Better for Less DISCUSSION PAPER SERIES

DP No. 09/2018

# Now you see it now you don't:

# The effect of teaching style and seniority on gender bias in teaching evaluations.

**Nigel Burnell**

School of Economics, University of Surrey, GU2 7XH, United Kingdom

N.Burnell@surrey.ac.uk

**Irina Cojuharenco**

Surrey Business School, University of Surrey, GU2 7XH, United Kingdom

I.Cojuharenco@surrey.ac.uk

**Zahra Murad**

Department of Economics and Finance, University of Portsmouth, PO1 3DE, United Kingdom

zahra.muard@port.ac.uk

School of Economics, Faculty of Arts and Social Sciences, University of Surrey, Elizabeth Fry Building 04 AD 00, Guildford, GU2 7XH, Surrey, UK. T: +44(0)01483686623, E-mail: deliveringbetter@surrey.ac.uk, Web: https: www.surrey.ac.uk/better-for-less

## Abstract

Gender bias in teaching evaluations leads to unfair decisions during academics' careers. In two controlled experiments, we examine whether gender bias is eliminated by an academic's high warmth teaching style and by seniority. We find that gender bias lowers recommendations to hire female academics delivering identical content as male academics, with the effect mediated by evaluations of the academic's warmth and/or competence. In Study 1, we test competing hypotheses regarding the effect of teaching style on gender bias. We find that a high warmth teaching style increases women's perceived warmth, but decreases their perceived competence, so gender bias in hiring recommendations remains. In Study 2, we find that gender bias disappears for senior academics. Finally, we find no evidence of less biased evaluations by those who anticipate gender bias. We discuss our findings in the higher education context and make recommendations to mitigate gender bias in teaching evaluations.

**Keywords:**

Gender bias, teaching evaluations, teaching style, seniority, bias awareness.

## Introduction

The recent decades have seen a surge of evidence in higher education settings pointing to a gender bias in teaching evaluations (Langbein, 1994; MacNell, Driscoll, & Hunt, 2015; Ottoboni, Boring, & Stark, 2016; Pounder, 2007; Wagner, Rieger, & Voorvelt, 2016; Young, Rush, & Shaw, 2009). This is a problem because there is a tradition to associate teaching evaluations with educational outcomes and to decide on the careers of academics based on teaching evaluations (Wild & Berger, 2016). To the extent that teaching evaluations assess academics in a biased way, based on their gender rather than specific behaviours, decisions that are key to academic careers may be unfair. For example, the under-representation of women in senior academic roles, especially in male-dominated disciplines, may be due to unfair decisions early in the careers of female academics (Dick & Nadin, 2006; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Newsome, 2008; Sheltzer & Smith, 2014; Way, Larremore, & Clauset, 2016).

A recent quasi-experimental study of 19,920 teaching evaluations at Maastricht University in the Netherlands showed that it is female academics who tend to be evaluated less positively, especially if they are junior, in male-dominated disciplines, and rated by male students (Mengel, Sauermann, & Zölitz, 2017). Interestingly, the gender bias in teaching evaluations mirrors a recent meta-analysis of gender bias in employment decision making (Koch, D'Mello, & Sackett, 2015). The latter typically examine decisions to hire a candidate, but have not looked at variables of relevance to the study of teaching evaluations, such as the teaching style or seniority of the candidate. What controlled experiments add to quasi-experimental or correlational designs is a ruling out of possible differences in such variables between female and male academics, and a possibility to estimate their effects in isolation and as moderators of the effect of gender (Arbuckle & Williams, 2003; Doubleday & Lee, 2016; MacNell et al., 2015; Ottoboni et al., 2016; Wagner et al., 2016).

In what follows, we build on the literatures in management, economics, and education, to formulate testable research hypotheses regarding the role of teaching style and seniority on gender bias in teaching evaluations and hiring recommendations. We perform controlled experiments testing competing hypotheses regarding the effects of teaching style and the de-biasing role of seniority. In addition, we survey lay intuitions of experimental participants regarding a possible gender bias in teaching evaluations and hiring recommendations, and examine whether the bias is expected, and if so, what is the effect of bias awareness. Overall, our work contributes to a more fine-grained understanding of the gender bias in teaching evaluations allowing us to identify conditions under which the bias disappears.

To summarize, we find that a gender stereotypical ("warm") teaching style improves perceptions of warmth for female academics but may backfire by lowering perceptions of their competence. Hence, hiring recommendations are lower for female academics (vs. male) irrespective of their style because of a double-bind nature of reactions to their teaching. If their teaching style is low in warmth, lower hiring recommendations are driven by lower perceptions of their warmth, and if their teaching style is warm, lower hiring recommendations are driven by lower perceptions of their competence. Fortunately, gender bias is sensitive to seniority, and we find no evidence of bias against senior female academics in hiring recommendations or warmth evaluations even when they teach in a low warmth style. In conjunction with findings from previous research, these results suggest a need to shield junior academics from decisions that rely on teaching evaluations, especially in the early stages of their careers (Mengel et al., 2017). Moreover, they highlight possible benefits from showcasing titles and other credentials that may indicate more senior standing for female academics. An unexpected finding has been to find lower perceptions of warmth for senior (vs. junior) male academics suggesting that senior male academics, unlike their female

colleagues, may not need to be concerned with showcasing seniority. Last but not least, those aware of gender bias hurting female academics are no more likely to correct their evaluations, suggesting caution in treating awareness alone as an effective remedy to the problem.

## Theory Development

### The role of teaching style

Academics are commonly evaluated on criteria that align with the two universal dimensions of social cognition: warmth and competence (Fiske, Cuddy, & Glick, 2007). For example, recommendations of research councils suggest assessments of warmth-related "enthusiasm", "consideration" and "accessibility" and competence-related "class structure", "mastery of material" and "level of preparation" (Hannover Research Council, 2009). Experimental evidence to date has found significant bias against female academics on both dimensions, including criteria such as enthusiasm, praise, respect and fairness (warmth) and promptness and professionalism (competence) (MacNell et al., 2015).

From a theoretical perspective, teaching evaluations are indeed ripe for gender bias. Teaching is a power relationship that highlights the dependence of the student on the goodwill, mastery and knowledge of the instructor (Schrodt, Witt, & Turman, 2007). The performance of the academic is highly salient to the student as the very reason why students enter the relationship. As a result, students are naturally inclined to judge various aspects of the academic's performance in the classroom. Often, the judgment is made under time pressure and intuitively (Bassett, Cleveland, Acorn, Nix, & Snyder, 2017; Pinto & Mansfield, 2010). Moreover, higher education is a credence-based service as students lack the knowledge necessary to confidently judge the academic, especially concerning competence (Darby & Karni, 1973; Kasnakoglu, 2016). This makes the evaluation of performance through teaching evaluations highly uncertain (Gruber & Frugone, 2011). Gender stereotypes

and considerations of gender-role congruity become an important source of information that helps address the uncertainty in the teaching relationship (Davison & Burke, 2000; Kunda & Spencer, 2003). Yet, the reliance on gender stereotypes and considerations of gender-role congruity is likely to favour male as opposed to female academics because women are typically believed to be less competent than men and less fit to occupy positions of power (Eagly & Karau, 2002). This is particularly true of more male-dominated disciplines, which reinforce the stereotype and established gender roles, making them more salient in judgment (Cejka & Eagly, 1999; Koch et al., 2015).

However, research has also shown that one effective way of generating more positive and accepting evaluations of competent women, such as female academics, is for the women to show warmth, a stereotypically female characteristic associated with care and the pursuit of communal goals (Carli, 2001). Unlike men, women need to show pro-sociality in addition to self-confidence in order to influence others based on their higher performance (Guillén, Mayo, & Karelaia, 2017).

In what follows we formulate competing hypotheses regarding a possible effect of a teaching style that is high on warmth and, hence, stereotypically "female". On the one hand, we suggest that in the context of teaching, evaluations of female academics may be enhanced if the style of lecture delivery is high rather than low on warmth, and more so than for male academics. Male academics who, from the start, are more likely to be perceived as fulfilling a gender-appropriate role, are simply less likely to be scrutinized in terms of their style. Our prediction is supported indirectly by content analysis of qualitative data, including comments on Ratemyprofessor.com. Adjectives that relate to high versus low warmth in teaching style (bossy, nice, caring, warm, etc.) are more likely mentioned in relation to female rather than male academics such that teaching style is more important in the assessment of female academics (Mitchell & Martin, 2018; Shen, 2015). So, if a male and a female academic teach

the same content, a teaching style that is high on warmth is likely to raise the warmth and, together with it, competence evaluations for female academics more than for male academics. This may happen to the point of possibly eliminating the gender bias in these evaluations, as well as their associated downstream consequences, such as hiring recommendations. We formulate the following research hypotheses:

***Hypothesis 1: (gender bias in hiring)*** Hiring recommendations will be lower for female than for male academics who teach the same content.

***Hypothesis 2: (mediation in hiring)*** Gender bias in hiring recommendations will be mediated by warmth and competence perceptions.

***Hypothesis 3: (effect of style on warmth and competence):*** Relative to a teaching style that is low on warmth, a teaching style that is high on warmth will raise perceptions of the academic's warmth and competence, and more so for female rather than male academics.

Collectively, Hypotheses 2-3 imply the possibility of a reduction in gender bias under a teaching style that is high on warmth. So,

***Hypothesis 4: (effect of style on gender bias in hiring):*** Gender bias in hiring recommendations will be reduced or eliminated under a teaching style that is high on warmth.

Although we predicted in Hypotheses 3-4 that a teaching style high on warmth may help overcome gender bias in the evaluations of female academics due to increasing perceptions of the female's warmth and competence, competing hypotheses are also possible. To formulate competing hypotheses, we note the specificity of the teaching context in that it is relatively easier to assess the academics' warmth rather than their competence. To the

extent that competence assessments are highly uncertain, they may be affected in the direction of the stereotype especially when the style of teaching reinforces the stereotype. In particular, because women who behave warmly reinforce the gender stereotype, observers are likely to rely more heavily on the idea that women are less competent than men, and less fit to occupy positions of power. As a result, female academics may benefit from higher perceptions of their warmth but at the same time suffer a competence penalty associated with the alignment of the style and the stereotype of someone less knowledgeable. If this was the case, then we would predict that a teaching style that is high on warmth may not diminish or eliminate the gender bias, but rather affect competence perceptions differently for male versus female academics. For women, a warm teaching style could decrease competence perceptions whereas no such effect would be expected for men. Hence, a warm teaching style would increase gender bias in competence evaluations rather than help diminish it.

***Hypothesis 3A: (competing, effect of style on warmth and competence)*** Relative to a teaching style that is low on warmth, a teaching style that is high on warmth will raise perceptions of the academic's warmth, and more so for female rather than male academics. However, it will diminish perceptions of the academic's competence, and more so for female rather than male academics.

As a result, female academics may continue to be recommended for hiring less because of their lower perceived competence and fit to the role. Depending on the weight placed on competence versus warmth as determinants of hiring recommendations, the bias may change either upward or downward, and we, therefore, limit our theorizing to the mediating role of warmth and competence for hiring recommendations.

**The role of seniority**

The fact that female academics may be doubted more in terms of their fit to the role than their male counterparts due to gender stereotypes and considerations of gender-role congruity invites the question of whether seniority has the potential to eliminate the gender bias. If, in a given setting, students require more convincing evidence to infer competence from female academics compared to male academics then a double standard exists (Rubin, 1981; Winocur, Schoen, & Sirowatka, 1989). Double standards are known to impede career advancement (Lyness & Thompson, 2000) but the attainment of a senior position implies therefore a higher level of skill or ability (Crocker & Major, 1989). Thus, where individuals reach senior positions despite the existence of double standards this may confer a positive advantage. Indeed, research shows that provided information that supports without ambiguity the high competence of candidates, gender bias disappears (Koch et al., 2015). As senior academics and especially in male-dominated disciplines, women may be judged unambiguously as highly competent. Moreover, it is likely that for senior female academics, both perceptions of warmth and competence will be high supporting their seniority proven fit to the role. Indeed leadership research has argued that, where warmth is perceived as advantageous in a role, women in top positions can be viewed as both warm and competent and enjoy an advantage in evaluations compared to male peers (Byron, 2007; Emmerik, Wendt, & Euwema, 2010; Rosette & Tost, 2010). There has been a lengthy literature on the benefits to female students of female approaches to teaching and of a role model effect (Bettinger & Long, 2005; Carrell, Page, & West, 2010; Lockwood & Kunda, 1997). This could provide the basis for a female seniority advantage in academia.

***Hypothesis 5: (effect of seniority on gender bias on hiring recommendation)*** Gender bias for junior academics will be reduced or eliminated for senior academics.

***Hypothesis 6: (effect of seniority on gender bias on warmth and competence)*** Relative to junior academics, senior academics delivering the same content will be perceived as more warm and more competent, and more so for female rather than male academics.

**Bias awareness**

A number of approaches have been suggested in the literature to overcome biases in decision-making including gender bias (Beshears & Gino, 2015). One important insight is that a more deliberate and thorough analysis of situations helps individuals control their tendency to rely on stereotypes or other faulty generalizations in judging an individual's performance on a particular task. Bias awareness could help trigger a more deliberate analysis to overcome biased evaluations of male versus female academics. Even though students cannot "blind" themselves to the gender of the instructor, they may mentally simulate counterfactual scenarios. For example, they may consider evaluations they would have given if the same content was delivered by an academic of a different gender, examine the relevance of gender as a factor in their evaluations, and correct their evaluations accordingly. In fact, taking control over tacitly learned reactions to various stimuli in our daily environments, and developing skills of speculation, testing, and generalization has been advocated as a way to "educate" intuitive judgment and overcome biases (Hogarth, 2001). Those who are aware of gender bias, may be in a better position to revise their judgment to more accurately reflect the quality of teaching and stray away from the considerations of the academic's gender-role congruity. Consistent with this argument, a field experiment finds that a factual awareness of the gender bias in past evaluations of similar students leads to a reduction in gender bias (Boring & Arnaud, 2017). In the same study, there is a null effect of being merely reminded that one should not discriminate against female academics in teaching evaluations. Consequently,

***Hypothesis 7: (effect of bias awareness):*** Those who are aware of the gender bias in teaching evaluations favouring male academics will be less likely to show gender bias in their teaching evaluations (warmth, competence, and hiring recommendations).

**Overview of studies**

We test our research hypotheses in two experimental studies. Study 1 tests for gender bias in the context of a male-dominated discipline (astronomy). We then examine evaluations of warmth, competence and hiring recommendations relative to male versus female academics who deliver the lecture in either a teaching style that is high or low on warmth.

In Study 2, we use the "low warmth" version of the same experimental materials to test the de-biasing effect of seniority. As in Study 1, we examine evaluations of the academic's warmth, competence, and hiring recommendations. In addition, we elicit intuitions regarding a possible bias, and examine how bias awareness affects teaching evaluations.

## STUDY 1

**Participants and Design**

We recruited 479 participants on the Prolific.com academic website ($M_{age}$ = 24.07, $SD_{age}$ = 3.17, 50.2% female, 61.2% with undergraduate or postgraduate degrees) for a study that asked them to assess a lecture by a candidate in the academic job market, and provide a hiring recommendation to the university. Participants were restricted to between 18 and 30 years old due to the teaching evaluation context of the study. All participants were from countries with female representation of less than 20% in physics departments. They were paid £1.40 for completing a 10 minute study (average completion time was 8 minutes 35 seconds). Data were gathered during September 2017.

The study consisted of a 2 (gender: male vs. female) x 2 (warmth: high vs. low) between-subjects design. Participants were randomly assigned to one of the four conditions. The number of participants required for the study was determined based on a-priori power analysis with anticipated small effect sizes (i.e., Cohen's $f = .15$; (Cohen, 1992)) which would require a sample size of 460 to be powered at 90%. All power calculations were conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

**Materials**

Participants read an astronomy lecture of around 900 words. The lecture was based on Professor Stephen Hawking's first Reith Lecture entitled "Do Black Holes Have No Hair?" (Hawking, 2016). In the version of the lecture which was high on warmth, the candidate appeared warm and accessible as a teacher. In the version of the lecture which was low on warmth, the candidate appeared to be cold and patronizing. A silhouette of either a male or female head, together with the academic's name (Steve Smith versus Sue Smith), was shown on each of the five screens of the lecture text to reinforce the salience of the academic's gender.

*Pilot study*: We conducted a pilot study to test whether the teaching context (astronomy lecture) was perceived as male-dominated and whether the high warmth version of the lecture was perceived as warmer than the low warmth version. Twenty one individuals ($M_{age} = 24.33$, $SD_{age} = 3.02$, 16 males) participated in this pilot study for a payment of £1.40. For the first test, the academic was described in gender-neutral terms (surname only without a silhouette) and participants rated how likely it was that the academic was male on a 5-point Likert-type scale anchored by 1 (definitely male) to 5 (definitely female). The result, compared to the middle of the scale, confirmed that the astronomy lecture was perceived as male-dominated ($t(21) = 1.92$, $p < .05$). For the second test, participants rated the academic's warmth on a 5-point Likert-type scale anchored by 1 (not at all) to 5 (very) and the high

warmth version was rated higher ($t(21) = 2.02$, $p < .05$) confirming our manipulation of teaching style.

## Procedure

Participants were randomly assigned to each of the four experimental conditions, and proceeded to read the astronomy lecture. Following the lecture they assessed the academic candidate in terms of warmth and competence, and provided a hiring recommendation. The survey finished with socio-demographic questions about the participants.

## Measures

**Warmth.** Participants were asked to assess the academic's warmth using the items "warm" and "accessible" (Fiske, Cuddy, Glick, & Xu, 2002). Participants had to consider the above adjectives and indicate the extent to which they believed the candidate to be each of these things on a 5-point Likert-type scales anchored by 1 (not at all) to 5 (very). The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach's $\alpha = .76$).

**Competence.** Participants were asked to assess the academic's competence using the items "professional" and "knowledgeable" (Fiske et al., 2002). Participants had to consider the above adjectives and indicate the extent to which they believed the candidate to be each of these things on a 5-point Likert-type scales anchored by 1 (not at all) to 5 (very). The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's $\alpha = .70$).

**Hiring recommendation.** Participants were asked whether the candidate who had given the lecture should be hired on a 5-point Likert-type scales anchored by 1 (definitely reject) to 5 (definitely hire).

**Control variables**. We controlled for age, gender, level of education, student status,

and cross-cultural differences operationalized as the World Economic Forum's Global Gender Gap Index for 2016 for the country of birth of each participant (World Economic Forum, 2016).

**Manipulation and attention checks.** At the conclusion of the study, participants were asked to indicate the gender of the academic that they had evaluated. They were also presented with an attention check question telling them to complete an answer with a Likert value of 1. A further check was made on outlying survey completion time of less than one standard deviation from the mean (3 minutes 46 seconds). As a result of these checks, a total of 7 participants (1.5%) were excluded from all subsequent analysis.

## Results

Descriptive statistics for all study variables are given in Table 1.

---

Insert Table 1 about here

---

Our manipulation of the teaching style worked as expected. The high warmth lecture was rated more highly on warmth than the low warmth lecture (M = 3.97, SD = 0.68 versus M = 3.53, SD = 0.80, $t(477) = 6.51$, $p < .001$). The manipulation of the candidate's gender was also successful. 94% of participants in the male condition remembered the academic delivering the lecture as male ($t$ (235) = 29.90, $p < .01$ compared to 50%), and 90% in the female condition remembered the academic as female ($t(242) = 20.92$, $p < .01$ compared to 50%).

To test Hypothesis 1, we conducted multiple regression analysis with the hiring recommendation as the dependent variable. The independent variables were dummies for the gender of the candidate (gender: 1=male, 0=female), the teaching style (warmth: 1=high warmth, 0=low warmth), and their interaction (gender × warmth). All control variables were

included (see Table 2, column 1). Consistent with Hypothesis 1, we found a gender bias: male academics were more likely to be recommended for hiring than their female peers ($\beta$ = .21, $p < .05$). The bias held under both high and low warmth teaching style. The effect size for the gender bias was small, and it did not differ substantially across the low warmth and high warmth scenarios (Cohen's D of -.27 and -.19 respectively).

---

Insert Table 2 about here

---

To test Hypothesis 2 (mediation), we analysed whether warmth and competence acted as mediators between the gender of the academic and the hiring recommendation (Kenny, 2016). Without controlling for warmth and competence, the hiring recommendation was correlated with gender (Table 2, column 1). For a low warmth teaching style, the indirect (mediated) effects of gender through warmth and competence were significant (standardized path coefficient = .10, $p < .05$) whilst the direct effect of gender lost significance (standardized path coefficient = .04, *ns*) (see Table 3, row 1). For a high warmth teaching style, the indirect (mediated) effects of gender through warmth and competence were significant (standardized path coefficient = .08, $p < .05$) whilst the direct effect of gender lost significance (standardized path coefficient = .00, *ns*) (see Table 3, row 2). The indirect effects of gender on the hiring recommendation remained significant in both scenarios when bootstrapping standard errors to allow for kurtosis (Preacher & Hayes, 2008). Thus we found support for Hypothesis 2 that gender bias in hiring recommendations was mediated by warmth and competence perceptions.

---

Insert Table 3 about here

---

Contrary to Hypothesis 3 but consistent with the competing Hypothesis 3A (effect of style on warmth and competence), we found that the high warmth style had different effects on the evaluations of warmth and competence of academics depending on their gender. For warmth, the high warmth style led to more positive evaluations of warmth for female academics, and the effect was larger than the same effect for male academics. We conducted the regression analysis with warmth as the dependent variable and the academic's gender, teaching style, and the interaction between the two as independent variables (see Table 2, column 2). The main effect of male gender was positive and significant ($\beta$ = .34, $p$ < .01) qualified by a negative and significant interaction term ($\beta$ = -.27, $p$ < .05). As for competence, the high warmth style led to somewhat more negative evaluations of competence for female versus male academics. We conducted the regression analysis with competence as the dependent variable and the academic's gender, teaching style, and the interaction between the two as independent variables (see Table 2, column 3). The interaction term was correctly signed but failed to reach statistical significance ($\beta$ = -.18, *ns*). We further examined the simple slopes for the effect of teaching style on competence evaluations depending on gender (Aiken, West, & Reno, 1991). Our results showed there was a statistically significant decrease in the evaluations of competence for women when they taught in a high warmth style ($\beta$ = -.19, $p$ < .05), but not for men ($\beta$ = -.01, *ns*) (see Figure 1).

---

Insert Figure 1 about here

---

Overall, our results show that gender bias persisted in the hiring recommendation in the high warmth scenario because of lower competence evaluations for female academics.

## Discussion

Study 1 showed that in a male-dominated discipline the delivery of the same teaching content led to greater hiring recommendations for male rather than female academics, irrespective of whether the style of delivery was low or high on warmth. Female academics benefited more than male academics from teaching in a style that was high on warmth (as opposed to low on warmth) in terms of evaluations of their warmth. However, they also suffered a somewhat greater penalty in terms of evaluations of their competence, which led to lower hiring recommendations.

# STUDY 2

In Study 2 we tested the de-biasing role of seniority. To date, many empirical studies of gender bias in teaching evaluations examine junior academics (Boring, 2017; MacNell et al., 2015), and a recent field study finds stronger effects of gender for junior as opposed to more senior academics (Mengel et al., 2017). In Study 2, we distinguished deliberately between junior (post-PhD) and senior (Professor Level) academics to test the debiasing role of seniority.

## Participants and Design

We recruited a further 478 participants on Prolific.com ($M_{age}$ = 24.40, $SD_{age}$ = 3.24, 49.9% female, 64.6% with undergraduate or postgraduate degrees) for a study that asked them to assess a lecture by a candidate in the academic job market, and provide a hiring recommendation to the university. Participants were restricted to between 18 and 30 years old and were paid £1.40 for completing a 10 minute study (average completion time was 9 minutes). Data were gathered during November 2017.

The study consisted of a 2 (gender: male vs. female) x 2 (seniority: Professor vs. junior) between subjects design. Participants were randomly assigned to one of the four

conditions. The number of participants required for the study was determined as in Study 1 to be powered at 90% with small effect sizes.

**Materials**

Participants read the low warmth version of the astronomy lecture used in Study 1. The academic was described as a post-PhD male/female candidate or as a Professor male/female candidate. A silhouette of either a male or a female head was shown on each of five screens of text to reinforce the gender manipulation. In addition, depending on the experimental condition, each screen showed the post-PhD candidate's name without the use of any titles, or the senior academic's name used next to the "Professor" title (e.g., Sue Smith versus Professor Sue Smith).

**Procedure**

Participants were randomly assigned to each of the four experimental conditions, and proceeded to read the astronomy lecture. Following the lecture they assessed the academic candidate in terms of warmth and competence, and provided a hiring recommendation. The survey finished with questions about gender bias and socio-demographic questions.

**Measures**

**Warmth.** Participants were asked to assess the academic's warmth using the items "warm" and "accessible" on a 5-point Likert-type scale as in Study 1. The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach's $\alpha$ = .77).

**Competence.** Participants were asked to assess the academic's competence using the items "professional" and "knowledgeable" on a 5-point Likert-type scale as in Study 1. The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's $\alpha$ = .67).

**Hiring recommendation.** Participants were asked whether the candidate who had given the lecture should be hired on a 5-point Likert-type scale as in Study 1.

**Bias awareness.** Following the survey questions, participants were asked whether they thought there is a male, female or no bias in evaluations of warmth, competence and the hiring recommendation generally. The order of the questions was randomized and we coded for bias awareness as 1 if participants believed in a male bias and 0 if participants did not believe in any bias or believed in a female bias (1 = Bias Aware, 0 = Not Aware).

**Control variables.** We controlled for age, gender, level of education, student status, and cross-cultural differences in the gender gap as in Study 1.

**Manipulation and attention checks.** Participants were asked the gender of the academic that they had evaluated. A total of 6 participants (1.25%) failed the manipulation check and further checks on outlying survey completion times less than one standard deviation from the mean (3 minutes 30 seconds). They were excluded from all subsequent analysis.

**Results**

Descriptive statistics for all study variables are given in Table 4.

---

Insert Table 4 about here

---

To test Hypothesis 5, we used multiple linear regressions of the hiring recommendation on dummies for the gender of the candidate (gender: 1 = male, 0 = female), the seniority of the candidate (seniority: 1 = Professor, 0 = Junior), and their interaction (gender × seniority) (see Table 5, column 1).

---

Insert Table 5 about here

---

There was no significant gender bias at Professor Level for the hiring recommendation ($\beta$ = -.03, *ns*, effect size Cohen's D = .05) supporting Hypothesis 5. Simple slopes analysis showed a significant improvement in the hiring recommendation at Professor Level, compared to junior levels, for female academics with little change for in the hiring recommendation for male academics ($\beta$ = .23, $p < .05$ for female academics, versus $\beta$ = 0.01, *ns* for male academics, see Figure 2). Gender bias for junior levels was comparable in size to the bias in Study 1 (Cohen's D -.21) but only weakly significant ($\beta$ = .19, $p < .10$). To clarify what our findings implied for testing the existence of a gender bias at junior academic levels, we performed the single paper meta-analysis (SPM) on the estimates of the gender bias using non-parametric tests for junior academics in Study 1 and Study 2 (McShane & Böckenholt, 2017). Pooling the results of Studies 1 and 2, SPM yielded evidence of a significant gender bias at junior level *($z(490) = -2.92, p < .01$,* Cohen's D = -.24, see Table 6). Comparing the gender bias in the hiring recommendation between the two studies ($\beta$ = -.03, *ns*) does not yield a significant difference between the size of the effects found (See Table 7).

---

Insert Table 6 about here

---

---

Insert Table 7 about here

---

As in Study 1, gender bias in the evaluations of the academic's warmth was statistically significant ($\beta$ = .22, $p < .05$) (see Table 5, column 2). The regression of warmth evaluations on dummies for the gender of the candidate (gender: 1 = male, 0 = female), the seniority of the candidate (seniority: 1 = Professor, 0 = Junior), and their interaction (gender

× seniority), included a significant interaction effect ($\beta$ = -.39, $p < .05$). At first sight, this seemed to be consistent with Hypothesis 6. However, simple slopes analysis showed that rather than significantly improving evaluations of warmth for female academics, seniority diminished the evaluations of warmth for male academics ($\beta$ = .15, *ns* for female academics, versus $\beta$ = .-24, $p < .05$ for male academics, see Figure 2). This was unexpected, and contrary to the rationale of Hypothesis 6 which predicted higher warmth evaluations for senior female academics (due to the fact that they overcame double standards) without any drop in the warmth evaluations of male academics.

For competence, seniority affected the evaluation of female and male academics the same (see Table 5, column 3).

---

Insert Figure 2 about here

---

As for bias awareness, we found that it was generally low. Irrespective of whether respondents considered warmth, competence or hiring recommendations, roughly 70% believed no gender bias existed. Participants who believed in a female advantage were most numerous when it came to warmth evaluations (24% versus 2% when competence was considered, and 5% when hiring recommendation was considered). To analyse whether awareness of gender bias helped participants correct their evaluations, we performed regressions of warmth, competence and the hiring recommendation on all independent and control variables from our previous analyses, adding the variable "bias aware" and the interaction between "bias aware" and the dummy for the academic's gender to the analysis (see Table 8). The coefficients for bias awareness and the interaction of bias awareness and the gender dummy were not significant in any of the regressions. Contrary to Hypothesis 7, we did not find that being aware of a male bias de-biases teaching evaluations.

---

Insert Table 8 about here

---

**Discussion**

In Study 2, we found that the gender bias against junior female academics, in warmth and in hiring recommendations, disappeared with seniority. The finding of a significant gender bias for junior academics proved wrong roughly 70% of respondents who considered that gender bias was not a factor in teaching evaluations and hiring.

Moreover, there was an unexpected bias against senior male academics such that their warmth evaluations diminished and became inferior to those of senior female academics while the latter did not improve in comparison to junior female academics. This pattern of results suggested a mechanism for the elimination of gender bias that we did not initially foresee. At senior levels, male academics seem to have lost the advantage that drove their hiring recommendations when academics were portrayed as juniors. The mechanism for this effect should be tested in future research.

Importantly, those who reported being aware of the bias did not show more accurate evaluations of the candidates on either warmth, competence, or hiring recommendations.

## GENERAL DISCUSSION

In two controlled experiments we showed evidence of gender bias in teaching evaluations and hiring recommendations for junior academics. These results are consistent with the predictions of the gender-role congruity theory (Cejka & Eagly, 1999; Eagly & Karau, 2002) and previous empirical findings (MacNell et al., 2015; Mengel et al., 2017; Ottoboni et al., 2016). Our work provides a more fine-grained understanding of the workings of gender bias and points to the settings in which the bias disappears, representing a novel

contribution to the existing literature. We discuss below the important practical and theoretical implications of our work.

**Practical Implications**

We submitted to an empirical test the idea that the academic's teaching style may matter for the magnitude of the bias (Mitchell & Martin, 2018). In particular, the hope was that a female-stereotypic high warmth style may shield female academics from lower evaluations of the same teaching content. Yet, our results provided support to a competing hypothesis whereby a high warmth teaching style brought a competence penalty which led to lower hiring recommendations for female academics despite the fact that evaluations of their warmth improved to the level of the evaluations for their male peers. These results may be specific to the higher education setting because there are important information asymmetries between students and academics in understanding the subject matter and hence, evaluating the competence of the academic. Because student evaluations of the academic's competence are therefore uncertain, they are particularly likely to be vulnerable to bias. Hence, whereas in other settings, a competent performance by a woman would be perceived more positively if the woman adopted a high warmth style (Carli, 2001), in the academic setting, her high warmth style triggered a greater reliance on the gender stereotype and considerations of gender-role congruity, exerting a downward pressure on the evaluations of her competence and the hiring recommendation.

We showed that the gender bias is sensitive to seniority and disappears for professors as opposed to junior academics in Study 2. This result supports recent calls in the literature to shield academics from decisions based on teaching evaluations, and qualifies it by the importance of doing so at least in the early stages of their careers. With seniority, the female academic's title begins to pave the way for her to be assessed on par with her male peer for equal performance. Unexpectedly, seniority produced a negative effect for warmth

evaluations of male academics. On the one hand, this suggests that male professors may stand nothing to gain from showcasing their senior status. On the other hand, it is important to understand the underlying mechanism for this effect. It may be that at junior levels, male academics experience an unfair advantage (rather than female academics experiencing a disadvantage), which is corrected at senior levels. However, it may also be that at junior levels, female academics are subjected to stereotype-driven unfair disadvantage (as argued in this paper) whereas at senior levels, a seniority-related stereotype produces a similar disadvantage for male academics. It remains to be investigated in future research which of these mechanisms applies and whether our findings are specific to academia or generalize beyond the higher education context.

Finally, we examined bias awareness among the very people who evaluated a given teaching content in our experiments, and tested the idea that bias awareness leads to less biased teaching evaluations. It was informative to find that the vast majority of our study participants did not believe that gender played a role in teaching evaluations. Regrettably, those who anticipated the gender bias failed to correct for it in their own teaching evaluations. Although many organizations may rely on building awareness about the gender bias as the bias mitigation strategy, this result suggests caution in relying on that kind of intervention alone without other forms of career support to junior female academics.

**Theoretical Implications**

Our work makes theoretical contributions and opens promising avenues for future research. First, we show that gender biases may benefit from a systematic study in credence versus non-credence settings (Darby & Karni, 1973; Gruber & Frugone, 2011; Kasnakoglu, 2016). Our prediction is that when violation of a misaligned (e.g. high warmth-low competence) stereotype on one of its dimensions affects overall performance evaluations,

behaving in a stereotype-consistent manner will be beneficial if the other dimension is non credence-based, and may not be beneficial if the other dimension is credence-based. In the latter case, behaving in a stereotype-consistent manner may simply reinforce the stereotype.

Second, most of the literature on gender bias in academia focuses on a female disadvantage (Carli, 2001). In contrast, we found evidence of a male disadvantage for senior academics in a male-dominated discipline. This finding merits further research attention. In fact, the field as a whole could benefit from a more thorough understanding of all the explanatory mechanisms behind gender biases that produce either male or female disadvantages. To date, the literature has generated an impressive list of possible mechanisms, pointing to the role of the considerations of "double standards" (Rubin, 1981; Winocur et al., 1989), the ease with which people come up with upward versus downward counterfactuals and the role of expectations (Epstude & Roese, 2008), halo effects (Kaplan, 1978; Landy & Sigall, 1974) and contrast effects (Moskowitz, 2005, pp. 388-437). Yet any combination of these may be particularly likely in a given setting or, as a function of an individual's particular characteristic, and we need to be able to understand the net effect.

Finally, we focused in our experimental work on the evaluations of the academic along two fundamental dimensions of social cognition (warmth and competence) which impact the career decisions of individual academics (Fiske et al., 2007). Yet, it is also instructive to shed light on the role that gender biases may play for the evaluations of learning outcomes and taught content which impact decisions regarding academic institutions. At the level of academic institutions, this may help further inform the impact of initiatives that aim to reduce possible gender biases and provide impetus for more active research on bias-reducing interventions.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions* (1st ed.). London: Sage.

Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*(9-10), 507-516. doi:http://dx.doi.org/10.1023/A:1025832707002

Bassett, J., Cleveland, A., Acorn, D., Nix, M., & Snyder, T. (2017). Are they paying attention? students' lack of motivation and attention potentially threaten the utility of course evaluations. *Assessment & Evaluation in Higher Education, 42*(3), 431-442. doi:http://dx.doi.org/10.1080/02602938.2015.1119801

Beshears, J., & Gino, F. (2015). Leaders as decision architects. *Harvard Business Review, 93*(5), 52-62. doi:http://dx.doi.org/10.1109/EMR.2016.7559057

Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *The American Economic Review, 95*(2), 152-157. doi:http://dx.doi.org/10.1257/000282805774670149

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41. doi:http://dx.doi.org/10.1016/j.jpubeco.2016.11.006

Boring, A., & Arnaud, P. (2017). De-biasing performance evaluations: Evidence from a field experiment on student evaluations of teaching. Retrieved from http://ices.gmu.edu/wp-content/uploads/2017/08/De-biasing-Biased-Performance-Evaluations-Evidence-from-a-Field-Experiment-on-Student-Evaluations-of-Teaching-Abstract-by-Boring-and-Philippe.pdf

Byron, K. (2007). Male and female managers' ability to read emotions: Relationships with supervisor's performance ratings and subordinates' satisfaction ratings. *Journal of Occupational and Organizational Psychology, 80*(4), 713-733. doi:http://dx.doi.org/10.1348/096317907X174349

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues, 57*(4), 725-741. doi:http://dx.doi.org/10.1111/0022-4537.00238

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics, 125*(3), 1101-1144. doi:http://dx.doi.org/10.1162/qjec.2010.125.3.1101

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*(4), 413-423. doi:http://dx.doi.org/10.1177/0146167299025004002

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi:http://dx.doi.org/10.1037/0033-2909.112.1.155

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608-630. doi:http://dx.doi.org/10.1037/0033-295X.96.4.608

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics, 16*(1), 67-88. doi:http://dx.doi.org/10.1086/466756

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*(2), 225-248. doi:http://dx.doi.org/10.1006/jvbe.1999.1711

Dick, P., & Nadin, S. (2006). Reproducing gender inequalities? A critique of realist assumptions underpinning personnel selection research and practice. *Journal of Occupational and Organizational Psychology, 79*(3), 481-498. doi:http://dx.doi.org/10.1348/096317905X68709

Doubleday, A. F., & Lee, L. M. (2016). Dissecting the voice: Health professions students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty. *Anatomical Sciences Education, 9*(6), 537-544. doi:http://dx.doi.org/10.1002/ase.1609

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573-598. doi:http://dx.doi.org/10.1037/0033-295X.109.3.573

Emmerik, H., Wendt, H., & Euwema, M. C. (2010). Gender ratio, societal culture, and male and female leadership. *Journal of Occupational and Organizational Psychology, 83*(4), 895-914. doi:http://dx.doi.org/10.1348/096317909X478548

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*(2), 168-192. doi:http://dx.doi.org/10.1177/1088868308316091

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:http://dx.doi.org/10.3758/BF03193146

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83. doi:http://dx.doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902. doi:http://dx.doi.org/10.1037/0022-3514.82.6.878

Gruber, T., & Frugone, F. (2011). Uncovering the desired qualities and behaviours of general practitioners (GPs) during medical (service recovery) encounters. *Journal of Service Management, 22*(4), 491-521. doi:http://dx.doi.org/10.1108/09564231111155097

Guillén, L., Mayo, M., & Karelaia, N. (2017). Appearing self-confident and getting credit for it: Why it may be easier for men than women to gain influence at work. *Human Resource Management, ,* 1-16. doi:http://dx.doi.org/10.1002/hrm.21857

Hannover Research Council. (2009). Best practices in student course evaluation. Retrieved from http://www.planning.salford.ac.uk/__data/assets/pdf_file/0015/20760/Best-Practices-in-Student-Course-Evaluation-Membership.pdf

Hawking, S. (2016). *Black holes: The reith lectures* Random House.

Hogarth, R. M. (2001). *Educating intuition* University of Chicago Press.

Kaplan, R. M. (1978). Is beauty talent? sex interaction in the attractiveness halo effect. *Sex Roles, 4*(2), 195-204. doi:http://dx.doi.org/10.1007/BF00287500

Kasnakoglu, B. (2016). Antecedents and consequences of co-creation in credence-based service contexts. *The Service Industries Journal, 36*(1-2), 1-20. doi:http://dx.doi.org/10.1080/02642069.2016.1138472

Kenny, D. A. (2016, ). Power analsis app MedPower. learn how you can do a mediation analysis and output a text description of your results: Go to mediational analysis using DataToText using SPSS or R. Retrieved from http://www.davidakenny.net/cm/mediate.htm

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100*(1), 128-161. doi:http://dx.doi.org/10.1037/a0036734

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*(4), 522-544. doi:http://dx.doi.org/10.1037/0033-2909.129.4.522

Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*(3), 299-304. doi:http://dx.doi.org/10.1037/h0036018

Langbein, L. I. (1994). The validity of student evaluations of teaching. *PS: Political Science & Politics, 27*(3), 545-553. doi:http://dx.doi.org/10.2307/420225

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73*(1), 91-103. doi:http://dx.doi.org/10.1037/0022-3514.73.1.91

Lyness, K. S., & Thompson, D. E. (2000). Climbing the corporate ladder: Do female and male executives follow the same route? *Journal of Applied Psychology, 85*(1), 86-101. doi:http://dx.doi.org/10.1037/0021-9010.85.1.86

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*(4), 291-303. doi:http://dx.doi.org/10.1007/s10755-014-9313-4

McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research, 43*(6), 1048-1063. doi:http://dx.doi.org/10.1093/jcr/ucw085

Mengel, F., Sauermann, J., & Zölitz, U. (2017, ). Gender bias in teaching evaluations. Retrieved from http://hdl.handle.net/10419/170984

Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, , 1-5. doi:https://doi.org/10.1017/S104909651800001X

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others* (1st ed.). New York: Guilford Press.

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

Newsome, J. L. (2008). *The chemistry PhD: The impact on women's retention.* ().

Ottoboni, K., Boring, A., & Stark, P. (2016, ). Student evaluations of teaching (mostly) do not measure teaching effectiveness. Retrieved from https://blendedtoolkit.wisc.edu/wp-content/uploads/2016/11/Science-Open_Stark-et-al.pdf

Pinto, M. B., & Mansfield, P. M. (2010). Thought processes college students use when evaluating faculty: A qualitative study. *American Journal of Business Education (AJBE), 3*(3), 55-62. doi:http://dx.doi.org/10.19030/ajbe.v3i3.399

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? an analytical framework for answering the question. *Quality Assurance in Education*, *15*(2), 178-191. doi:http://dx.doi.org/10.1108/09684880710748938

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891. doi:http://dx.doi.org/10.3758/BRM.40.3.879

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, *95*(2), 221-235. doi:http://dx.doi.org/10.1037/a0018204

Rubin, R. B. (1981). Ideal traits and terms of address for male and female college professors. *Journal of Personality and Social Psychology, 41*(5), 966-988. doi:http://dx.doi.org/10.1037/0022-3514.41.5.966

Schrodt, P., Witt, P. L., & Turman, P. D. (2007). Reconsidering the measurement of teacher power use in the college classroom. *Communication Education*, *56*(3), 308-332. doi:http://dx.doi.org/10.1080/03634520701256062

Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences of the United States of America, 111*(28), 10107-10112. doi:http://dx.doi.org/10.1073/pnas.1403334111

Shen, A. (2015, ). Students see male professors as brilliant geniuses, female professors as bossy And Annoying. Retrieved from https://thinkprogress.org/students-see-male-professors-as-brilliant-geniuses-female-professors-as-bossy-and-annoying-5dd018d5a785#.5t2dzr2v0

Wagner, N., Rieger, M., & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *ISS Working Paper Series/General Series, 617*(617), 1-32. doi:http://dx.doi.org/10.1016/j.econedurev.2016.06.004

Way, S. F., Larremore, D. B. & Clauset, A. (2016). Gender, productivity, and prestige in computer science faculty hiring networks. Retrieved from https://arxiv.org/pdf/1602.00795.pdf;

Wild, C., & Berger, D. (2016). The proposed teaching excellence framework (TEF) for UK universities. *International Journal of Teaching and Education, 4*(3), 33-50. doi:http://dx.doi.org/10.20472/TE.2016.4.3.004

Winocur, S., Schoen, L. G., & Sirowatka, A. H. (1989). Perceptions of male and female academics within a teaching context. *Research in Higher Education, 30*(3), 317-329. doi:http://dx.doi.org/10.1007/BF00992607

World Economic Forum. (2016). World economic forum 2016 global gender gap report, cologny, switzerland . Retrieved from http://www3.weforum.org/docs/GGGR16/WEF_Global_Gender_Gap_Report_2016.pdf

Young, S., Rush, L., & Shaw, D. (2009). Evaluating gender bias in ratings of university instructors' teaching effectiveness. *International Journal for the Scholarship of Teaching and Learning, 3*(2, Article 19), 1-14. doi:http://dx.doi.org/10.20429/ijsotl.2009.030219

TABLE 1

*Means (and standard deviations in brackets) for Study 1 variables*

| | Low Warmth Scenario | | High Warmth Scenario | |
|---|---|---|---|---|
| | Female | Male | Female | Male |
| Warmth | 3.36 | 3.70 | 3.94 | 4.00 |
| | (0.78) | (0.79) | (0.68) | (0.69) |
| Competence | 4.37 | 4.36 | 4.17 | 4.34 |
| | (0.55) | (0.56) | (0.67) | (0.59) |
| Hiring | 3.81 | 4.02 | 3.97 | 4.13 |
| | (0.85) | (0.75) | (0.85) | (0.73) |
| Control Variables | | | | |
| Male Respondent | 0.51 | 0.43 | 0.55 | 0.49 |
| | (0.50) | (0.50) | (0.50) | (0.50) |
| Graduate Degree | 0.46 | 0.43 | 0.46 | 0.47 |
| | (0.50) | (0.50) | (0.50) | (0.50) |
| Postgraduate Degree | 0.16 | 0.19 | 0.19 | 0.14 |
| | (0.37) | (0.39) | (0.39) | (0.35) |
| Over 24 years old | 0.46 | 0.46 | 0.44 | 0.48 |
| | (0.50) | (0.50) | (0.50) | (0.50) |
| Gender Index | 0.54 | 0.57 | 0.52 | 0.59 |
| | (0.50) | (0.50) | (0.50) | (0.49) |
| Observations | 125 | 127 | 118 | 109 |

TABLE 2

*Regressions for Hiring Recommendation, Warmth and Competence in Study 1*

| Variables | Hiring | Warmth | Competence |
|---|---|---|---|
| Male Academic | 0.21** | 0.34*** | -0.02 |
| | (0.10) | (0.09) | (0.08) |
| High Warmth Scenario | 0.17 | 0.58*** | -0.19** |
| | (0.10) | (0.09) | (0.08) |
| Male Academic x High Warmth Scenario | -0.06 | -0.27** | 0.18 |
| | (0.15) | (0.14) | (0.11) |
| Male Respondent | 0.03 | 0.06 | -0.08 |
| | (0.08) | (0.07) | (0.06) |
| Graduate Degree | -0.02 | 0.00 | -0.05 |
| | (0.08) | (0.08) | (0.06) |
| Postgraduate Degree | 0.04 | 0.20* | 0.02 |
| | (0.11) | (0.10) | (0.08) |
| Over 24 years old | 0.04 | -0.06 | 0.04 |
| | (0.08) | (0.07) | (0.06) |
| Gender Index | 0.06 | 0.09 | 0.04 |
| | (0.08) | (0.07) | (0.06) |
| Constant | 3.74*** | 3.27*** | 4.39*** |
| | (0.11) | (0.10) | (0.08) |
| Observations | 479 | 479 | 479 |
| R-squared | 0.02 | 0.12 | 0.03 |

*Note*: * = p < .10, ** = p < .05, *** = p < .01

TABLE 3

*Standardized Mediation Effects for Warmth and Competence on the Hiring Recommendation in Study 1.*

| | DV = Hiring Recommendation | | | | |
|---|---|---|---|---|---|
| | | Indirect effect of IV on DV | | | |
| Teaching Style | Direct Effect of IV on DV | Via Warmth | Via Competence | Total | Total Effect IV on DV |
| Gender - Low Warmth | .04 | **.10** | .00 | **.10** | **.14** |
| Gender - High Warmth | .00 | .01 | **.07** | **.08** | .08 |

*Note*: Low Warmth N=252, High Warmth N=227, DV = dependent variable, IV = independent variable (Gender),

Values in **bold** = significant.

TABLE 4
*Means (and standard deviations in brackets) for Study 2 variables*

| | Junior Level | | Professor Level | |
|---|---|---|---|---|
| | Female | Male | Female | Male |
| Warmth | 3.38 | 3.62 | 3.55 | 3.37 |
| | (0.78) | (0.78) | (0.91) | (0.83) |
| Competence | 4.30 | 4.25 | 4.39 | 4.35 |
| | (0.59) | (0.57) | (0.52) | (0.57) |
| Hiring | 3.83 | 4.02 | 4.06 | 4.02 |
| | (0.86) | (0.89) | (0.83) | (0.76) |
| Control Variables | | | | |
| Male Respondent | 0.40 | 0.55 | 0.58 | 0.47 |
| | (0.49) | (0.50) | (0.50) | (0.50) |
| Graduate Degree | 0.50 | 0.48 | 0.49 | 0.51 |
| | (0.50) | (0.50) | (0.50) | (0.50) |
| Postgraduate Degree | 0.13 | 0.16 | 0.18 | 0.14 |
| | (0.33) | (0.37) | (0.38) | (0.35) |
| Over 24 years old | 0.48 | 0.54 | 0.46 | 0.59 |
| | (0.50) | (0.50) | (0.50) | (0.49) |
| Gender Index | 0.57 | 0.55 | 0.51 | 0.60 |
| | (0.50) | (0.50) | (0.50) | (0.49) |
| Observations | 118 | 126 | 130 | 104 |

TABLE 5
*Regressions for Warmth, Competence and Hiring Recommendation in Study 2*

| Variables | Hiring | Warmth | Competence |
|---|---|---|---|
| Male Academic | 0.19* | 0.22** | -0.04 |
| | (0.11) | (0.11) | (0.07) |
| Professor Level | 0.23** | 0.15 | 0.11 |
| | (0.11) | (0.11) | (0.07) |
| Male Academic x Professor Level | -0.22 | -0.39** | -0.02 |
| | (0.16) | (0.15) | (0.10) |
| Male Respondent | -0.00 | 0.07 | -0.05 |
| | (0.08) | (0.08) | (0.05) |
| Graduate Degree | -0.12 | 0.00 | -0.05 |
| | (0.09) | (0.09) | (0.06) |
| Postgraduate Degree | -0.13 | 0.09 | -0.07 |
| | (0.12) | (0.12) | (0.08) |
| Over 24 years old | -0.05 | 0.02 | 0.03 |
| | (0.08) | (0.08) | (0.05) |
| Gender Index | -0.04 | -0.06 | 0.06 |
| | (0.08) | (0.08) | (0.05) |
| Constant | 3.95*** | 3.36*** | 4.30*** |
| | (0.11) | (0.11) | (0.08) |
| Observations | 478 | 478 | 478 |
| R-squared | 0.02 | 0.02 | 0.02 |

*Note*: * = p < .10, ** = p < .05, *** = p < .01. Low warmth scenario only.

TABLE 6

*Single-Paper Meta-Analysis (SPM) for the Effect of Male Gender*

| Study | t | df | p | z | Cohen's D |
|---|---|---|---|---|---|
| Study 1 | -2.16 | 250 | .03 | -2.19** | -.27 |
| Study 2 | -1.66 | 240 | .10 | -1.95* | -.21 |
| SPM | | | | -2.92*** | -.24 |

*Note*: * = p < .10, ** = p < .05, *** = p < .01

TABLE 7

*Comparison of Male Academic Coefficients in Study 1 and Study 2*

| Variables | Hiring | Warmth | Competence |
|---|---|---|---|
| Male Academic | 0.22** | 0.33*** | 0.01 |
| | (0.10) | (0.10) | (0.07) |
| Study 2 | 0.03 | 0.04 | -0.08 |
| | (0.11) | (0.10) | (0.07) |
| Male Academic x Study 2 | -0.04 | -0.11 | -0.02 |
| | (0.15) | (0.14) | (0.10) |
| Male Respondent | 0.07 | 0.08 | -0.08 |
| | (0.08) | (0.07) | (0.06) |
| Graduate Degree | -0.03 | -0.00 | 0.02 |
| | (0.08) | (0.08) | (0.06) |
| Postgraduate Degree | 0.02 | 0.15 | 0.04 |
| | (0.12) | (0.11) | (0.08) |
| Over 24 years old | -0.10 | -0.11 | -0.03 |
| | (0.08) | (0.07) | (0.06) |
| Gender Index | 0.09 | 0.08 | 0.09 |
| | (0.08) | (0.07) | (0.06) |
| Constant | 3.79*** | 3.31*** | 4.35*** |
| | (0.11) | (0.10) | (0.08) |
| Observations | 494 | 494 | 494 |
| R-squared | 0.02 | 0.04 | 0.02 |

*Note*: * = p < .10, ** = p < .05, *** = p < .01. Low warmth scenario only.

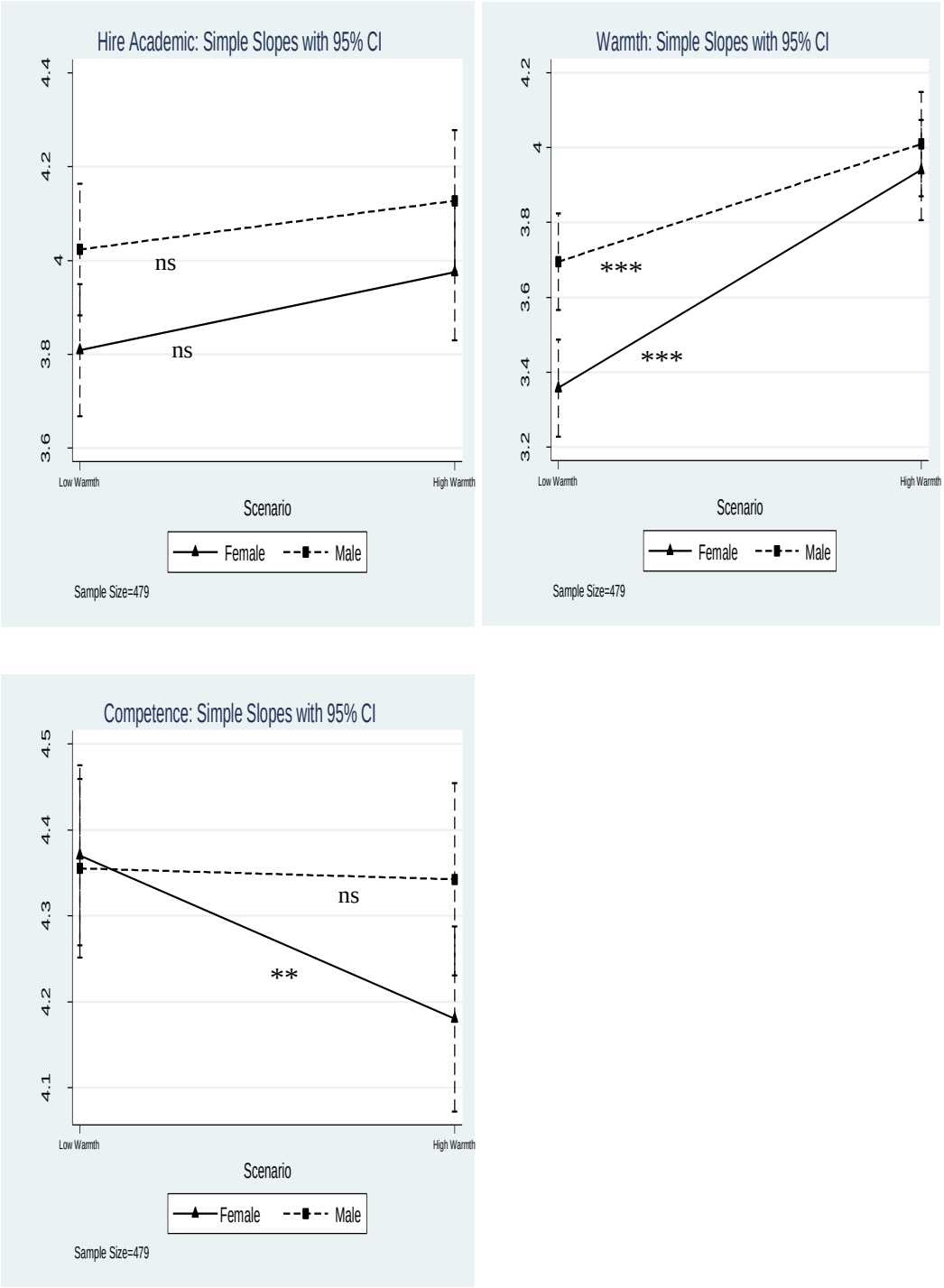
TABLE 8
*Regression Analysis of the Role of Bias Awareness on Gender Bias in Study 2*

| Variables | Hiring | Warmth | Competence |
|---|---|---|---|
| Male Academic | 0.21* | 0.20* | -0.05 |
| | (0.12) | (0.11) | (0.08) |
| Professor Level | 0.23** | 0.15 | 0.11 |
| | (0.11) | (0.11) | (0.07) |
| Male Academic x Professor Level | -0.21 | -0.40** | -0.02 |
| | (0.16) | (0.15) | (0.10) |
| Bias Aware | -0.01 | -0.31 | -0.09 |
| | (0.12) | (0.22) | (0.08) |
| Male Academic x Bias Aware | -0.10 | 0.40 | 0.05 |
| | (0.17) | (0.32) | (0.12) |
| Male Respondent | -0.01 | 0.07 | -0.05 |
| | (0.08) | (0.08) | (0.05) |
| Graduate Degree | -0.12 | 0.00 | -0.05 |
| | (0.09) | (0.09) | (0.06) |
| Postgraduate degree | -0.13 | 0.09 | -0.07 |
| | (0.12) | (0.12) | (0.08) |
| Over 24 years old | -0.04 | 0.02 | 0.03 |
| | (0.08) | (0.08) | (0.05) |
| Gender Index | -0.04 | -0.07 | 0.05 |
| | (0.08) | (0.08) | (0.05) |
| Constant | 3.96*** | 3.38*** | 4.32*** |
| | (0.12) | (0.11) | (0.08) |
| Observations | 478 | 478 | 478 |
| R-squared | 0.02 | 0.03 | 0.02 |

*Note*: * = p < .10, ** = p < .05, *** = p < .01

FIGURE 1

*Study 1 –Simple Slopes Analysis for the Hiring Recommendation, Warmth and Competence.*

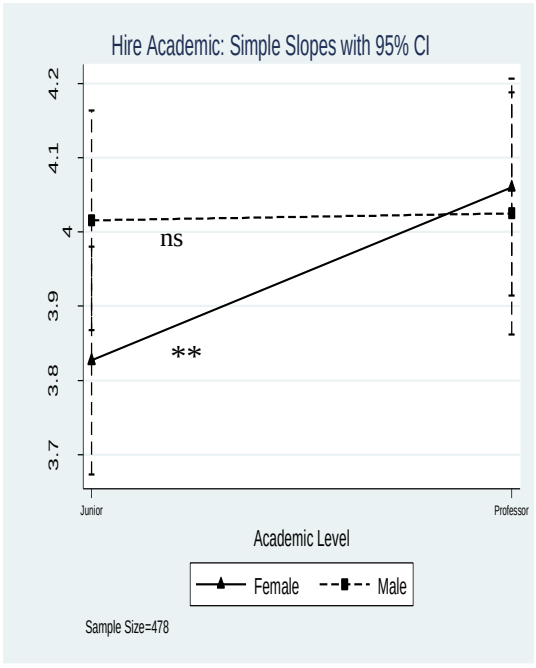


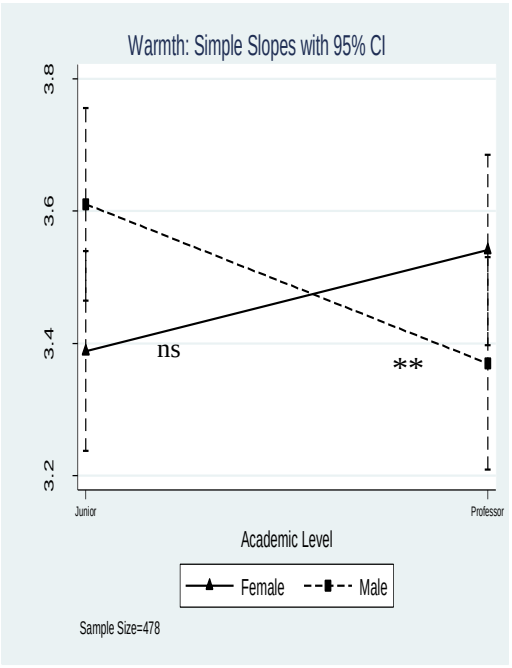

*Note*: 95% confidence interval shown as a dotted line.

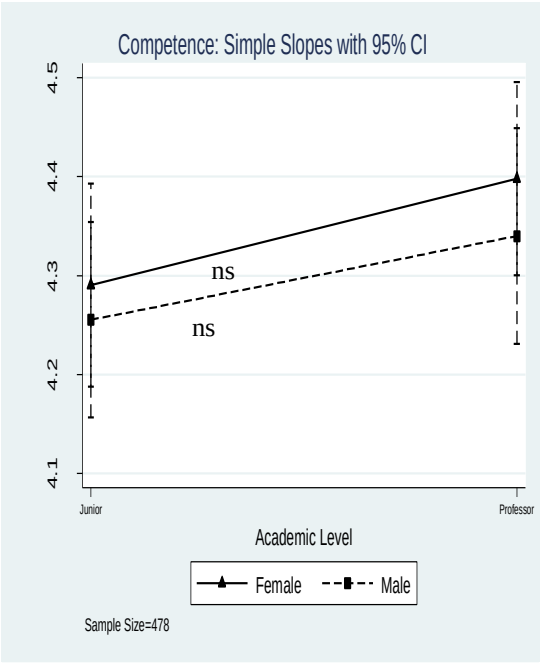
*Note*: * = p < .10, ** = p < .05, *** = p < .01

FIGURE 2

*Study 2 – Simple Slopes Analysis for the Hiring Recommendation, Warmth and Competence.*




*Note*: 95% confidence interval shown as a dotted line.

*Note*: * = p < .10, ** = p < .05, *** = p < .01