary predictor of the selection decision for male candidates.

Method

Pre-Tests

In Study 3, student participants from a large Italian university were required to imagine that they were members of the Teaching Board of their Department, which had to select a Master's student for a tutoring job. A pre-test was run to assure that the position was considered suitable for both women and

men. Thirty students enrolled at the same Italian university as participants of Study 3 were given the following instructions:

> The Teaching Board of your Department has to select a Master's student for a tutoring job. The main tasks of the tutor involve assisting undergraduates in their career, supporting them in planning exams and in the preparation of the final paper.

Participants were then asked, using two separate items, "Please rate the extent to which this tutoring job is suitable for a woman [a man]." Findings revealed no differences in the suitability of the position (1 = *not at all*; 7 = *very much*) for women ($M = 5.27, SD = 1.17$) and men ($M = 5.20, SD = 1.25$), $t(29) = .63, p = .536$.

Two further pre-tests were run to assure that candidates' evaluation profiles employed as stimuli in Study 3 were perceived as intended. The profiles, which had apparently been completed by candidates' university supervisors, reported evaluations of candidates along two clusters of traits: one comprising competence-related and the other comprising morality-related traits. Evaluations along the two clusters varied so that candidates presented either a high-competence/low-morality profile or a low-competence/high-morality profile. First, 44 Italian university students were presented with the competence-related traits (capable, efficient, smart, competent) and the morality-related traits (trustworthy, loyal, honest, sincere) and then were asked: "Please rate the extent to which the following traits are related to competence/to morality" (1 = *not at all*; 7 = *very much*). Findings supported that the traits included in the competence cluster were overall perceived as more related to competence ($M = 5.66, SD = 1.16$) than to morality ($M = 2.48, SD = 1.64$), $t(43) = 11.21, p < .001$, whereas those included in the morality cluster were perceived as more related more to morality ($M = 5.50, SD = 1.30$) than to competence ($M = 2.45, SD = 1.25$), $t(43) = 12.20, p < .001$.

Furthermore, 44 Italian university students were presented with the profiles of two candidates who were identified by their initials. Each candidate had apparently been evaluated on the two clusters of traits (composed as explained before), which were presented as "Evaluation domain: 1" and "Evaluation domain: 2." Evaluations of each candidate on the eight traits could range from 1 (*not at all*) to 10 (*very much*). One profile reported relatively higher evaluations on competence-related traits (with ratings ranging from 8 to 9; overall $M = 8.50$) and relatively lower evaluations on morality-related traits (with ratings ranging from 6 to 7; overall $M = 6.50$); the other profile reported opposite ratings on the two dimensions. For each profile, participants to the pre-test were asked: "Please rate the extent to which the candidate is competent [moral]"; the response scale ranged from 1 (*not at all*) to 7 (*very much*). Supporting the manipulation of profiles,

candidates with the high-competence/low-morality profile were attributed higher competence ($M = 5.70, SD = 1.13$) and lower morality ($M = 4.53, SD = .79$) than the candidates with the low-competence/high-morality profile ($M_{competence} = 4.46, SD = .87; M_{morality} = 5.70, SD = .88; ts > 6.03, ps < .001, ds > 1.04$).

Participants and Design

Participants to Study 3 were 123 Italian undergraduate students (78, 62.9% women; $M_{age} = 2.96, SD = 2.36$, range = 19–31 years-old) attending a Psychology course at a large Italian university, who were asked to fill in a paper-and-pencil questionnaire at the end of a class and on a voluntary basis. The resulting sample size was in line with Harris (1985) guideline of $n > 50$ + number of predictor variables for regression sample sizes (see also Green 1991), as well as with Wilson Van Voourhis and Morgan's (2007) guideline of 30 participants per cell, considering a 2 (evaluator gender) × 2 (candidate gender) × 2 (evaluation profile: high-competence/low-morality, low-competence/high-morality; within participants) design. A sensitivity analysis conducted with G*Power (Faul et al. 2007) showed that our sample was sufficient to detect medium-size effects of $f^2 = .17$ (corresponding to $R^2 = .14$), assuming an alpha of .05, and power of .95 for a multiple regression model.

Procedure

Study 3 was approved by the Bioethical Committee of the first author's institution in May 2014. Before being presenting with the questionnaire, participants provided informed consent on a separate form. Then, participants were given the questionnaires, where they read the following instructions:

> Imagine that you are member of the Teach Board of your Department, which is by statute composed by equal numbers of academic staff and students. The Board has to select a Master's student for a tutoring job. The main tasks of the tutor involve assisting undergraduates in their career, supporting them in planning exams and in the preparation of the final paper. You will be now presented with the profiles of two candidates to the tutoring position. Each candidate has already been evaluated from his/her university supervisor on a series of characteristics. For each candidate, please read supervisor's ratings and answer the questions that follow thereafter.

Each participant was then presented with the evaluations of two female or two male candidates along the two clusters of competence- and morality-related traits described in the pre-tests section, which were named "Evaluation domain: 1" and

"Evaluation domain: 2"; these profiles were presented in randomized order across participants. Evaluations on the two clusters of traits varied so that one of the two candidates appeared to have a high-competence/low-morality profile, whereas the other candidate presented a low-competence/high-morality profile (see the pre-tests section for detailed descriptions of the profiles). The order of candidate profiles was randomized across participants. After being presented with each evaluation profile, participants filled in measures of perceived competence, perceived morality, and selection decision. Manipulation checks were presented at the end of the questionnaire.

Perceived Competence For each candidate's profile, participants rated on 7-point scales from 1 (*very unlikely*) to 7 (*very likely*) the extent to which if hired, the candidate would show competence. The items were: "How likely is it that if hired, the candidate will work in an efficient manner"; "How likely is it that if hired, the candidate will be able to make a personal contribution to the project?" ($\alpha_{\text{high-competence/low-morality}} = .76$, $\alpha_{\text{low-competence/high-morality}} = .73$).

Perceived Morality For each profile, participants rated the perceived morality of the candidate on two items: "How likely is it that if hired, the candidate will behave in a correct manner with the people involved in the project?"; "How likely it is that if hired, the candidate will be able to gain the trust of the other people involved in the project?" (1 = *very unlikely*; 7 = *very likely*; $\alpha_{\text{high-competence/low-morality}} = .80$, $\alpha_{\text{low-competence/high-morality}} = .74$).

Selection Decision Selection decision was measured through the following items: "In your view, how likely is it that the Committee would select the candidate?"; "Would you select the candidate?," with each rated on a scale from 1 (*very unlikely*) to 7 (*very likely*) ($\alpha_{\text{high-competence/low-morality}} = .78$; $\alpha_{\text{low-competence/high-morality}} = .85$).

Manipulation Checks Participants rated the suitability of the tutoring job for men and women on two separate items: "Please rate the extent to which this position is suitable for a woman [a man]" (1 = *not at all*; 7 = *very much*). As in Study 2, they were also asked to report the gender of the candidates they had been presented.

Results

Manipulation Checks

All participants correctly indicated the gender of the candidates they had been asked to evaluate. A 2 (evaluator gender) × 2 (candidate gender) × 2 (gender suitability of the job; within-participants) ANOVA showed that the job was

considered equally suitable for men ($M = 5.46$, $SD = 1.04$) and women ($M = 5.54$, $SD = 1.00$), $F(1, 120) = 1.741$, $p = .190$. No other effects were significant (Fs < 2.910, ps > .091).

Perceived Competence

The 2 (evaluator gender) × 2 (candidate gender) × 2 (evaluation profile: high-competence/low-morality, low-competence/high-morality; within-participants) ANOVA on competence ratings showed a main effect of the profile, $F(1, 120) = 15.94$, $p < .001$, $\eta_p^2 = .117$. High-competence/low-morality candidates were rated as more competent than low-competence/high-morality candidates. As predicted, across profiles female candidates were rated as less competent than male candidates, $F(1, 120) = 9.07$, $p = .003$, $\eta_p^2 = .070$, despite the identical profiles (see Table 2a). No other effects were significant (Fs < 1.87, ps > .174).

Perceived Morality

The analysis showed a main effect of the evaluation profile, $F(1, 120) = 280.190$, $p < .001$, $\eta_p^2 = .700$. Low-competence/high-morality candidates were rated higher in morality than high-competence/low-morality candidates (see Table 2a). Again, across profiles female candidates obtained lower scores than male candidates, $F(1, 120) = 6.54$, $p = .012$, $\eta_p^2 = .052$. No other effects were significant (Fs < 2.20, ps > .141).

Selection Decision

The ANOVA on selection decision revealed a significant evaluator gender × candidate gender interaction, $F(1, 120) = 11.33$, $p < .001$, $\eta_p^2 = .086$. Male ($M = 4.65$, $SD = 1.01$) and female ($M = 5.03$, SD = .74) evaluators did not differ with respect to decisions about male candidates ($p = .077$). Male evaluators ($M = 5.09$, $SD = 1.00$), compared to female evaluators ($M = 4.38$, $SD = .74$), indicated a higher likelihood of selection for female candidates ($p < .001$, $d = .81$). As a result, male evaluators indicated a higher likelihood of selection for female versus male candidates ($p = .045$, $d = .44$), whereas female evaluators reported a higher likelihood of selection for male over female candidates ($p < .001$, $d = .88$). No other effects were significant (Fs < 1.22, ps > .272).

Regression Analyses

Table 2a shows correlations among perceived competence, perceived morality, and selection decision. A series of regression analyses were run to test the relative importance of perceived morality and competence as predictors of the selection decision and to assess whether candidates' gender moderated

Table 2 Descriptive statistics and correlations among study variables for female and male candidates, studies 3 and 4

Variables	Female Candidates M (SD)	Male Candidates M (SD)	Total M (SD)	Correlations 1.	2.	3.	4.	5.	6.
(a) Study 3									
High-competence/low-morality profile									
1. Perceived Competence	5.00 (1.20)	5.37 (.88)	5.19 (1.07)	–	.42***	.46***	.54***	.37***	.09
2. Perceived Morality	4.06 (.96)	4.53 (.69)	4.30 (.87)		–	−.02	.54***	.32***	.30***
3. Selection Decision	4.68 (1.13)	4.84 (1.17)	4.76 (1.15)			–	−.09	.12	.05
Low-competence-high morality profile									
4. Perceived Competence	4.45 (1.08)	5.06 (.83)	4.75 (1.01)				–	.56***	.13
5. Perceived Morality	5.48 (1.04)	5.79 (.80)	5.64 (.94)					–	.33***
6. Selection Decision	4.64 (1.24)	4.96 (1.29)	4.80 (1.27)						–
All Candidates									
7. Perceived Competence	4.73 (1.02)	5.22 (.64)	4.97 (0.89)						
8. Perceived Morality	4.77 (0.89)	5.17 (.61)	4.97 (.79)						
9. Selection Decision	4.66 (.92)	4.90 (.90)	4.78 (.91)						
(b) Study 4									
High-competence-low morality profile									
1. Perceived Competence	5.06 (1.22)	4.72 (1.57)	4.91 (1.39)	–	.26**	.28**	−.24**	.02	.07
2. Perceived Morality	3.17 (1.38)	2.67 (1.25)	2.96 (1.34)		–	.27**	.33**	−.15	.26**
3. Renewal Decision	4.14 (1.33)	4.57 (1.06)	4.33 (1.24)			–	−.07	.12	.05
Low-competence/high-morality profile									
4. Perceived Competence	3.95 (1.37)	3.96 (1.48)	3.95 (1.41)				–	.11	.38***
5. Perceived Morality	5.97 (1.07)	6.24 (.92)	6.09 (1.01)					–	.11
6. Renewal Decision	3.64 (1.25)	3.65 (1.42)	3.64 (1.32)						–
All Candidates									
7. Perceived Competence	4.33 (.99)	4.37 (.74)	4.35 (.87)						
8. Perceived Morality	4.40 (.86)	4.64 (.68)	4.52 (.78)						
9. Selection Decision	3.64 (.96)	4.32 (.76)	3.99 (.93)						

Correlations for female candidates are reported above the diagonal of the correlation matrix; for male candidates, below

*$p < .05$. **$p < .01$. ***$p < .001$

these effects. Considering the two profiles separately, perceived competence (centered), perceived morality (centered), candidate gender (0 = man, 1 = woman), the interaction between perceived competence and gender, and the interaction between perceived morality and gender were entered in a linear regression as predictors of selection decision.

With respect to high-competence/low-morality candidates, findings yielded a significant main effect of perceived competence ($\beta = .655$, $p < .001$), but no effects of perceived morality ($\beta = .151$, $p = .294$) or candidates' gender ($\beta = .068$, $p = .420$). Interestingly, the significant perceived competence × gender interaction revealed that candidates' gender moderated the influence of perceived competence on selection decision ($\beta = -.533$, $p < .001$, $R^2 = .258$, $R^2_{adjusted} = .227$), $F(5, 123) = 8.21$, $p < .001$. Further analyses run separately for male and female candidates (see Table 3a) showed that for male candidates, perceived competence was the only significant predictor of

selection decision. Conversely, for female candidates perceived competence was not significantly related to the selection decision, which was only predicted by perceived morality. The perceived morality × gender interaction was not significant.

Considering the low-competence/high-morality candidates ($R^2 = .237$, $R^2_{adjusted} = .204$), $F(5, 123) = 8.37$, $p < .001$, perceived competence turned out to be the only significant predictor of selection decision ($\beta = .503$, $p = .002$), whereas perceived morality was not significantly related to the decision ($\beta = .234$, $p = .131$). There was no effect of candidates' gender ($\beta = .028$, $p = .744$). Neither the perceived competence × candidate gender interaction ($\beta = -.122$, $p = .441$) nor the perceived morality × candidate gender interaction ($\beta = -.153$, $p = .331$) was significant. Additional analyses confirmed that for low-competence/high-morality candidates, perceived competence was the only significant predictor of selection decision for both male and female candidates (see Table 3a).

Table 3 Regression analyses for selection decision (study 3) and renewal decision (study 4) concerning male and female candidates

Variables	(a) Selection decision (Study 3)		(b) Renewal decision (Study 4)	
	Female Candidates β (SE)	Male Candidates β (SE)	Female Candidates β (SE)	Male Candidates β (SE)
High-competence/low-morality profile				
Perceived competence	- .008 (.131)	.528 (.147) ***	- .067 (.116)	.502 (.100)***
Perceived morality	.466 (.164)***	.117 (.188)	.402** (.129)	.031 (.097)
F	8.09***	12.15***	4.47*	9.07***
R^2	.212	.290	.152	.259
$R^2_{adjusted}$	.186	.270	.136	.230
Low-competence/high-morality profile				
Perceived competence	.376 (.167)*	.409 (.197)**	.466 (.120)***	.289 (.116)*
Perceived morality	.046 (.173)	.196 (.205)	.122 (.149)	.031 (.190)
F	5.87**	11.43***	5.87***	2.52
R^2	.164	.283	.240	.088
$R^2_{adjusted}$	.136	.258	.210	.053

Study 3 included ratings of 62 female candidates and 60 male candidates. Study 4 included ratings of 52 female candidates and 54 male candidates

$*p < .05. **p < .01. ***p < .001$

Discussion

Study 3 assessed evaluators' reactions to specific candidates for whom information on morality and competence were provided. Unexpectedly, there were no differences in selection decisions concerning female and male candidates. This finding might be due to the gender-neutral characteristic of the job position. Moreover, the explicit request of selection, as well as the between-participants design, might have prevented the occurrence of gender bias. Notably, however, female candidates were perceived as lower in competence and morality than male candidates, despite candidates displayed identical qualifications.

As expected, decisions concerning women and men were based on different dimensions. Perceived competence—irrespective of whether the candidate has been presented as high or low in competence—was the only predictor of selection decisions concerning male candidates. Ratings of both perceived competence and morality informed decisions concerning female candidates. These findings therefore support our hypothesis that men are mainly evaluated on competence whereas multiple criteria are considered important for female candidates. However, Study 3 also specifies the multiple criteria assumption, highlighting that selection decisions concerning female candidates were systematically based on their weak points: Decisions were based on perceived morality when female candidates were described as high in competence, but were based on perceived competence when female candidates were presented as high in morality.

Study 4

Previous research has suggested that evaluators rely less on gender stereotypical expectations when they can base decisions on actual observations of candidates' performance (Bosak and Sczesny 2011; Davison and Burke 2000). Thus, in Study 4 participants were provided with information about the prior performance of female and male candidates in the job under consideration. Candidates had supposedly been evaluated by their supervisor as relatively high in competence-related traits and low in morality-related traits, or vice-versa. Participants reported their perception of candidates' competence and morality, and they indicated the likelihood that each of them would obtain a contract renewal. Thus, besides providing a conceptual replication of Study 3, Study 4 examined whether women are evaluated along multiple criteria even when evaluators can rely on more concrete information about job candidates—a condition which is known to reduce the influence of gender-based expectations on evaluative decisions (Heilman et al. 2015).

In general terms, we expected a greater likelihood of contract renewal for candidates with a high-competence/low-morality profile. Basing on the findings of Study 3, we anticipated that overall female candidates would be judged less favorably than male candidates, despite having received identical performance evaluations. Moreover, we expected that decisions about women would be influenced mainly by ratings on their relatively weaker characteristics. Conversely, we expected that for men, perceived competence would be the primary determinant of the renewal decision.

Method

Participants

Undergraduate students attending a Psychology course at a large Italian university filled in a paper-and-pencil questionnaire at the end of a class and on a voluntary basis. This recruitment resulted in a convenient sample of 108 participants (61, 56.5%; women; $M_{age} = 21.73$, $SD = 4.88$, range = 19–66 years-old). The sample size was deemed sufficient considering Harris's (1985) guideline of $n > 50 +$ the number of predictors for regression sample sizes, and with 23–32 participants per cell, it was a little below the recommended number of 30 participants per cell for ANOVAs (Wilson Van Voourhis and Morgan 2007). A sensitivity analysis showed that our sample was sufficient to detect medium-size effects of $f^2 = .19$ (corresponding to $R^2 = .16$), assuming an alpha of .05 and power of .95 for a multiple regression model.

Procedure

The present study was approved by the Bioethical Committee of the first author's institution in May 2014. Participants provided informed consent before being presented with the questionnaire. Procedure and measures were similar to Study 3. In the questionnaire, participants were asked to imagine that they were part of the Teaching Board of their Department. The instructions, including the description of the tutoring job, were the same as Study 3, except that they referred to the renewal of the contract to candidates. Specifically, the instructions reported:

> The Teaching Board has to decide about the renewal of contract to Master's students who have been working as tutors at the university in the past year. Each candidate has been evaluated from his/her supervisor on a series of characteristics. You will be presented with the profiles of two candidates for contract renewal. For each candidate, please read the supervisor's evaluations and answer the questions that follow thereafter.

Each participant was provided with the performance evaluations of two (female or male) candidates. Evaluations were given on the same two clusters of traits pertaining to morality and competence used in Study 3. In the present study, the contrast between the two clusters was however slightly increased. In one cluster, scores ranged from 8 to 10 (indicating high competence/morality; overall $M = 8.75$); in the other, scores ranged from 4 to 6 (indicating low competence/morality; overall $M = 5.00$). Each participant was presented with a high-competence/low-morality and a low-competence/high-morality candidate in a randomized order.

Perceived Competence For each candidate, perceived competence was measured by asking participants: "If the candidate's contract is renewed, how likely is it that the candidate will work in an efficient manner?"; "If the candidate's contract is renewed, how likely is it that the candidate will be able to make a personal contribution to the project?" (1 = *very unlikely*; 7 = *very likely*; $\alpha_{high\text{-}competence/low\text{-}morality} = .79$; $\alpha_{low\text{-}competence/high\text{-}morality} = .82$).

Perceived Morality Perceived morality was measured with two items: "If the candidate's contract is renewed, how likely is it that the candidate will behave in a correct manner with the people involved in the project?"; "If the candidate's contract is renewed, how likely is it that the candidate will be able to gain the trust of the other people involved in the project?" (1 = *very unlikely*; 7 = *very likely*; $\alpha_{high\text{-}competence/low\text{-}morality} = .86$, $\alpha_{low\text{-}competence/high\text{-}morality} = .74$).

Renewal Decision and Manipulation Checks Two items measured the likelihood that each candidate had his/her contract renewed: "To what extent do you think the Teaching Board will renew the contract to the candidate?", "How likely is it that you would renew the contract to the candidate?" (1 = *very unlikely*; 7 = *very likely*; $\alpha_{high\text{-}competence\text{-}low\ morality} = .61$; $\alpha_{low\text{-}competence/high\text{-}morality} = .80$). The manipulation checks were the same as Study 3.

Results

Manipulation Checks

Participants correctly reported the gender of the candidates they had been asked to evaluate. The job was perceived as equally suitable for women ($M = 5.69$, $SD = 1.10$) and men ($M = 5.78$, $SD = 1.08$), $F(1, 104) = 1.88$, $p = .173$.

Perceived Competence

A 2 (evaluator gender) × 2 (candidate gender) × 2 (evaluation profile) ANOVA showed that high-competence/low-morality candidates were rated more competent than low-competence/high-morality candidates, $F(1, 104) = 17.65$, $p < .001$, $\eta_p^2 = .145$ (see Table 2b). No other effects were significant ($Fs < 3.05$, $ps > .084$).

Perceived Morality

Low-competence/high-morality candidates were perceived as higher in morality than high-competence/low-morality candidates (Table 2b), $F(1, 104) = 338.80$, $p < .001$, $\eta_p^2 = .765$. The evaluator gender × profile interaction was also significant, $F(1, 104) = 4.93$, $p = .029$, $\eta_p^2 = .045$. However, pairwise comparisons between ratings provided by female and male

evaluators were not significant ($ps > .060$). No other effects were significant ($Fs < 2.18$, $ps > .143$).

Renewal Decision

Participants indicated a higher likelihood of renewal decision for high-competence/low-morality candidates than for low-competence/high-morality candidates, $F(1, 104) = 17.02$, $p < .001$, $\eta_p^2 = .141$ (see Table 2b). Moreover, across profiles male candidates were more likely to have the contract renewed than female candidates, $F(1, 104) = 16.15$, $p < .001$, $\eta_p^2 = .134$. No other effects were significant ($Fs < 2.20$, $ps > .141$).

Regression Analyses

Table 2b shows correlations among measures. As in Study 3, a series of regression analyses examined the impact of perceived competence and perceived morality on renewal decisions and tested for moderation of candidates' gender. With respect to the high-competence/low-morality candidates, findings yielded main effects of perceived competence ($\beta = .462$, $p < .001$) and candidate gender ($\beta = -.315$, $p < .001$). Perceived morality did not predict renewal decisions ($\beta = .027$, $p = .820$). The analysis revealed significant perceived competence × gender ($\beta = -.388$, $p = .003$) and perceived morality × gender ($\beta = .268$, $p = .027$, $R^2 = .285$, $R^2_{adjusted} = .250$), $F(5, 109) = 8.51$, $p < .001$, interactions, indicating that candidates' gender moderated the impact of ratings of both perceived competence and morality. Further analyses run separately for high-competence/low-morality male and female candidates (see Table 3b) showed that as far as male candidates are concerned, perceived competence was the only significant predictor of the renewal decision. Conversely, decisions concerning female candidates were predicted by perceived morality.

With respect to the low-competence/high-morality candidates ($R^2 = .224$, $R^2_{adjusted} = .186$), $F(5, 109) = 8.61$, $p < .001$, perceived competence predicted the renewal decision ($\beta = .265$, $p = .035$) whereas there was no effect of perceived morality ($\beta = .034$, $p = .819$). Candidates' gender was also significantly related to renewal decision ($\beta = -.215$, $p = .016$), reflecting the general pattern that female candidates were less likely than male candidates to have their contract renewed. Neither the perceived competence × gender interaction ($\beta = .155$, $p = .211$) nor the perceived morality × gender interaction ($\beta = .070$, $p = .634$) was significant. Perceived competence was the only significant predictor of the likelihood of contract renewal for both male and female low-competence/high-morality candidates.

Discussion

In Study 4, no differences were found with respect to perceptions of competence and morality of female and male candidates, possibly because participants reacted to information relative to competent and moral behaviors which candidates had supposedly shown in the job under consideration. However, findings revealed a preference for male over female candidates, despite information documenting identical past performance on the job as well as the fact that the job itself was equally suitable for men and women.

As expected, decisions concerning male candidates were mainly predicted by ratings of perceived competence, whereas decisions concerning female candidates were based on evaluators' perceptions of the characteristics where females appeared to be relatively weaker. Namely, perceived competence predicted decisions concerning female candidates who had been presented as relatively low in competence and high in morality. Perceived morality predicted decisions concerning candidates who had been presented as relatively low in morality but high in competence. Thus, overall Studies 3 and 4 are consistent in suggesting that decisions concerning women are based on multiple criteria and are especially influenced by the dimension along which women do not reach high standards of performance.

General Discussion

Our research examined the relative importance of different dimensions—competence, morality and sociability (Brambilla and Leach 2014)—in evaluations of female and male job candidates. Overall, competence played a key role in evaluation and employment decisions. However, the findings revealed that women are evaluated against more criteria than men are and that women's weaknesses along a single dimension are likely to affect employment decisions. Decisions concerning men are instead more clearly influenced by their level of competence.

In Study 1, professional selectors relied on a wider range of qualities in their evaluations of female candidates compared to evaluations of male candidates: Even though competence-related terms were the most mentioned, morality- and sociability-related terms were mentioned more often in justifying selection as well as rejection decisions involving women. Selection decisions concerning male candidates were mainly motivated on grounds of positive competence-related qualities; rejection, on grounds of their less-than-optimal competence.

In Study 2, competence was judged as the most important quality candidates should have. However, competence, morality, and sociability were all considered relatively more important by naïve selectors who had been required to think of

female rather than male candidates. Studies 3 and 4 revealed that hiring and retention decisions concerning male candidates were predicted primarily by evaluators' perception of their (high or low) level of competence. Perceived competence was instead the main predictor of decisions concerning female candidates only when female candidates had been presented as relatively low in competence. Decisions concerning female candidates were in fact influenced mainly by their perceived morality when candidates had been presented as relatively low in morality.

In this respect, it is important to note that evaluators' focus on women's weakness in the moral domain might imply an additional risk for female candidates. Indeed, previous research has revealed that negative information about others' morality are more diagnostic and harder to modify than negative information on competence or sociability (Brambilla et al. 2016). This emphasis might represent a further means through which women's achievements in the workplace might be hindered: Not only women are excluded from current professional opportunities for shortcomings that are forgiven in men, but their overall reputation can be damaged by this focus on their perceived flaws in the moral domain.

Overall, our evidence highlights that individuals seem to rely on a more complex and elaborate set of requirements to select female candidates. They also appear to be more influenced by information concerning women's flaws. This can be seen as a consequence of the higher attention paid to multiple criteria: Because all dimensions are important and are therefore taken into consideration in evaluations of female candidates, a weakness on a given dimension is more likely to capture the attention of evaluators and to influence their final decision. The finding that employment decisions concerning women were mainly influenced by their weakness (rather than by information on both the dimensions we considered) can be accounted for by the greater power of negative over positive information in impression formation (Baumeister et al. 2001).

More generally, one might say that evaluators follow two different strategies in decisions concerning men and women. In evaluations of men, they search for work-related information that might support (or, eventually, not support) their general belief that to the extent that men are competent, they suit the work context. Namely, evaluators rely on a heuristic information processing anchored to the idea that being competent is enough for men. In evaluations of women, evaluators might be influenced by their intuitive assumption that women are likely to encounter more difficulties than men in their job. This assumption might include the perception of lack of fit between women's qualities and specific job requirements (Heilman et al. 2015). However, even though evaluators think that a job is perfectly suitable for a woman as for a man (as participants did in our studies), as well as recognize that a female candidate has the right credentials to get the job, they might still think that a woman is likely to encounter more

obstacles in the workplace. For instance, they might think that a woman will be less easily accepted and less respected by her (especially men) colleagues, supervisors, or clients, who might feel more comfortable with a man (see van den Brink and Benschop 2011, for a similar reasoning). Additionally, evaluators might also figure that, in the long term, women might decrease their commitment to the job for personal reasons (for instance, because of "maybe baby" expectations; Gloor et al. 2018, p. 45). Overall, the assumption that women are likely to encounter more obstacles in the working environment might lead evaluators to carry out a more comprehensive scrutiny of information concerning female candidates. In other words, evaluators might adopt more systematic information processing in decisions concerning women. However, this more comprehensive and systematic scrutiny of information is likely to end up sustaining evaluators' focus on women's weaknesses and therefore to lead to the exclusion of women.

Our speculation is consistent with van den Brink and Benschop's (2011) explanation of gender practices in academia. Examining how women and men are appointed to full professor positions, the authors highlighted that men's individual qualities (e.g., physical appearance, perceived personality) contribute to their likeability (e.g., being nice), whereas the same qualities seem to work as disqualifiers for women (e.g., being too nice). In a similar fashion, in our studies evaluators might have scrutinized women's—but not men's—characteristics in search of negative information that can justify the exclusion of women. Obviously, ours is just a tentative account of the possible mechanisms driving the effects we observed in our research, which would deserve more thoroughly investigation.

Limitations and Future Research Directions

A first objection that might be advanced to our interpretation of results is that these studies did not consistently demonstrate that women are systematically preferred to men in hiring (with the exception of Study 4). This could be at least in part due to the fact that, in Studies 1 and 2, we did not directly examined hiring decisions. Moreover, with the exception of Study 1, we purposefully decided to consider jobs that were equally suitable for men and women. Moreover, the use of between participants designs—with participants considering either female or male candidates (Studies 2 to 4)—might have not favored the occurrence of more explicit gender-based discrimination. As mentioned, we detected an explicit preference for male over female candidates in retention decisions (Study 4). Thus, the present studies do not allow us to establish a causal relationship between the differences in evaluation that we detected and actual discrimination against women in hiring or retention decisions. Future research should address this issue by examining whether the reliance on multiple criteria in

evaluating female candidates has a direct influence on (disadvantaging) selection decisions.

A related limitation is that our evidence (with the exception of Study 1) was mostly based on laboratory studies. Future research should therefore consider the mechanisms we documented here in different real-life settings and should provide further evidence of evaluators' reliance on multiple criteria for women by considering a wider range of judgments. From a theoretical perspective, it would be interesting including information about candidates' sociability (e.g., Leach et al. 2007). More generally, it is crucial to test whether information about female and male candidates' behaviors in the private domain—information that nowadays can easily be found through social networks—are more influential on selectors' impressions and final decisions when candidates are women rather than men. In a related way, future studies might employ a more comprehensive design whereby positive and negative information about job candidates are fully crossed (for instance, by comparing evaluations of women who are highly qualified along multiple dimensions and evaluations of men who are competent but less qualified than women on other dimensions). Moreover, future studies might examine more in depth whether individuals follow different evaluative strategies in decisions about female and male candidates. In particular, it is important to test whether evaluators are more interested in gathering information that justify women's exclusion or whether they spend more time processing and discussing negative information concerning women than negative information concerning men.

Finally, it would be worth examining how differences in evaluation of women and men can be prevented and counteracted. This exploration might be done by assessing whether specific conditions (e.g., avoiding time pressure, consideration of relevant criteria before assessing candidates, provision of sufficient individuating information; e.g., Prati et al. 2015) might lead to higher gender equality in evaluation.

Implications for Theory and Research

Theoretically, our studies extend research on gender stereotypes (e.g., Ellemers 2018; Heilman 2012) and impression formation (e.g., Brambilla et al. 2011, 2012). In fact, they highlight that morality—distinct from communality/warmth—is gendered: Women are expected to show high morality, and flaws on morality can compromise women's career prospects. Moreover, morality, when made explicit as an evaluation criterion, seems to play a more important role than sociability in employment decisions.

More generally, the present findings go beyond previous evidence of gender bias in hiring. Indeed, they suggest that besides having to provide more evidence of skills than men (see also Biernat and Fuegen 2001), women might be expected to be simultaneously strong on *more* dimensions. In a

related way, our research extends evidence on backlash effects (Phelan et al. 2008; Rudman and Glick 2001) because it shows that when female candidates are highly competent, evaluators focus on their moral weakness, as distinct from generic social skills, instead of appreciating their performance-related qualities.

Practice Implications

Our findings have important implications for personnel selection practices as well as for policies aiming at counteracting gender biases in organizations. From the perspective of candidates, women who want to apply for a job not only need to show high qualifications on competence, but also should be sure to provide explicit information about their moral values and social skills. Recommendations like these, however, unfairly place the burden of proving their merit in all domains on the shoulders of women.

These findings might also be fruitfully used to improve selection procedures. Selectors should be trained to develop assessment strategies that prevent them from using more criteria in their evaluations of women. For instance, selectors could be encouraged to establish specific evaluation criteria and specify their importance for the job under consideration before starting the selection process. Then, they should rely on these pre-defined criteria in evaluation prior to disclosure of candidates' gender and before developing overall impressions (for similar reasoning, see Uhlmann and Cohen 2005).

As we mentioned, in our studies both female and male evaluators set multiple criteria for women. In Study 3, female evaluators were even more critical of women candidates than male evaluators were. Overall, these findings lead us to argue that having more women in charge of the selection process might not in itself be sufficient to eliminate biased treatment of female candidates (see also Faniko et al. 2017; Prati et al. 2019). Nevertheless, it is possible that, if trained, female selectors would more easily recognize and accept evidence of gender differences than men would and might possibly be more willing to adapt their evaluations and decisions accordingly. This conjecture is in line with prior findings that men (including university professors) consider evidence on gender bias in scientific faculties as less convincing than women do (Handley et al. 2015).

In a related way, because team diversity enhances members' awareness of others' possible diverging views and therefore the quality of team decision-making (Hoever et al. 2012), it is possible that having selection committees composed by both women and men would result in more balanced evaluations of male and female candidates (Prati et al. 2019). More generally, it is important that managers and other stakeholders unambiguously promote and effectively manage gender diversity. In order to realize the performance benefits of increased organizational diversity (e.g., Ellemers and Rink 2016),

organizations cannot just rely on the assumption that decisionmakers are able to judge the merits of male and female candidates fairly, but should actively guard against the risk that women end up having to meet higher expectations on more criteria in order to be given the same opportunities as men.

Conclusions

The present studies showed that evaluators rely on more dimensions in judging female than male job candidates so that women's weakness on a single dimension might influence the employment decision. Conversely, men are mainly judged on competence. In our view, these findings suggest that women must "have it all" to have a chance to be selected and, if they do not, they might be targets of a *perfection bias*: Because women are judged on multiple dimensions, they might be required to excel in every domain against which they are evaluated. In other words, women not only have to demonstrate more competence in their professional domain than their male counterparts do—as highlighted by previous research (e.g., Biernat and Fuegen 2001)—but also have to meet additional requirements not set for men to be hired and to make a career. This might represent a powerful mechanism that can hinder women's possibilities of being hired and enhance women's risk to be rejected if they do not reach high standards of qualification in all the considered domains.

References

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*, 1–17. https://doi.org/10.3389/fpsyg.2016.01810.

Baez, S., Flichtentrei, D., Prats, M., Mastandueno, R., Garcia, A. M., Cetkovich, M., … Ibañez, A. (2017). Men, women … who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLoSONE, 12*(6), e0179336. https://doi.org/10.1371/journal.pone.0179336.

Barreto, M., & Ellemers, N. (2005). The perils of political correctness: Men's and women's responses to old-fashioned and modern sexist views. *Social Psychology Quarterly, 68*, 75–88. https://doi.org/10.1177/019027250506800106.

Barreto, M., Ryan, M., & Schmitt, M. (2009). *The glass ceiling in the 21st century: Understanding barriers to gender inequality* (pp. 182–303). Washington, DC: American Psychological Association.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370. https://doi.org/10.1037//1089-2680.5.4.323.

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*, 707–724. https://doi.org/10.1111/0022-4537.00237.

Bosak, J., & Sczesny, S. (2011). Gender bias in leader selection? Evidence from a hiring simulation study. *Sex Roles, 65*, 234–242. https://doi.org/10.1007/s11199-011-0012-7.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition, 32*, 397–408. https://doi.org/10.1521/soco.2014.32.4.397.

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*, 135–143. https://doi.org/10.1002/ejsp.744.

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbit, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology, 51*, 149–166. https://doi.org/10.1111/j.2044-8309.2010.02011.x.

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*, 171–186. https://doi.org/10.1007/s10919-016-0229-9.

Bruckmüller, S., Ryan, M. K., Haslam, S. A., & Peters, K. (2013). Ceilings, cliffs, and labyrinths: Exploring metaphors for workplace gender discrimination. In M. K. Ryan & N. R. Branscombe (Eds.), *The Sage handbook of gender and psychology* (pp. 450–465). Thousand Oaks, CA: Sage.

Budig, M. (2002). Male advantage and the gender composition of jobs: Who rides the glass escalator? *Social Problems, 49*, 258–277. https://doi.org/10.1525/sp.2002.49.2.258.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology, 107*, 943–963. https://doi.org/10.1037/a0037245.

Crocetti, E., Moscatelli, S., Kaniušonytė, G., Meeus, W., Žukauskienė, R., & Rubini, M. (2019). Developing morality, competence, and sociability in adolescence: A longitudinal study of gender differences. *Journal of Youth and Adolescence, 48*, 1009–1021. https://doi.org/10.1007/s10964-019-00996-2.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225–248. https://doi.org/10.1006/jvbe.1999.1711.

de Oliveira, S. H., Ksenofontov, K., & Becker, J. C. (2015). Explicit but not implicit sexist beliefs predict benevolent and hostile sexist behavior. *European Journal of Social Psychology, 45*, 702–715. https://doi.org/10.1002/ejsp.2128.

Deszö, C. L., & Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal, 33*, 1072–1089. https://doi.org/10.1002/smj.1955.

Ellemers, N. (2017). *Morality and the regulation of social behavior*. Milton Park: Routledge.

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

Ellemers, N., & Rink, F. (2016). Diversity in work groups. *Current Opinion in Psychology, 11*, 49–53. https://doi.org/10.1016/j.copsyc.2016.06.001.

European Commission. (2018). *2018 report on equality between women and men in the EU*. Retrieved from https://ec.europa.eu/newsroom/just/document.cfm?doc_id=50074

Faniko, K., Ellemers, N., Derks, B., & Lorenzi-Cioldi, F. (2017). Nothing changes, really: Why women who break through the glass ceiling end up reinforcing it. *Personality and Social Psychology Bulletin, 43*, 638–651. https://doi.org/10.1177/0146167217695551.

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*, 175–191. https://doi.org/10.3758/BF03193146.

Franke, G. R., Crown, D. F., & Spake, D. F. (1997). Gender differences in ethical perceptions of business practices: A social role theory perspective. *Journal of Applied Psychology, 82*, 920–934. https://doi.org/10.1037/0021-9010.82.6.920.

Fruhen, L., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial attractiveness, dominance and trustworthiness predict managerial pay awarded in experimental tasks. *Leadership Quarterly, 26*, 1005–1016. https://doi.org/10.1016/j.leaqua.2015.07.001.

Fulton, L., & Sechi, C. (2018). ETUC annual gender equality survey 2018. *European Trade Union Institute*. Retrieved from https://www.etuc.org/system/files/circular/file/2018-06/ETUC%20Annual%20Gender%20Equality%20Survey%202018.pdf

Geiger, A. W., & Parker, K. (2018, March). For Women's history month, a look at gender gains – And gaps – In the U.S. *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2018/03/15/for-womens-history-month-a-look-at-gender-gains-and-gaps-in-the-u-s/

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge: Harvard University Press.

Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512. https://doi.org/10.1037/0022-3514.70.3.491.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology, 55*, 178–186. https://doi.org/10.1037/0022-3514.55.2.178.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J., Abrams, D., Masser, B., … López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology, 79*, 763–775. https://doi.org/10.1037/0022-3514.79.5.763.

Gloor, J. L., Li, X., Lim, S., & Feierabend, A. (2018). An inconvenient truth? Interpersonal and career consequences of "maybe baby" expectations. *Journal of Vocational Behavior, 104*, 44–58. https://doi.org/10.1016/j.jvb.2017.10.001.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148–168. https://doi.org/10.1037/a0034726.

Gorman, E. H. (2005). Gender stereotypes, same-gender preferences, and organizational variation in the hiring of women: Evidence from law firms. *American Sociological Review, 70*, 702–728. https://doi.org/10.1177/000312240507000408.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 499–510. https://doi.org/10.1207/s15327906mbr2603_7.

Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *PNAS, 112*, 13201–13206. https://doi.org/10.1073/pnas.1510649112.

Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). New York: Academic Press.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416–427. https://doi.org/10.1037/0021-9010.89.3.416.

Heilman, M. E., Manzi, F., & Braun, S. (2015). *Presumed incompetent*: *Perceived lack* of fit and *gender bias* in *recruitment* and *selection*. In A. M. Broadbridge & S. L. Stirling (Eds.), *Handbook of gendered careers in management: Getting in, getting on, getting out* (pp. 90–104). Cheltenham: Edward Elgar.

Hoever, I. J., van Knippenberg, D., van Ginkel, W. P., & Barkema, H. G. (2012). Fostering team creativity: Perspective taking as key to unlocking diversity's potential. *Journal of Applied Psychology, 97*, 982–996. https://doi.org/10.1037/a0029159.

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*, 234–249. https://doi.org/10.1037/0022-3514.93.2.234.

Leach, C. W., Carraro, L., Garcia, R. L., & Kang, J. J. (2017). Morality stereotyping as a basis of women's in-group favoritism: An implicit approach. *Group Processes & Intergroup Relations, 20*, 153–172. https://doi.org/10.1177/1368430215603462.

Levin, I., Rouwenhorst, R., & Trisko, H. (2005). Separating gender biases in screening and selecting candidates for hiring and firing. *Social Behavior and Personality, 33*, 793–804. https://doi.org/10.2224/sbp.2005.33.8.793.

Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people–things dimension. *Journal of Personality and Social Psychology, 74*, 996–1009. https://doi.org/10.1037/0022-3514.74.4.996.

Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology, 94*, 1591–1599. https://doi.org/10.1037/a0016539.

Manganelli Rattazzi, A. M., Volpato, C., & Canova, L. (2008). L'atteggiamento ambivalente verso donne e uomini. Un contributo alla validazione delle scale ASI e AMI [the ambivalent attitudes towards women and men. A contribution to the validation of ASI and AMI scales]. *Giornale Italiano di Psicologia, 35*, 217–243. https://doi.org/10.1421.26601.

Menegatti, M., & Rubini, M. (2017). Gender bias and sexism in language. In H. Giles & J. Harwood (Eds.), *Oxford Research Encyclopedia, Communication* (Vol. 1, pp. 451–468). New York: Oxford University Press.

Menegatti, M., Crocetti, E., & Rubini, M. (2017). Do gender and ethnicity make the difference? Linguistic evaluation bias in primary school. *Journal of Language and Social Psychology, 36*, 415–437. https://doi.org/10.1177/0261927X17694980.

Menegatti, M., Moscatelli, S., Brambilla, M., & Sacchi, S. (2019). *The honest mirror: Morality as a moderator of spontaneous behavioral mimicry*. Manuscript submitted for publication.

Moscatelli, S., Menegatti, M., Albarello, F., Pratto, F., & Rubini, M. (2019). Can we identify with a nation low in morality? The heavy weight of (im) morality in international comparison. *Political Psychology, 40*, 93–110. https://doi.org/10.1111/pops.12504.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109.

Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25–29. https://doi.org/10.1016/j.econlet.2014.12.019.

Pagliaro, S., Brambilla, M., Sacchi, S., D'Angelo, M., & Ellemers, N. (2013). Initial impressions determine behaviors: Morality predicts the willingness to help newcomers. *Journal of Business Ethics, 117*, 37–44. https://doi.org/10.1007/s10551-012-1508-y.

Pagliaro, S., Ellemers, N., Barreto, M., & Di Cesare, C. (2016). Once dishonest, always dishonest? The impact of perceived pervasiveness of moral evaluations of the self on motivation to restore a moral reputation. *Frontiers in Psychology, 7*, 586. https://doi.org/10.3389/fpsyg.2016.00586.

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash

towards agentic women. *Psychology of Women Quarterly, 32*, 406–413. https://doi.org/10.1111/j.1471-6402.2008.00454.x.

Prati, F., Vasiljevic, M., Crisp, R. J., & Rubini, M. (2015). Some extended psychological benefits of challenging social stereotypes: Decreased dehumanization and a reduced reliance on heuristic thinking. *Group Processes & Intergroup Relations, 18,* 801–816. https://doi.org/10.1177/1368430214567762

Prati, F., Moscatelli, S., Van Lange, P. A. M., Van Doesum, N. J., & Rubini, M. (2018). The central role of morality in perceived humanness and unselfish behaviors. *Social Psychology, 49*, 330–343. https://doi.org/10.1027/1864-9335/a000352.

Prati, F., Menegatti, M., Moscatelli, S., Kana Kenfack, C. S., Pireddu, S., Crocetti, E., Mariani, M. G., & 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*. Advance online publication. https://doi.org/10.1177/0261927X19844808

Ramos, M. R., Barreto, M., Ellemers, N., Moya, M., & Ferreira, L. (2018). What hostile and benevolent sexism communicate about men's and women's warmth and competence. *Group Processes & Intergroup Relations, 21*, 159–177. https://doi.org/10.1177/1368430216656921.

Rubini, M., & Menegatti, M. (2014). Hindering women's careers in academia: Gender linguistic bias in personnel selection. *Journal of Language and Social Psychology, 33*, 632–650. https://doi.org/10.1177/0261927X14542436.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743–762. https://doi.org/10.1111/0022-4537.00239.

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*, 165–179. https://doi.org/10.1016/j.jesp.2011.10.008.

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly, 27*, 446–455. https://doi.org/10.1016/j.leaqua.2015.10.008.

Sackett, P. R., & Schmitt, N. (2012). On reconciling conflicting meta-analytic findings regarding integrity test validity. *Journal of Applied Psychology, 97*, 550–556. https://doi.org/10.1037/a0028167.

Sheppard, L. D., & Johnson, S. K. (2019). The femme fatale effect: Attractiveness is a liability for businesswomen's perceived truthfulness, trust, and deservingness of termination. *Sex Roles*. Advance online publication. https://doi.org/10.1007/s11199-019-01031-1

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*, 474–480. https://doi.org/10.1111/j.0956-7976.2005.01559.x.

van den Brink, M., & Benschop, Y. (2011). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization, 19*, 507–524. https://doi.org/10.1177/1350508411414293.

Williams, C. (1992). The glass escalator: Hidden advantages for men in the "female" professions. *Social Problems, 39*, 253–267. https://doi.org/10.2307/3096961.

Willis, J., & Todorov, A. (2006). First impression: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x.

Wilson Van Voourhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology, 3*, 43–50. https://doi.org/10.20982/tqmp.03.2.p043.

Zaikman, Y., & Marks, M. J. (2014). Ambivalent sexism and the sexual double standard. *Sex Roles, 71*, 333–344. https://doi.org/10.1007/s11199-014-0417-1.