Working Papers in Economics & Finance
2020-05

He Taught, She Taught: The effect of teaching style, academic credentials, bias awareness and academic discipline on gender bias in teaching evaluations

Nigel Burnell, University of Surrey
Irina Cojuharenco, University of Surrey
Zahra Murad, University of Portsmouth

Portsmouth Business School
https://www.port.ac.uk/about-us/structure-and-governance/organisational-structure/our-academic-structure/faculty-of-business-and-law/portsmouth-business-school

He Taught, She Taught:

The effect of teaching style, academic credentials, bias awareness and academic discipline 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

23 Apr. 20

Abstract

Gender bias in teaching evaluations leads to unfair decisions during academics' careers. In four controlled experiments, we examine the role of academics' teaching style, academic credentials, academic discipline and bias awareness on gender bias in teaching evaluations. 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 female academics' perceived warmth, but decreases their perceived competence, so gender bias in evaluations persists. In Study 2, we find that gender bias disappears for academic with senior credentials. Additionally, we find no evidence of less biased evaluations by those who anticipate gender bias. In Study 3 and Study 4, we test the robustness of our results in a different academic discipline and using different evaluation measures. In these latter studies, we do not find any evidence of gender bias in evaluations. 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, academic credentials, 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; Boring, Ottoboni, & Stark, 2016; Pounder, 2007; Wagner, Rieger, & Voorvelt, 2016; Young, Rush, & Shaw, 2009; Boring 2017; Mengel, Sauermann & Zolitz 2019). This finding constitutes a major 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-stereotyped disciplines, maybe 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).

Recent field studies using more than nineteen thousand teaching evaluations from major European Universities showed that it is female academics who tend to be evaluated less positively, especially if they are junior, in male-stereotyped disciplines, and rated by male students (Boring 2017; Mengel, et al. 2019). It is worthwhile to note that 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 academic credentials of the candidate. What controlled experiments add to field studies is ruling out of possible differences in between female and male academics in such variables, and a possibility to estimate the effects of each variable in

isolation and interacting with the effect of gender (Arbuckle & Williams, 2003; Doubleday & Lee, 2016; MacNell et al., 2015; Boring et al., 2016; Wagner et al., 2016).

In what follows, we build on the works of literature in management, economics, and education, to formulate testable research hypotheses regarding the role of teaching style, academic credentials and discipline on gender bias in teaching evaluations. We perform four controlled studies testing the hypotheses regarding the effects of academics' teaching style on gender bias and the de-biasing role of academics' credentials. In addition, we survey lay intuitions of experimental participants regarding possible gender bias in teaching evaluations 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 appears.

In Studies 1-2, we find that a female-stereotyped ("warm") teaching style improves perceptions of warmth for female academics but backfires by lowering perceptions of their competence. Hence, academic evaluation in the form of hiring recommendations are lower for female academics (vs. male) irrespective of their teaching style because of a double-bind nature of reactions to their teaching: if female academics' 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, we find that 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 the need to shield junior academics from decisions that rely on teaching evaluations, especially in the early stages of their careers. 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. Finally, those experimental participants aware of gender bias hurting female academics are no more likely to correct their evaluations, suggesting caution in treating bias awareness alone as an effective remedy to the problem (Boring & Philippe 2019).

In preregistered Studies 3-4, we test the robustness of our Study 1-2 results using alternative teaching evaluation measures in conjunction with warmth and competence evaluations. We test the effect of teaching style and academic credentials in different academic disciplines. We find no gender bias in these studies to start with and hence no effect of teaching style and credentials. Most published literature has so far focused and reported only significant gender bias results. We think that it is important to report insignificant gender bias results in the published literature to give more prominence to the possibility that the appearance of the gender bias may be context- and study-dependent.

Hypotheses

The role of teaching style

Academics, like any other employee, 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 lecturer (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-stereotyped 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).

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 the 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.

Hypothesis 1: (gender bias in evaluations) Evaluations will be lower for female than for male academics who teach the same content.

Hypothesis 2: (effect of style on gender bias in evaluations): Gender bias in evaluations will be reduced or eliminated under a teaching style that is high on warmth.

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. We also will test whether gender bias in evaluations is

operating through evaluations of warmth and competence. We hypothesize that the evaluations of warmth and competence will fully explain teaching evaluations and hence the gender bias will disappear when we control for these.

Hypothesis 4: (explaining evaluations) Gender bias in evaluations will be explained by warmth and competence perceptions of the academic.

Although we predict in Hypotheses 2-3 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 academics' 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 female academics, a warm teaching style could decrease competence perceptions whereas no such effect would be expected for male academics. 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 evaluated lower because of their lower perceived competence and fit to the role. Depending on the weight placed on competence versus warmth as determinants of evaluations, the bias may change either upward or downward. Therefore, we will estimate and test the significance of both the direct and indirect effects of academics' gender on their evaluations through both warmth and competence.

The role of academic credentials

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 academic 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-stereotyped 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 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; Beaman, Duflo, Pande & Topalove 2012). This could provide the basis for a female seniority advantage in academia.

Hypothesis 5: (effect of seniority on gender bias on evaluations) 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 warmer and more competent, and more so for female rather than male academics.

In addition to seniority of the academic, that is a proxy of their skill, gender bias may be affected by the academic's background information. Many studies have shown that statistical and taste-based discrimination can coexist and removing statistical discrimination decreases or eliminates discrimination per se (Neilson and Ying 2016; Guryan and Charles 2013). This provides the basis for an additional hypothesis that focuses more directly on studying the effect of academic's credentials in terms of their performance measures on gender bias.

Hypothesis 7: (effect of credentials): Relative to the academics with high credentials, academics with no credentials delivering the same content will be evaluated worse, and more so for female rather than male academics.

Bias awareness

Several 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. 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 & Philippe, 2019). In the same study, there is a null effect of being merely reminded that one should not discriminate against female academics in teaching evaluations.

Hypothesis 8: (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 evaluations.

The role of academic discipline

Given the reliance on gender stereotypes and considerations of gender-role congruity generating and nurturing gender biases, a natural question to ask is whether gender bias is more pronounced in more male-stereotyped than female-stereotyped academic disciplines. Previous literature has mostly focused on gender biases in stereotypically male fields such as life sciences, economics or in work environments using competitive tournament incentive schemes (Sheltzer and Smith 2014; Basow and Silberg 1987; Boring et al. 2016). In such male-stereotyped disciplines, evaluators are likely to use congruency expectations to evaluate the academics and punish the deviations from these expectations that would manifest itself as lower evaluations of females. Given this, it is natural to expect that if the discipline is male-stereotyped then females will be evaluated lower than males in this discipline. The reverse might be true in disciplines that are female-stereotyped: indeed a study by Hesselbart (1977) has shown that male nurses are rated as more unattractive, less skilled and unambitious than female nurses.

Hypothesis 9: (effect of academic discipline): Female academics teaching in male-stereotyped disciplines and delivering the same content will be evaluated worse than male academics, whereas reverse gender bias will be observed in female-stereotyped disciplines.

OVERVIEW OF STUDIES

We test our research hypotheses in four experimental studies. In Study 1, we 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. The lecture is in the field of astronomy - a male-stereotyped discipline.

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 subjects' intuitions regarding a possible bias and examine how bias awareness affects teaching evaluations.

In Study 3, we test the robustness of Study 1-2 results by altering experimental materials. We use a different male-stereotyped discipline and alter the measures that the academics are evaluated in: we ask participants to evaluate academics in warmth, competence, the quality of the taught content, the likelihood of enrolment success of the academic and overall evaluation of the academic.

In Study 4, we test the effect of gender stereotypes about the academic discipline on gender bias in academic evaluations. We conduct two pre-tests to test and assess gender stereotypes about 20 academic disciplines. As a result, we choose two academic discipline as the most male/female-stereotyped ones and test for the existence of gender bias in each discipline using similar methodology as in Study 3.

All experimental instructions are in the appendix of the paper.

STUDY 1

Participants and Design

We recruited 479 participants on the www.Prolific.ac[1] website (M_{age} = 24.07, SD_{age} = 3.17, 50.2% female) 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 ages between 18 and 30 years old as being of the age of potential students due to the teaching evaluation context of the study. All participants were from countries with a 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.

[1] The website is a designated platform for conducting academic research with research participants from across the globe.

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).

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, we manipulated the text so that the candidate appeared warm and accessible as a teacher. In the version of the lecture which was low on warmth, we manipulated the text so that 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 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-stereotyped and whether the high warmth version of the lecture text was perceived as warmer than the low warmth version. Twenty one individuals participated in this pilot study for the payment of £1.40 at www.Prolific.ac website. 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-stereotyped ($t(21) = 1.92$, $p < .05$). 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 successful 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.

Measure of 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 $\alpha = .76$).

*Measure of 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 $\alpha = .70$).

Measure of Hiring Recommendation. Participants were asked whether the university should hire the candidate who had given the lecture on a 5-point Likert-type scales anchored by 1 (definitely reject) to 5 (definitely hire).

We controlled for age, gender of participants, 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).

At the conclusion of the study, participants were asked to indicate the gender of the academic that they had evaluated. 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. Our manipulation of the teaching style worked as expected. The high warmth lecture was rated more highly on warmth than the low warmth lecture (Mean = 3.97, SD = 0.68 versus Mean = 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%).

Insert Table 1 about here

To test Hypothesis 1, we conducted multiple regression analysis with robust standard errors of the hiring recommendation as the dependent variable. The independent variables were dummies whether the academic candidate was male (vs female), whether the teaching style was High Warmth (vs Low Warmth), and their interaction (Male × HighWarmth). 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 = -.27 and -.19 respectively).

Insert Table 2 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 the male dummy was positive and significant (β = .34, p < .01) qualified by a negative and significant interaction term (β = -.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 (β = -.18, *ns*). We further examined the predicted marginal effects of teaching style on competence evaluations depending on gender (Aiken, West, & Reno, 1991). Post estimation Wald-statistics showed there was a statistically significant decrease in the evaluations of competence for women when they taught in a high warmth style (β = -.19, p < .05), but not for men (β = -.01, *ns*) (see Figure 1). 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. In an additional regression analysis we checked whether the bias manifests more for male participants than for female participants using interaction terms between the gender treatment and the gender of the participant (Boring 2017; Mengel et al. 2019). We find no significant effect of participant gender on gender bias.

Insert Figure 1 about here

To test Hypothesis 4, we analysed whether warmth and competence acted as predictors between the gender of the academic and the hiring recommendation (Kenny, 2016). Without controlling for warmth and competence, the hiring recommendation was predicted by the gender of the academic (Table 2, column 1). Including warmth and competence measures in the regression separately (Table 2, Column 4 and 5), we observe that the gender of the academic lost significance when we controlled for the warmth evaluation ($\beta = .04$, *ns*) while both warmth and competence evaluations significantly predicted hiring recommendation ($\beta = .53$, $p < .001$ for warmth and $\beta = .75$, $p < .001$ for competence). This was the first indication that the gender bias in hiring recommendations was explained by the warmth and competence evaluations of the academic. Additionally, we conducted multiple mediation analysis on the observed coefficients to quantify the direct and indirect effects of warmth and competence evaluations on hiring recommendations. We use bootstrapped standard errors to allow for kurtosis in the data and use percentile confidence intervals to test for significance levels. These confidence intervals are non-symmetric reflecting the skewness of the sampling distribution of the product coefficients. If the confidence interval does not contain zero then the indirect effect is considered to be statistically significant (Preacher and Hayes 2008). We find that for a low warmth teaching style, the indirect effects of gender through warmth and competence were significant (standardized coefficient = .10, $p < .05$) whilst the direct effect of gender lost significance (standardized 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 coefficient = .08, $p < .05$) whilst the direct effect of gender lost significance (standardized coefficient = .00, *ns*) (see Table 3, row 2). Thus we found support for Hypothesis 4 that gender bias in hiring recommendations was mediated by warmth and competence perceptions.

Insert Table 3 about here

Discussion

Study 1 showed that in a male-stereotyped 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 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 academic seniority on hiring recommendations. 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., 2019). In Study 2, we distinguished deliberately between junior (post-PhD) and senior (Professor Level) academics to exogenously test the de-biasing role of seniority on hiring recommendations.

Participants and Design

We recruited a further 478 participants on Prolific.ac (M_{age} = 24.40, SD_{age} = 3.24, 49.9% female) 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.

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.

*Measure of 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 α = .77).

Measure of 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 α = .67).

Measure of 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.

Measure of 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 in at least two out of three measures (1 = Bias Aware, 0 = Not Aware).

We controlled for age, gender, level of education, student status, and cross-cultural differences in the gender gap as in Study 1. Participants were asked the gender of the academic that they had evaluated. A total of 6 participants (1.25%) failed this 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 with robust standard errors 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 (β = -.03, *ns*, effect size Cohen's d = .05) supporting Hypothesis 5. Predicted marginal effects

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 (β = .23, p < .05 for female academics, versus β = .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 marginally significant (β = .19, p < .10). Parametrically comparing the size of the gender bias in the hiring recommendation between the two studies (β = -.03, *ns*) does not yield a significant difference (see Table 6, Column 1).

Insert Table 6 about here

As in Study 1, gender bias in the evaluations of the academic's warmth was statistically significant (β = .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 (β = -.39, p < .05). This seems to be consistent with Hypothesis 6. However, the predicted marginal effects showed that the pattern of change of warmth evaluations was operating mainly through decrease evaluations of male academics rather than increase in females (β = .15, *ns* for female academics, versus β = .-24, p < .05 for male academics, see Figure 2). For competence, seniority affected the evaluation of female and male academics the same (see Table 5, column 3): β = .11, *ns* and β = .09, *ns* respectively)

Insert Table 7 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 8, we did not find that being aware of a male bias de-biases teaching evaluations. In an additional regression analysis, we checked whether the bias manifests more for male participants than for female participants using interaction terms between the gender treatment and the gender of the participant (Boring 2017; Mengel et al. 2019). We find no significant effect of participant gender on gender bias.

Insert Figure 2 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 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.

STUDY 3

In Study 3, we tested whether gender bias exists in another male-stereotyped academic discipline and whether alternative de-biasing technique using academic credentials is successful. We used finance as a context of academic's prepared lecture that students had to evaluate. The seniority effects that we tested in Study 2 might have age effects that we would not be able to control for. Hence in Study 3, we kept the seniority of lecturers the same but varied whether participants could see lecturer's previous academic credentials in terms of their research and teaching excellence. We pre-registered this study prior to data collection which can be accessed at https://osf.io/82tmy.

Participants and Design

We recruited 1400 participants on Prolific.ac (M_{age} = 22.1, SD = 3.16, 50% female) for a study that asked them to assess a lecture by a candidate in the academic job market. Participants were restricted to be 18-30 years old, holding student status, not have participated in any of our previous experiments and were paid £1 for completing a 10-minute study (average completion time was 8 minutes). Data were gathered in July 2019.

The study consisted of a 2 (gender: male vs female) x 2 (teaching style: high vs low warmth) x 2 (credentials: High vs No) between-subject design. Participants were randomly assigned to one of the eight conditions. The number of participants required for the study was determined to be powered at 80% with a small effect size (i.e., Cohen's f = .15; Cohen, 1992). All power calculations were conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Participants read either the low or the high warmth version of a finance lecture describing the Capital Asset Pricing Model using mathematical formulas and examples. As in previous studies in the high warmth version of the lecture, we manipulated the text so that the candidate appeared warm and accessible as a teacher. In the low warmth version of the lecture, we manipulated the text so that the candidate appeared to be cold and patronizing. The academic was described as a post-PhD male/female candidate being considered for a lecturer position at a university. A silhouette of either a male of a female head together with the academic's name (Sue Smith or Steve Smith) was shown on each of four screens of the text to reinforce the gender manipulation. In addition, depending on the experimental condition, below the silhouette on each screen there was a 4 line summary of academic's previous credentials, such as their previous teaching evaluations, awards, admin experience and grants. In the high credentials treatment these were quite positive (4.9 out of 5 teaching evaluation score, best conference paper and innovator of the year awards, being a committee member and two grants) while in the no credentials treatment these were replaced with "information pending" statement.

Procedure: Participants were randomly assigned to one of the eight experimental conditions and proceeded to read the finance lecture. Following the lecture, they assessed the academic candidate in terms of warmth (6 items), competence (6 items), enrolment in academic's classes (4 items), overall evaluation of the academic (4 items) and the lecture (4 items). The items were either framed as more objective or more subjective in nature. Participants were asked to express their assessments on a 7-point Likert Scale ranging from strongly disagree to strongly agree. We pre-registered the aim of comparing whether gender bias manifests itself more in objective than in subjective evaluation measures in line with shifting standards theory by Biernat (1995). However, given that we do not find any difference between gender and objective and subjective evaluations, we pool the description of our items and the results.

The survey finished with manipulation check questions (check of academic discipline, credentials, gender and teaching style) and questions about participants' socio-demographic characteristics.

Measure of Warmth: Participants were asked to express their agreement to the statements that the candidate was "warm", "approachable", "enthusiastic", "will be sought advice on their teaching style", "would be ranked top in terms of friendliness and approachability", "would volunteer and build a supportive environment for student learning". The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach's α = 0.93).

Measure of Competence: Participants were asked to express their agreement to the statements that the candidate was "professional", "knowledgeable", "mastered the content of the lecture", "would progress up the academic ladder faster than average", "would bring large research grants", "will reach Full Professor position". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.82).

Measure of Enrolment: Participants were asked to express their agreement to the statements that the respondent "would enrol in any class taught by this candidate", "would join the waiting list if the classes were oversubscribed", and that candidate's classes "would be oversubscribed", "would be the most oversubscribed classes in the respective program". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.89).

Measure of Overall Academic: Participants were asked to express their agreement to the statements that "they would recommend the candidate's classes to other students", "candidate should be hired", "anyone would recommend the candidate's classes to other students" and "the university will definitely hire the candidate". The items were averaged together to form a

single composite score, where higher scores indicated greater competence (Cronbach's α = 0.93).

Measure of Overall Content: Participants were asked to express their agreement to the statements that "the lecture would rank top among all lectures on the topic", "that finance degree would be incomplete without the content of this lecture", "the material was intellectually stimulating" and "material was of high quality". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.73).

We controlled for age, gender, level of education and the country of residence of the respondents in all of our analysis. A very high number of participants failed gender and credentials manipulation checks in our study: 585 failed the gender manipulation check and 322 failed credentials manipulation check. We present the results with the inclusion of those participants who failed manipulation checks, as leaving them out would leave the study underpowered. However, we later discuss the results in the light of these manipulation check failures and the possible reasons behind these failures and how we address them in Study 4. We excluded 82 participants that were outliers in terms of survey completion time less than one standard deviation from the mean (3 minutes 10 seconds).

Results

The descriptive statistics of all elicited variables are in Table 8.

Insert Table 8 about here

We tested Hypotheses 1-3 and Hypothesis 7 by running a multiple linear regression with robust standard errors of the five dependent variables on three independent variables, their interactions and controls. The three independent variables were the dummies for the gender of the candidate (gender: 1=male, 0=female), the presence of high credentials

information (credentials: 1=high credentials, 0= no credentials) and whether the teaching style was high warmth (warmth: 1= high warmth, 0 = low warmth). We find that the teaching style and credentials manipulation checks worked as predicted. The effect of high credentials is positive and significant for the evaluations of enrolment (β = .36, $p < 0.001$) and competence (β = .29, $p < 0.05$). The teaching style had a positive and significant effect on all of the dependent variables: academics in high-warmth scenario were evaluated higher than academic in low-warmth scenario in terms of enrolment (β = .96, $p < 0.001$), overall academic evaluation (β = 1.20, $p < 0.001$), overall content evaluation (β = .51, $p < 0.001$), warmth (β = 1.73, $p < 0.001$) and competence (β = .44, $p < 0.001$).

However, contrary to the Hypotheses 1-3 and Hypothesis 8, we do not find any gender bias in our data in either of our treatments. The coefficient of the gender variable and all of its interactions with treatment manipulations is statistically null.

Insert Table 9 about here

Discussion

Study 3 aimed to test whether gender biases extend to other male-stereotyped academic disciplines such as finance and elicited a more extensive list of dependent variables such as enrolment, content evaluation and overall evaluation of the academic. It also asked whether the presence of high credentials will de-bias students' gender biases. While high warmth teaching style and presence of high credentials did improve the evaluations of the academic, contrary to our hypotheses we did not find any gender bias in neither of the treatments.

We noted that in this study we observed a very high number of manipulation check failures in terms of gender (40% failed to remember what gender the academic candidate was). Thus one potential reason why we did not observe gender bias could be inattention of participants who completed the study and received the fixed payment. Another potential issue

could be that students did not perceive finance as a male-stereotyped academic discipline. We address these two potential issues in Study 4 and using a pre-test to elicit gender stereotypes.

PRETEST STUDY ON GENDER STEREOTYPES

We ran two strands of pre-tests to determine what academic disciplines are the most male-stereotyped. We employ the method developed by Babin (2019) using incentivized coordination games to elicit gender stereotypes. Both pre-tests were preregistered prior to data collection which can be accessed at https://osf.io/rsv7b .

In the first pre-test, we elicited the gender stereotypes of students about academic disciplines. We recruited 90 participants on www.Prolific.ac (M_{age} = 22-24 bracket, SD = 1.9, 62% female). Participants were restricted to be between 18-30 year old, currently residing in the UK and being a student at a UK higher education institution. Participants were paid £0.30 fixed payment for their on average 6-minute participation and additionally were paid an average of £0.50 bonus payment that depended on their answers. The data was collected in November 2019.

Participants were asked to read the same finance lecture as in Study 3. First part of the bonus payment depended on guessing whether the majority of the participants thought that the (i) lecture was prepared by a man/woman/either gender, (ii) that the material of the lecture fits a male/female lecturer more than a female/male lecturer or fits both gender, (iii) whether the lecture could be mastered by male/female academic more than the female/male academic or mastered equally by both genders and (iv) whether it was more of a male/female type of a lecture or gender-neutral. The second part of the bonus payment depended guessing the answers of the majority of participants of whether each of the given 20 different academic disciplines was “more appropriate for men”, “appropriate for all, irrespective of gender” or

"more appropriate for women". Participants earned £0.05 for coordinating on the most common answer given by the majority of the other participants.

Most participants coordinated on the answer that the lecture was prepared by a man (70% versus 25% either gender), was more fitting to a male academic (60% versus 35% fitting to both genders), and was male type (57% versus 38% gender-neutral). Of the 20 academic disciplines, participants coordinated on Finance, Physics, Maths, Computer Science and Engineering being in the top 5 of male-stereotyped academic fields (majority choosing "more appropriate for men"). On the other hand, Health Sciences, Hospitality & Tourism, Environmental Studies, Sociology and Psychology were in the top 5 of female-stereotyped academic fields (majority choosing "more appropriate for women").

In the second pre-test, we recruited 50 participants on www.Prolific.ac (M_{age} = 22-24 age bracket, SD = .88, 52% female). Participants we restricted to be between 18-30 year old, currently residing in the UK and being a student at a UK higher education institution. Participants were paid £0.50 fixed payment for their on average 5-minute participation and additionally were paid an average of £0.30 bonus payment. The data was collected in December 2019.

The participants were asked to read the same finance lecture but framed as a part of Financial Engineering degree and guess the most common answer of other participants whether the (i) lecture is more likely to be prepared by a man/woman/either gender, (ii) lecture is more of male/female/gender-neutral type and (iii) financial engineering is more of a male/female/gender-neutral discipline. Most participants coordinated on the answer that the lecture was more likely prepared by a man (70% versus 28% by either gender), the lecture was a more male type (58% versus 38% gender-neutral) and financial engineering was more of a male type of discipline (84% versus 16% gender-neutral).

Given the results of the two pre-tests, we chose two academic disciplines as our male- and female-stereotyped treatments in Study 4: Financial Engineering as the male-stereotyped and Social Psychology as the female-stereotyped academic discipline.

STUDY 4

In Study 4, we tested whether the gender stereotypes about the academic discipline has an effect on students' gender bias in teaching evaluations. There has been some evidence that some discipline-specific effects may exist such that the gender bias is more prominent in male-stereotyped disciplines than female-stereotyped disciplines; for example disciplines with less or more math content (Boring et al. 2016; Mengel et al. 2019; Basow and Silberg 1987). In Study 4, we explicitly test for this conjecture by varying the discipline of the lecture using the findings from the pre-test studies on UK students' ratings of academic disciplines as being more appropriate for men or women. Additionally, we address the shortcoming of Study 3 where we observed a large number of manipulation check failures by incentivizing the participation in the experiment given participants' attention to the experiment. We pre-registered this study prior to data collection which can be accessed at https://osf.io/pjdyc.

Participants and Design

We recruited 490 participants on Prolific.ac (M_{age}=22.08, SD=3.42; 53% female) for a study that asked them to read and assess a lecture by a candidate in the academic job market. Participants we restricted to be between 18-30 year old and being a student at a UK higher education institution. Participants were paid £0.75 fixed payment for their on average 9-minute participation and additionally were paid an average of £0.20 bonus payment. The data was collected in February 2020.

The study consisted of a 2 (gender: male vs female) x 2 (discipline: Financial Engineering vs Social Psychology) between-subject design. Participants were randomly

assigned to one of the four conditions. The number of participants for the study was determined to be powered at 90% with small effect sizes (Cohen's f = .15) and allowing for 5% of data being unusable due to missing values, manipulation failures and inattention. Participants read the low warmth version of either the lecture used in Study 3 but framed as Financial Engineering or the lecture of similar length and introductory content in Social Psychology. A silhouette of either a male or a female head was shown on each screen of text to reinforce the gender manipulation. In addition, each screen showed the candidate's name below the silhouette (Sue Smith versus Steve Smith).

Procedure: Participants were randomly assigned to each of the four experimental conditions, and proceeded to read the lecture. Following the lecture, they assessed the academic candidate in terms of warmth (3 items), competence (3 items), enrolment in academic's classes (2 items), overall evaluation of the academic (3 items) and of the lecture content (2 items). Participants were asked to express their assessments on a 7-point Likert Scale ranging from strongly disagree to strongly agree. The study finished with manipulation check questions (check of academic discipline, gender and teaching style) and questions about participants' socio-demographic characteristics.

Measure of Warmth: Participants were asked to express their agreement to the statement that the candidate was "warm", "approachable" and "enthusiastic". The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach's α = 0.91).

Measure of Competence: Participants were asked to express their agreement to the statement that the candidate was "professional", "knowledgeable" and "mastered the content of the lecture". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.73).

Measure of Enrolment: Participants were asked to express their agreement to the statement that they "would enrol in any class taught by this candidate" and "would join the waiting list if the classes were oversubscribed". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.86).

Measure of Overall Academic: Participants were asked to express their agreement to the statement that "they would recommend the candidate's classes to other students", "candidate should be hired", and "the candidate is very skilled". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.86).

Measure of Overall Content: Participants were asked to express their agreement to the statement that "the material was intellectually stimulating" and "material was of high quality". The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach's α = 0.78).

We controlled for age, gender, level of education and the country of residence of the respondents in all of our analysis. Participants were paid an additional £0.05 if they correctly answered each of five manipulation check questions. Three questions measured if participants perceived the lecture as low-warmth: 80% of the participants had at least two out of three answers correct about lecturer's unavailability, condescending tone and prohibition of questions during the lecture. More importantly, we find that 91.4% of participants perceived the gender correctly and 96.5% perceived the academic discipline belonging to a correct school (business school versus social sciences). A total of 62 participants that failed the gender, discipline manipulation checks or did not complete the study were excluded from the analysis.

Results

The descriptive statistics of all elicited variables are in Table 10.

Insert Table 10 about here

We tested Hypothesis 9 by running multiple linear regression analysis with robust standard errors of the five dependent variables on two independent variables, their interactions and controls. The two independent variables were the dummies for the gender of the candidate (gender: 1=male; 0=female) and the academic discipline their lecture was on (financial engineering: 1=Financial Engineering, 0=Social Psychology). We find that candidates in Financial Engineering discipline are evaluated less positively in terms of willingness to enrol in academic's classes (β = -.84, p < 0.001), overall evaluation of the academic (β = -.51, p < 0.050), the content of the lecture (β = -.79, p < 0.001), warmth (β = -.93, p < 0.001) and competence (β= -.34, p < 0.050) of the academic. However, we do not find support for our Hypothesis 9. There is no evidence of gender bias in either discipline: the coefficient of the gender variable and its interaction with academic discipline is null for all of our dependent variables, except competence where males are evaluated lower than females (β = -.35, p < 0.50) .

Insert Table 11 about here

Discussion

The aim of Study 4 was to address the possible issues of Study 3 in identifying gender bias. By incentivizing participants' answers to attention check questions we addressed the issue of high manipulation failures: more than 90% of participants passed the gender and academic discipline manipulation checks. Moreover by framing the lecture as financial engineering – the combination of two male-stereotyped disciplines we aimed to reach the same level of male-stereotype that astronomy lecture could have delivered in Studies 1-2. However, neither

of these resulted in identifying significant gender bias. There may be several reasons why there was no gender bias in Study 4 (as in Study 3 in contrast to Study 1-2) we can only speculate about. We discuss these in the concluding discussion section of the paper.

CONCLUDING DISCUSSION

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-stereotyped high warmth style may shield female academics from lower evaluations of the same teaching content. Yet, in one of our studies, 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 even though 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 downward pressure on the evaluations of her competence and the hiring recommendation. We did not replicate this result in a subsequent study that changed the academic discipline of the academic from astronomy to finance and evaluated the academic in more dimensions. Hence this result, although interesting, should be taken with a pinch of salt.

We showed that 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. We also aimed to test whether a different type of credentials information may mitigate gender bias. In Study 3 we provided the student evaluators with a piece of brief information about academic's credentials in terms of previous teaching evaluations, awards and grants. Given that there was no gender bias in Study 3 to start with, the positive credentials information affected male and female academics' evaluations similarly.

We also examined whether 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 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.

Finally, we examined whether the gender stereotyping about the academic discipline may affect the gender bias in Study 4. Students evaluated an academic teaching Financial

Engineering or Social Psychology degree. We did not find any gender bias in either of the academic disciplines and hence there was no differential effect of academic discipline on the gender bias.

We may only speculate why the significant gender bias in Studies 1-2 did not manifest itself in Studies 3-4. In pre-test studies, we showed that both physics (astronomy) and finance (financial engineering) were male-stereotyped disciplines; hence we can rule out the possibility of the stereotype of the discipline causing the gender bias. Another possibility could be that lower fixed payment amounts of Study 3 (compared to Studies 1-2) may have caused more inattention to survey and noisier evaluations. We shortened the survey length significantly to make it comparable to Study 1-2 payments and make attention checks incentivized in Study 4 and yet did not find any gender bias. Hence, we can rule out the payment/attention deficiency possibility as the potential drivers for insignificant gender bias in later studies. One last potential explanation for the differences between the earlier and later studies could be the student status of the participants. While in the earlier studies, we recruited participants pre-screening them to be aged 18-30 who are *potentially* of student status, in the later studies we pre-screened participants to be 18-30 year old and *currently holding* a student status. Unfortunately, we do not hold any information about participants' student status in Studies 1-2, and hence we cannot test whether the gender bias is moderated by the student status. However, given the general profile of 18-30 year olds on Prolific.ac we expect that 35% of our participants in Study 1-2 to be students. Given that in Studies 3-4, student participants rated the male and female candidates equally, our results challenge the results of significant gender bias present in literature.

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 not 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-stereotyped 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.

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.

Babin, J. (2019). Detecting Group Gender Steretypes: Opinion Mining vs. Incentivized Coordination Games. *Journal of Economic Insight*, 45 (1), 21-42.

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently?. *Journal of educational psychology, 79*(3), 308.

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.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, 1212382.

Beshears, J., & Gino, F. (2015). Leaders as decision architects. *Harvard Business Review, 93*(5), 52-62.

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.

Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In Y.-T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (p. 87–114). American Psychological Association.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41.

Boring, A., & Philippe, A.. (2019). Reducing Discrimination in the Field: Evidence from an Awareness Raising Intervention Targeting Gender Biases in Student Evaluations of Teaching. Retrieved from http://www.anneboring.com/uploads/5/6/8/5/5685858/boring_philippe_reducing_gender_biases_in_sets_9_11_19.pdf

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

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.

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues, 57*(4), 725-741.

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.

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.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608-630.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics, 16*(1), 67-88.

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.

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.

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.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573-598.

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.

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*(2), 168-192.

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.

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

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

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.

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.

Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal*, *123*(572), 417-432.

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.

Hesselbart, S. (1977). Women doctors win and male nurses lose: A study of sex role and occupational stereotypes. *Sociology of Work and Occupations*, *4*(1), 49-62.

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.

Kasnakoglu, B. (2016). Antecedents and consequences of co-creation in credence-based service contexts. *The Service Industries Journal, 36*(1-2), 1-20.

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

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.

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.

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.

Langbein, L. I. (1994). The validity of student evaluations of teaching. *PS: Political Science & Politics, 27*(3), 545-553.

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.

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.

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.

Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association, 17*(2), 535-566.

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

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.

Neilson, W., & Ying, S. (2016). From taste-based to statistical discrimination. *Journal of Economic Behavior & Organization, 129*, 116-128.

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

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.

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.

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.

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.

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.

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.

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.

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.

Way, S. F., Larremore, D. B., & Clauset, A. (2016). Gender, productivity, and prestige in computer science faculty hiring networks. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 1169-1179).

Wild, C., & Berger, D. (2016). The proposed teaching excellence framework (TEF) for UK universities. *International Journal of Teaching and Education*, *4*(3), 33-50.

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.

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.

Table 1: Means (and st. dev) for Study 1 variables

	Low Warmth Condition		**High Warmth Condition**	
Candidate:	Female	Male	Female	Male
Dependent Variables				
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	51%	43%	55%	49%
Graduate Degree	46%	43%	46%	47%
Postgraduate Degree	16%	19%	19%	14%
Over 24 years old	46%	46%	44%	48%
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	Hiring	Hiring
	(1)	(2)	(3)	(4)	(5)
Male Academic	0.21**	0.34***	-0.02	0.21**	0.04
	(0.10)	(0.09)	(0.08)	(0.09)	(0.09)
High Warmth Scenario	0.17	0.58***	-0.19**	0.30***	-0.15
	(0.10)	(0.09)	(0.08)	(0.09)	(0.15)
Male Academic ×	-0.06	-0.27**	0.18	-0.19	0.06
High Warmth Scenario	(0.15)	(0.14)	(0.11)	(0.12)	(0.13)
Warmth					0.53***
					(0.05)
Competence				0.75***	
				(0.05)	
Constant	3.74***	3.27***	4.39***	0.46***	2.02***
	(0.11)	(0.10)	(0.08)	(0.24)	(0.19)
Control variables	Yes	Yes	Yes	Yes	Yes
Observations	479	479	479	479	479
Adj R-squared	0.02	0.10	0.02	0.31	0.23

Note: * = $p < .10$, ** = $p < .05$, *** = $p < .01$ with robust standard errors

TABLE 3

Standardized Mediation Analysis of Warmth and Competence on the Hiring Recommendation in Study 1.

	Indirect effects on Hiring			
Teaching Style	Direct Effect of Gender on Hiring	Via Warmth	Via Competence	Total	Total Effect of Gender on Hiring
Gender - Low Warmth	0.04	**0.10****	0.00	**0.10****	**0.14****
Gender - High Warmth	0.00	0.01	**0.07****	**0.08****	0.08

Note: Low Warmth N=252, High Warmth N=227.

TABLE 4

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

	Junior Level		**Professor Level**	
Candidate:	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	40%	55%	58%	47%
Graduate Degree	50%	48%	49%	51%
Postgraduate Degree	13%	16%	18%	14%
Over 24 years old	48%	54%	46%	59%
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
	(1)	(2)	(3)
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 × Professor Level	-0.22	-0.39**	-0.02
	(0.16)	(0.15)	(0.10)
Constant	3.95***	3.36***	4.30***
	(0.11)	(0.11)	(0.08)
Controls	Yes	Yes	Yes
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

Comparison of Male Academic Coefficients in Study 1 and Study 2

Variables	Hiring	Warmth	Competence
	(1)	(2)	(3)
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 × Study 2	-0.04	-0.11	-0.02
	(0.15)	(0.14)	(0.10)
Constant	3.79***	3.31***	4.35***
	(0.11)	(0.10)	(0.08)
Control	Yes	Yes	Yes
Observations	494	494	494
R-squared	0.02	0.04	0.02

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

TABLE 7

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

Variables	Hiring	Warmth	Competence
	(1)	(2)	(3)
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)
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: * = $p < .10$, ** = $p < .05$, *** = $p < .01$

TABLE 8

Means (and standard deviations in brackets) for Study 3 variables

	Low-Warmth Condition				**High Warmth Condition**			
	No Credential		**High Credential**		**No Credential**		**High Credential**	
Candidate:	Female	Male	Female	Male	Female	Male	Female	Male
Warmth	3.36 (1.53)	3.19 (1.59)	3.46 (1.50)	3.56 (1.43)	5.06 (0.96)	5.07 (1.04)	5.08 (0.94)	5.10 (0.83)
Competence	4.87 (1.10)	4.92 (0.97)	5.17 (0.91)	5.22 (0.92)	5.29 (0.74)	5.31 (0.86)	5.49 (0.78)	5.43 (0.81)
Enrolment	3.18 (1.42)	3.13 (1.48)	3.53 (1.50)	3.45 (1.42)	4.13 (1.27)	4.10 (1.29)	4.17 (1.24)	4.10 (1.09)
Overall Academic	4.02 (1.69)	3.89 (1.68)	4.22 (1.65)	4.39 (1.50)	5.20 (1.03)	5.19 (1.13)	5.18 (1.08)	5.20 (1.08)
Overall Content	4.49 (1.24)	4.59 (1.13)	4.59 (1.18)	4.68 (1.10)	4.99 (1.02)	5.05 (1.01)	5.09 (1.03)	5.02 (1.00)
Control Variables								
Male Respondent	49%	53%	56%	51%	51%	50%	45%	51%
Graduate Degree	85%	88%	83%	85%	85%	95%	91%	86%
Postgraduate Degree	15%	12%	17%	15%	15%	5%	9%	14%
Over 24 years old	25%	25%	20%	28%	24%	16%	18%	28%
Observations	164	165	158	185	154	152	191	159

TABLE 9

Regressions for Enrolment, Overall Academic, Overall Content, Warmth and Competence in Study 3

Variables	Enrolment	Overall Academic	Overall Content	Warmth	Competence
	(1)	(2)	(3)	(4)	(5)
Male Academic	-0.02	-0.14	0.10	-0.16	0.05
	(0.16)	(0.18)	(0.13)	(0.17)	(0.11)
High Credential	0.36**	0.21	0.11	0.13	0.29***
	(0.16)	(0.18)	(0.13)	(0.17)	(0.11)
High Warmth Scenario	0.96***	1.20***	0.51***	1.73***	0.44***
	(0.15)	(0.16)	(0.130	(0.14)	(0.11)
Male Academic × High Credentials	-0.10	0.30	-0.03	0.24	0.01
	(0.23)	(0.25)	(0.18)	(0.23)	(0.15)
Male Academic × High Warmth Sc.	-0.01	0.09	-0.05	0.16	-0.06
	(0.220	(0.22)	(0.18)	(0.21)	(0.15)
High Credentials × High Warmth	-0.33	-0.25	-0.03	-0.12	-0.11
	(0.210	(0.22)	(0.18)	(0.20)	(0.14)
Male Academic × High Credentials × High Warmth	0.05	-0.25	-0.07	-0.24	-0.05
	(0.29)	(0.30)	(0.24)	(0.28)	(0.20)
Constant	3.49***	4.32***	4.78***	3.61***	4.95***
	(0.18)	(0.19)	(0.15)	(0.18)	(0.13)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	1,328	1,328	1,328	1,328	1,328
R-squared	0.10	0.16	0.06	0.33	0.06

Note: * = $p < .10$, ** = $p < .05$, *** = $p < .01$ with robust standard errors in brackets.

TABLE 10

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

	Financial Engineering		**Social Psychology**	
Candidate:	**Female**	**Male**	**Female**	**Male**
Warmth	2.25	2.49	3.148	2.87
	(1.35)	(1.45)	1.58	(1.44)
Competence	4.82	4.68	5.13	4.76
	(1.30)	(1.01)	1.163	(1.27)
Enrolment	2.23	2.42	3.048	2.88
	(1.55)	(1.48)	1.58	(1.56)
Overall Academic	3.47	3.49	3.957	3.86
	(1.55)	(1.45)	1.528	(1.61)
Overall Content	4.08	4.11	4.854	4.59
	(1.50)	(1.34)	1.335	(1.33)
Control Variables				
Male Respondent	50%	45%	45%	41%
Graduate Degree	38%	44%	39%	40%
Postgraduate Degree	16%	11%	17%	16%
Over 24 years old	25%	19%	24%	26%
Observations	116	108	103	101

TABLE 11

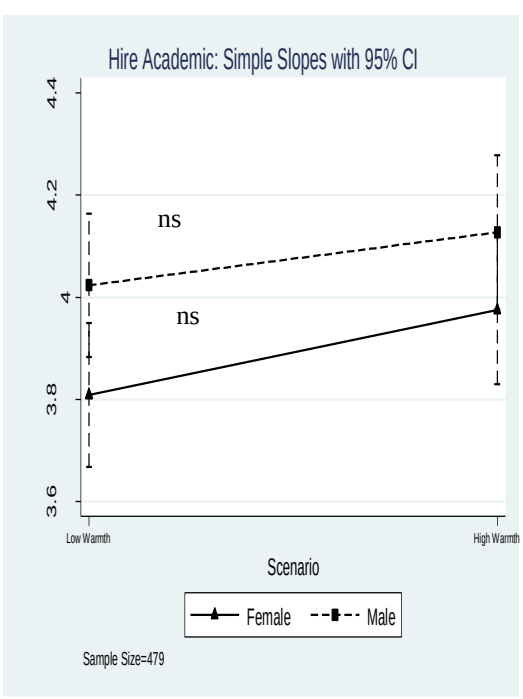
Regressions for Enrolment, Overall Academic, Overall Content, Warmth and Competence in Study 4

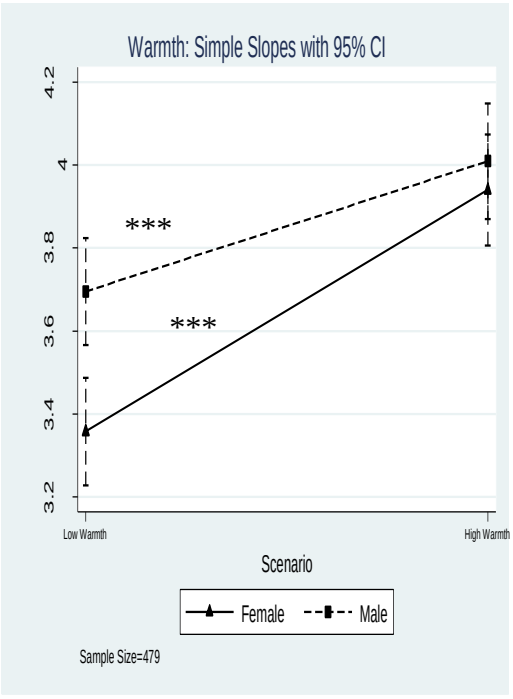
Variables	Enrolment	Overall Academic	Overall Content	Warmth	Competence
	(1)	(2)	(3)	(4)	(5)
Male Academic	0.14	-0.05	-0.22	-0.24	-0.35**
	(0.22)	(0.22)	(0.19)	(0.21)	(0.16)
Financial Engineering	-0.84***	-0.51**	-0.79***	-0.93***	-0.34**
	(0.21)	(0.20)	(0.19)	(0.20)	(0.16)
Male Academic × Financial Engin.	0.32	0.06	0.24	0.49*	0.20
	(0.30)	(0.29)	(0.26)	(0.28)	(0.23)
Constant	4.44***	5.23***	5.95***	3.88***	6.18***
	(0.68)	(0.64)	(0.53)	(0.62)	(0.48)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	428	428	428	428	428
R-squared	0.08	0.07	0.10	0.10	0.07

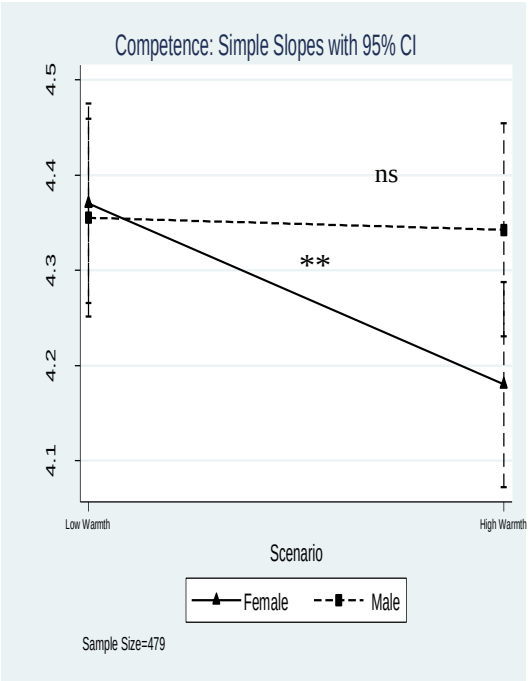
Note: * = $p < .10$, ** = $p < .05$, *** = $p < .01$ with robust standard errors in brackets.

FIGURE 1

Study 1 –Predicted Marginal Effects for the Hiring Recommendation, Warmth and Competence.





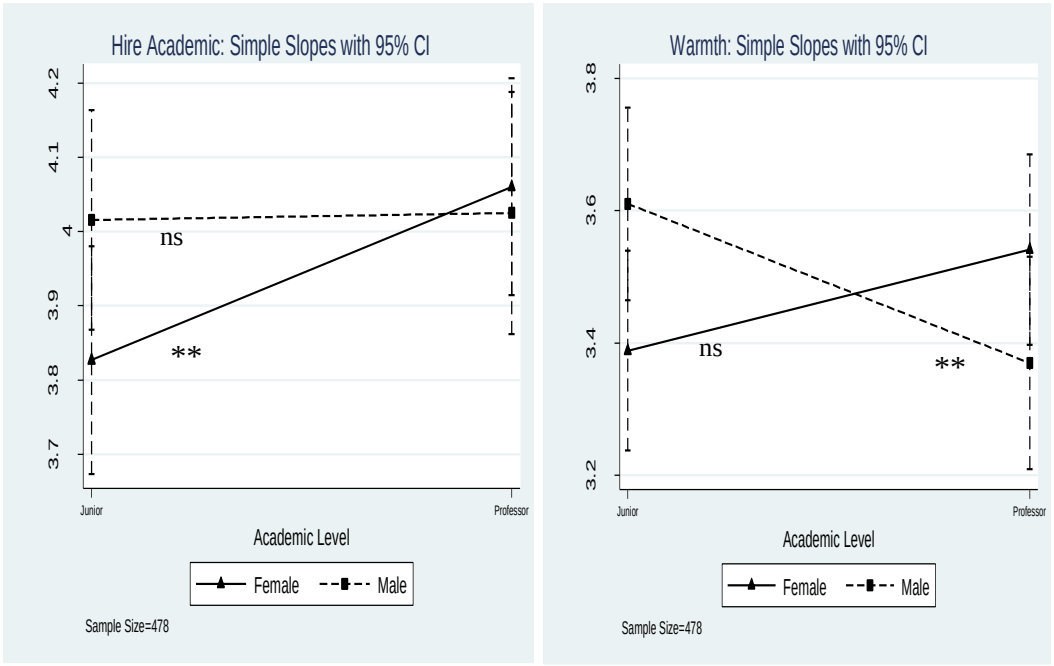


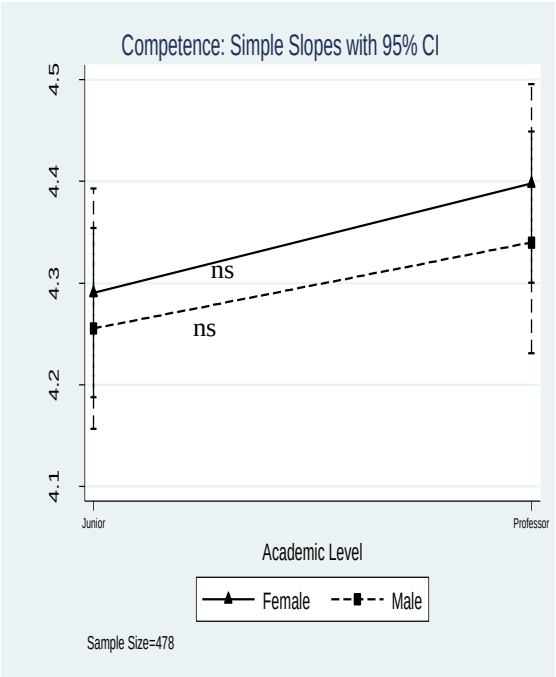
Note: 95% confidence interval shown as a dotted line.

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

FIGURE 2

Study 2 Predicted Marginal Effects for the Hiring Recommendation, Warmth and Competence.





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

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

Experimental Materials

STUDY 1:

You are about to see the text of an astronomy lecture (of about 900 words) presented by a graduate student who is applying for a job at a university. The candidate's performance in this lecture will be crucial to the recruitment committee's decision. Your role, once you have seen the lecture, is to rate the candidate's performance and indicate whether you would either hire or reject them.
When you are ready to proceed, please press NEXT PAGE

[Male x Female Manipulation High-Warmth x Low-Warmth Teaching Style]

Page Headings:

 Candidate Sue Smith or  Candidate Steve Smith

Page 1

Welcome to today's introductory lecture. My name is Sue/Steve Smith and I am a graduate student in astronomy. I hope you will find my talk interesting. I will be very happy to discuss complex areas with individual students afterwards. / I would appreciate quiet throughout the lecture so please turn mobile phones off. I will take questions after the lecture.
My talk is on the history of black holes. It is said that fact is sometimes stranger than fiction, and nowhere is that more true than in the case of black holes.
Black holes are stranger than anything dreamed up by science fiction writers, but they are firmly matters of science fact. The scientific community was slow to realize that massive stars could collapse in on themselves, under their own gravity, and how the object left behind would behave.
I understand that / In fact Albert Einstein even wrote a paper in 1939, claiming stars could not collapse under gravity, because matter could not be compressed beyond a certain point. Many scientists shared Einstein's gut feeling.

Page 2

The principal exception was the American scientist John Wheeler, who in many ways is the hero of the black hole story. In his work in the 1950s and '60s, he emphasized that many stars would eventually collapse, and the problems that posed for theoretical physics.
He also foresaw many of the properties of the objects which collapsed stars become, that is, black holes. The phrase 'black hole' is simple enough but it's hard to imagine one out there in space. Think of a giant drain with water spiralling down into it. Once anything slips over the edge or 'event horizon', there is no return.
Maybe we could describe / NASA describes stars as rather like pressure-cookers. The explosive force of nuclear fusion inside them creates outward pressure which is constrained by gravity pulling everything inwards. Eventually, however, the star will exhaust its nuclear fuel. The star will contract. In some cases, it may be able to support itself as a white dwarf star. However it was shown before the war, that the maximum mass of a white dwarf star exceeds that of the Sun. A similar maximum mass was calculated by a Soviet physicist for a star made entirely of neutrons.

In 1939, Robert Oppenheimer, of later atom bomb fame, showed that when a massive star exhausted its nuclear fuel it could not be supported by pressure. And that if one neglected pressure, a uniform spherically systematic symmetric star would contract to a single point of infinite density. Such a point is called a singularity.

Page 3

Take my word for it, / Evidentially all our theories of space are formulated on the assumption that spacetime is smooth and nearly flat, so they break down at the singularity, where the curvature of space-time is infinite. In fact, it marks the end of time itself. That is what Einstein found so objectionable.

Then the war intervened.

Most scientists, including Robert Oppenheimer, switched their attention to nuclear physics, and the issue of gravitational collapse was largely forgotten. Interest in the subject revived with the discovery of distant bright objects, called quasars. The first of these was discovered in the early 1960's.

Nuclear processes could not account for their energy output, because they release only a percent fraction of their rest mass as pure energy. As we'll discover in the next few weeks the only alternative was gravitational energy, released by gravitational collapse.

Gravitational collapses of stars were re-discovered. It was clear that a uniform spherical star would contract to a point of infinite density, a singularity.

Page 4

When John Wheeler introduced the term black hole in 1967, it replaced the earlier name, frozen star. Wheeler's coinage emphasized that the remnants of collapsed stars are of interest in their own right, independently of how they were formed.

So, what do we need to know about black holes? / So, what can anyone tell me about the properties of black holes?

From the outside, you can't tell what is inside a black hole. You can throw television sets, diamond rings, or even your worst enemies into a black hole, and all the black hole will remember is the total mass, and the state of rotation.

A black hole has a boundary, called the event horizon. It is where gravity is just strong enough to drag light back, and prevent it escaping.

Because nothing can travel faster than light, everything else will get dragged back also.

Falling through the event horizon is a bit like going over Niagara Falls in a canoe. If you are above the falls, you can get away if you paddle fast enough, but once you are over the edge, you are lost. There's no way back.

Page 5

It appears that the number of configurations that could form a black hole of a given mass, although very large, may be finite. You'll have to accept my word for this. Jacob Bekenstein suggested that from this finite number, you could interpret what we call the entropy of a black hole. This would be a measure of the amount of information that was irretrievably lost during the collapse when a black hole was created.

The apparently fatal flaw in Bekenstein's suggestion was that if information is lost, which is apparently what is happening in a black hole, there should be some release of energy - but that flies in the face of the theory that nothing comes out of black holes.

This is a paradox. And it's one which I am going to return to in my next lecture, when I'll be exploring how black holes challenge the most basic principle about the predictability of the universe, and the certainty of history, and asking what would happen if you ever got sucked into one.

My name is Sue/Steve Smith and I'd like to thank you for attending today.

I am available to discuss this lecture now and during office hours on the complex topics that we've covered today / For anyone who was not able to follow this introductory lecture there is an extensive reading list on my website.

Candidate Evaluation and Recruitment Decision

You will now be asked some questions on how the candidate came across to you. Please give a rating for each question below including your recruitment decision.

From course profile that the candidate has prepared, did they appear professional?

Not at all Professional	Unprofessional	Neither Professional or Unprofessional	Professional	Very Professional
○	○	○	○	○

From the course profile that the candidate has prepared, did they appear knowledgeable?

Not at all Knowledgeable	Not Knowledgeable	Neither Knowledgeable or Unknowledgeale	Knowledgeable	Very Knowledgeable
○	○	○	○	○

From the course profile that the candidate has prepared, did they appear to be an accessible person?

Not at all Accessible	Not Accessible	Neither Accessible or Unaccessible	Accessible	Very Accessible
○	○	○	○	○

From the course profile that the candidate has prepared, did they appear to be a warm person?

Not at all Warm	Not Warm	Neither Warm or Cold	Warm	Very Warm
○	○	○	○	○

Should the university hire the candidate to present this course?

Definitely Reject	Probably Reject	Undecided	Probably Hire	Definitely Hire
○	○	○	○	○

[Manipulation Checks]
Was the candidate Male/ Female/ I don't remember
Was the lecture taught in a masculine/ feminine/ gender-neutral way.

Finally, please provide some information about yourself.

What is your gender? Male/Female
Please enter your age in years __
What is your highest level of qualification?

Where were you born?
Where do you currently live?
Please enter your prolific ID.

STUDY 2:

Seniority manipulation as below:

You are about to see the text of an astronomy lecture (of about 900 words) presented by a candidate who is applying for a junior/senior position at a university. The candidate you are about to assess is a graduate student / Professor of Astronomy.

The candidate's / professor's performance in this lecture will be crucial to the recruitment committee's decision. Your role, once you have seen the lecture, is to rate the candidate's / professor's performance and indicate whether you would either hire or reject them.

When you are ready to proceed, please press NEXT PAGE.

Candidate/Professor Sue Smith  or Candidate/Professor Steve Smith 

Lecture same as in Study 1, but only Low-Warmth version.

Candidate Evaluation Questions and Manipulation Check Questions as in Study 1.

STUDY 3:

Male x female; High credential x No Credential information; Low-Warmth x High-Warmth

You are about to see the text of a finance lecture (appx. 500 words) presented by an academic who is applying for a lecturer position at a business school.

The candidate's performance in this lecture will be crucial to the recruitment committee's decision. Your role, once you have seen this lecture, is to rate the candidate's performance.

When you are ready to proceed, please press NEXT PAGE.

Appearing on each page of the lecture on top left corner.

 Candidate Sue Smith or  Candidate Steve Smith

Previous Teaching Evaluations: 4.9 out of 5 in 2018.
Awards: Best Conference Paper, Innovator of the Year.
Admin Experience: Teaching and Learning Committee Member 2016-2018.
Grants: British Academy (£79,000), Leverhulme Foundation (£233,000).

Previous Teaching Evaluations: information pending.
Awards: information pending
Admin Experience: information pending.
Grants: information pending

Page 1:

Welcome to today's introductory lecture. My name is Sue/Steve Smith and I am a Lecturer of Finance. I hope you will find my talk interesting and please ask questions at any point. I will be very happy to discuss points with individual students afterwards. / I insist on quiet throughout my lectures so keep your questions till the end. I might have the time to discuss points with individual students afterwards.

The capital asset pricing model is a model that describes the relationship between systematic risk and expected return for assets, particularly stocks. CAPM is widely used throughout finance for the pricing of risky securities, generating expected returns for assets given the risk of those assets and calculating costs of capital. The formula for calculating the expected return of an asset given its risk is as follows:

$$\bar{r}_a = r_f + \beta_a(\bar{r}_m - r_f)$$

Where:
r_f = Risk free rate
β_a = Beta of the security
$\bar{r}_m$ = Expected market return

Page 2:

To sum up, and simplify the jargon, / In simple words for those unable to understand this, the general idea behind CAPM is that investors need two forms of compensation: time value of money and risk. The risk-free rate in the formula represents the time value of money and compensates the investors for placing money in any investment over time. The risk-free rate is customarily the yield on government bonds like U.S. Treasuries.

The other half of the CAPM formula represents risk and calculates the amount of compensation the investor needs for taking on additional risk. You can calculate this by taking a risk measure (beta) that compares the returns of the asset to the market over time and to the market premium (Rm-rf): the return of the market in excess of the risk-free rate. Beta reflects how risky an asset is compared to overall market risk and is a function of the volatility of the asset and the market, and the correlation between the two. For stocks, the S&P 500 usually represents the market, but more robust indexes can represent it too.

Page 3:

The CAPM model says the expected return of a security or a portfolio equals the rate on a risk-free security plus a risk premium. If this expected return does not meet or beat the required return, then the investment should not be undertaken. The security market line plots the results of the CAPM for all different risks (betas).

Let's give an example that will help us understand this / for those of you who still have not managed to understand this.

Page 4:
Using the CAPM model and the following assumptions, we can compute the expected return for a stock:
The risk-free rate is 2 percent, and the beta (risk measure) of a stock is 2. The expected market return over the period is 10 percent, so that means that the market risk premium is 8% (10% - 2%) after subtracting the risk-free rate from the expected market return. Plugging in the preceding values into the CAPM formula above, we get an expected return of 18 percent for the stock:

18% = 0.18 = 0.02 + [2 x (0.10 - 0.02)]

I'd like to thank you for attending today. I'm free to discuss this lecture now and during office hours on the complex topics that we've covered today. / For anyone who was not able to follow this introductory lecture there is an extensive reading list on my website.

Please answer the questions below evaluating the job candidate.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's classes?
If the candidate were hired, the classes offered by this candidate would be oversubscribed.
The classes offered by this candidate would feature among the top three most oversubscribed classes in their respective programs.
If I could, I would enrol in any class taught by this candidate.
If the class taught by this candidate was oversubscribed, I would join the waiting list hoping to take it eventually.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate?
Anyone who took a class with this candidate would recommend the candidate to other students.
The university will definitely hire the candidate.
I would recommend the candidate to other students who might be interested in finance.
I think that the candidate should be hired.

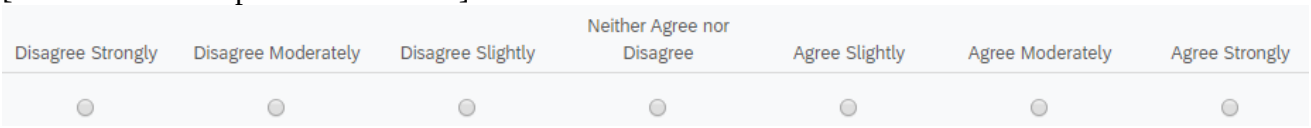
Based on the candidate's lecture, to what extent do you agree with the following statements about the content of the candidate's teaching?
The preparation of a finance degree student would be incomplete without mastering the content of this lecture.
The lecture would rank top among all lectures that introduce the CAPM in under 20 minutes.
The lecture material was intellectually stimulating.
The lecture material was of high quality.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's competence and expertise?
If the candidate were hired, the candidate would progress up the academic career at a faster-than-average rate.

If the candidate were hired, the candidate would be able to bring large research grants to the university in the next 5 years.
The candidate will reach the Full Professor position during their academic career.
The candidate was professional.
The candidate was knowledgeable.
The candidate mastered the content of the lecture.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's teaching style?
If the candidate were hired, the candidate will be sought for advice on teaching style by many colleagues.
If the candidate were hired, the candidate would rank among the top academic staff for friendliness and approachability.
If the candidate were hired, the candidate would never miss an opportunity to volunteer for events that help build a supportive environment for student learning.
The candidate was a warm person.
The candidate was an approachable person.
The candidate was enthusiastic.

[All assessed in 7 point Likert scale]



[Manipulation Checks]

What position did the candidate apply for at the university?
a lecturer position at the business school
a lecturer position in the sociology department
a professor position at the business school

What were the candidate's previous teaching evaluation scores, awards and grant history?
Very positive
Very negative
Unknown

Which of the following was TRUE about the candidate?
The candidate discouraged questions during the lecture
The candidate allowed questions at any point in the lecture

Which of the following was TRUE about the candidate?
The candidate did not seem condescending at any point during the lecture
The candidate seemed condescending at times during the lecture

Which of the following was TRUE about the candidate?
After the lecture, the candidate did not show availability to discuss lecture content in person
After the lecture, the candidate showed availability to discuss lecture content in person

Was the candidate?
Male/ Female/ I don't remember

Please provide some information about yourself. [Same as in Study 1 and 2 plus:]

Do you study business/economics or a related field?
Overall, the lecture I evaluated was ...

Very unrealistic	○ ○ ○ ○ ○ ○ ○	Very realistic
Very difficult to imagine	○ ○ ○ ○ ○ ○ ○	Very easy to imagine

STUDY 4:

[Female x Male; Financial Engineering x Social Psychology Manipulation; Low-Warmth**]**

INSTRUCTIONS
Welcome to this study. The purpose of the study is to explore the students' perspective on academic job candidates. Before commencing the study you will be presented with a consent form. You should note that your bonus payment will depend on the correctness of the attention check questions. If you do not agree to this, please leave the study now.
When you are ready, please proceed to the next page.

INSTRUCTIONS
You are about to see the text of a lecture (appx. 400 words) presented by an academic who is applying for a lecturer position at a university.
The candidate's performance in this lecture will be crucial to the recruitment committee's decision. Your role, once you have seen this lecture, is to answer some questions about the candidate and the lecture. We will not ask you about the specific terminology of the lecture but rather its general theme and context. Your bonus payment will depend on the correctness of your answers.

The lecture you are about to read is given by the following candidate who is applying for a lecturer position at a business school / school of social sciences.

 Candidate Sue Smith or  Candidate Steve Smith

[Financial Engineering Lecture]
Page 1:
My name is Sue/Steve Smith and I am a Lecturer of Finance. I have prepared an introductory lecture on Financial Engineering. I insist on quiet throughout my lectures so keep your questions till the end. I might have the time to discuss points with individual students afterwards.
The lecture is about the capital asset pricing model which is central to Financial Engineering. The capital asset pricing model is a model that describes the relationship between systematic risk and expected return for assets, particularly stocks. CAPM is widely used throughout finance for the pricing of risky securities, generating expected returns for assets

given the risk of those assets and calculating costs of capital. The formula for calculating the expected return of an asset given its risk is as follows:

$$\overline{r_a} = r_f + \beta_a(\overline{r_m} - r_f)$$

Where:

$r_f =$ *risk free rate*

$\beta_a =$ *beta of the security*

$\overline{r_m} =$ *Expected market return*

In simple words for those unable to understand this, the general idea behind CAPM is that investors need two forms of compensation: time value of money and risk. The risk-free rate in the formula represents the time value of money and compensates the investors for placing money in any investment over time. The risk-free rate is customarily the yield on government bonds like U.S. Treasuries.

The other half of the CAPM formula represents risk and calculates the amount of compensation the investor needs for taking on additional risk. You can calculate this by taking a risk measure (beta) that compares the returns of the asset to the market over time and to the market premium (Rm-rf): the return of the market in excess of the risk-free rate. Beta reflects how risky an asset is compared to overall market risk and is a function of the volatility of the asset and the market, and the correlation between the two. For stocks, the S&P 500 usually represents the market, but more robust indexes can represent it too.

Page 2

The CAPM model says the expected return of a security or a portfolio equals the rate on a risk-free security plus a risk premium. If this expected return does not meet or beat the required return, then the investment should not be undertaken. The security market line plots the results of the CAPM for all different risks (betas).

Let’s give an example, for those of you who still have not managed to understand this.

Example:

$r_f = 2\%,\ \overline{r_m} = 10\%, \beta_a = 2$

Then:

$$0.18 = 0.02 + 2 \times (0.10 - 0.02)$$

That is, the expected return of the stock is 18%

That's all for today. For anyone who was not able to follow this introductory lecture there is an extensive reading list on my website. I would prefer you just check that at www.sue-smith.com / www.steve-smith.com if you struggle rather than come talk to me in person.

[Social Psychology Lecture]

Page 1:

My name is Sue / Steve Smith and I am a Lecturer of Social Psychology. I have prepared an introductory lecture on Social Psychology. I insist on quiet throughout my lectures so keep your questions till the end. I might have the time to discuss points with individual students afterwards.

Social Psychology is defined as the scientific study of the way in which our thoughts, feelings, and behaviors are influenced by the real or imagined presence of other people. And in this lecture we will focus on interactions between individuals and how cognitive biases influence these decisions and opinions. Today we talk about the power of the situation. A big idea in social psychology; that people's behavior can be phenomenally altered by the situation they're in. Furthermore, we'll talk about attribution, how we decide whether a person behaves a certain way because of their character or the situation they're in. We engage all the time as humans as interpreting the actions and words of others. What is driving that person's thoughts? What's in their mind that's making them behave that way?
There's at least 2 explanations. One of them is character. So personality psychology is all about that. The character of the person is making them behave that way because they're outgoing or because they're shy. So Jeremiah helps the elderly person across the street because he's a really nice guy. That's the personality predisposition interpretation. Or they're behaving that way because they're in a particular situation. So Jeremiah helped the elderly man across the street because his friends are breaking in to the apartment the guy's leaving behind. It's about character versus situation. Fundamental Attribution Error is a cognitive bias to interpret others actions as a revelation of their character. Whereas, we're biased to interpret our own actions as being subject to our situation.
Page 2
In simple words for those unable to understand this consider the Stanford Zimbardo Prison experiment. The aim of this experiment was to exemplify our predisposition to the Fundamental Attribution Error. Here Stanford students signed up to participate, where they would be randomly assigned the role of prisoners or guards. Students were in the basement of the psychology building and used cupboards as makeshift cells. The results showed those assigned to be guards became sadistic and 'prisoners' were stressed. So stressed in fact they stopped the study after only six days instead of the allotted of twenty-four. Nobody has been allowed to rerun this because the behavior was so horrible, so fast. Yet it truly exemplified the power of our cognitive biases and attribution to character.
That's all for today. For anyone who was not able to follow this introductory lecture, there is an extensive reading list on my website. I would prefer you just check that at www.sue-smith.com / www.steve-smith.com if you struggle rather than come talk to me in person.

Now that you have read the lecture, we will ask you a number of questions about the candidate and the lecture. For every correct answer, you will receive £0.05 as bonus payment. Please press Proceed to continue to the questions.

Please answer the following questions about the candidate and the lecture. You will receive £0.05 for every correct answer you give as a bonus payment.

What position did the candidate apply for at the university?
a lecturer position at the business school
a lecturer position in the school of social sciences
a lecturer position in the school of history

Which of the following was TRUE about the candidate?
allowed questions at any point in the lecture
discouraged questions during the lecture
did not make any references to student questions

Which of the following was TRUE about the candidate?
The candidate was condescending at times during the lecture
The candidate was not condescending at any point during the lecture

Which of the following was TRUE about the candidate?
After the lecture, the candidate discouraged students from talking to them in person
After the lecture, the candidate encouraged students to talk to them in person
After the lecture, the candidate did not mention anything about talking to students

Was the candidate?
Male
Female
Gender was not clear

Please answer the questions below evaluating the job candidate.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's classes?
If I could, I would enrol in any class taught by this candidate.
If the class taught by this candidate was oversubscribed, I would join the waiting list hoping to take it eventually.

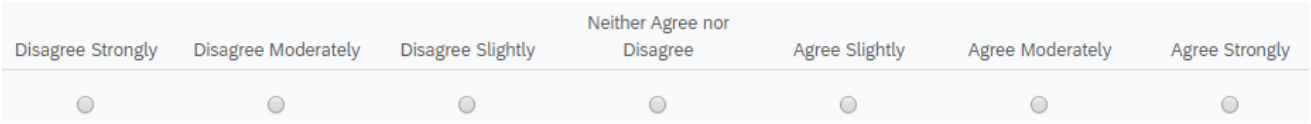
Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate?
I would recommend the candidate to other students who might be interested in finance.
I think that the candidate should be hired.
I think the lecturer is very skilled.

Based on the candidate's lecture, to what extent do you agree with the following statements about the content of the candidate's teaching?
The lecture material was intellectually stimulating.
The lecture material was of high quality.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's competence and expertise?
The candidate was professional.
The candidate was knowledgeable.
The candidate mastered the content of the lecture.

Based on the candidate's lecture, to what extent do you agree with the following statements about the candidate's teaching style?
The candidate was a warm person.
The candidate was an approachable person.
The candidate was enthusiastic.

[All assessed in 7-point Likert scale as below]



This completes the study. To be eligible for the bonus payment, please complete the following questionnaire about yourself.
[Same as in Study 3]