

Familiarity and sex based stereotypes on instant impressions of male and female faculty

Joel T. Nadler · Seth A. Berry ·
Margaret S. Stockdale

Received: 2 February 2012 / Accepted: 23 March 2013 / Published online: 19 April 2013
© Springer Science+Business Media Dordrecht 2013

Abstract To address the stranger-to-stranger critique of stereotyping research, psychology students ($n = 139$) and law students ($n = 58$) rated photographs of familiar or unfamiliar male or female professors on competence. Results from Study 1 indicated that familiar male psychology faculty were rated as more competent than were familiar female faculty, whereas unfamiliar female faculty were rated as more competent than unfamiliar male faculty. By contrast, in Study 2, familiarity had a stronger positive effect on competence ratings of female faculty than it did for male faculty. Among psychology students, familiarity increased sex bias against female faculty, whereas among law students familiarity decreased sex bias. Together, these studies call into question the stranger-to-stranger critique of stereotyping research. Our findings have direct implications for the context of student evaluations. In male-dominated disciplines it is important for students to be exposed to female instructors in order to reduce pre-existing biases against such instructors.

Keywords Stereotypes · Gender · Familiarity · College professors · Impression formation · Evaluations

J. T. Nadler (✉)
Department of Psychology, Southern Illinois University Edwardsville,
Edwardsville, IL 62026, USA
e-mail: jnadler@siue.edu

S. A. Berry
University of Notre Dame, Notre Dame, IN, USA

M. S. Stockdale
Indiana University Purdue University Indianapolis, Indianapolis, IN, USA

1 Introduction

Student evaluations of instructor performance are the primary source of teaching performance evaluation in colleges and universities (Sprague and Massoni 2005). Seldin (1999) found that 88.1 % of 4-year undergraduate universities relied on students' teaching evaluations in their performance reviews of instructors. These evaluations are often the basis for promotions and pay raises and are a factor in tenure decisions for faculty in academia (Sprague and Massoni 2005). Determining sources of bias in ratings in evaluations of college professors has important implications.

Women remain an under-represented group in academia, especially in higher-level positions. West (1995) found an increase of only 26–31 % female faculty from the years 1920–1995. In 2007, the National Center for Education Statistics reported that women comprised 41.8 % of faculty at public research universities but only 26.4 % of full professors across all types of higher education institutions (National Center for Education Statistics 2009). Comparatively, 52 and 55 % of non-tenure track instructor and lecturer positions respectively were held by women (National Center for Education Statistics 2009). In addition to this disparity in tenure status, female professors also receive less pay than male professors. West (1995) found that women who were full-professors earned only 88 % of male full professors' earnings. Considering these disparities and the widespread use of student evaluations of teaching, research has examined the extent to which students may be biased in favor of male professors.

Yet, the validity of applying stereotype research to personnel decisions has recently been called into question. Copus (2005) argued that having participants evaluate people who are unknown to them is an ecologically questionable research paradigm. Landy (2008) further argued that college students are naïve and untrained evaluators of work performance. Naïve raters (i.e., raters who are unaware of specific job-related characteristics) in a "stranger-to-stranger" paradigm would justifiably rely on stereotypes in their evaluations because they have no other information to rely upon. However, experienced raters, presumably such as supervisors, who are rating their own employees, would not need to resort to stereotypes for their judgments because they have sufficient individuated information on which to rely. Indeed, raters tend to rely less on stereotypes when they have job-relevant information about targets (Tosi and Einbender 1985).

We review the literature and theoretical explanations that apply to sex bias in evaluations of college professors. We include research on first impressions demonstrating that college students' initial impressions of a faculty member strongly predict their end-of-term evaluations as well as their decisions about whether to take a course from a given professor. We then describe two studies of psychology and law students' instant impressions of familiar or unfamiliar male and female faculty in their respective disciplines. The student evaluation setting is used to address Landy's (2008) concern that college students are naïve raters. Indeed, many college students may be naïve raters of fictitious employees (i.e., unaware of specific performance standards), but they are not naïve raters of college professors. College students frequently provide course and professor ratings during the course of their academic careers; therefore, they are not naïve raters. Our research, therefore, addresses important criticisms of stereotyping research and performance evaluations.

1.1 Sex bias in evaluations of college professors

Studies have found differences in evaluations of male and female professors (Basow and Silberg 1987; Centra and Gaubatz 2000; Winkler 2000) and in scenario research (i.e., research in which participants read or view a scenario involving hypothetical targets) evaluating class materials attributed to male or female professors (Abel and Meltzer 2007; Burns-Glover and Veith 1995). For example, Abel and Meltzer (2007) found that college students significantly rated a written lecture that was attributed to a woman lower than when the same lecture was attributed to a man. Studies of law students' evaluations also show more hostility toward female faculty than toward male faculty (Farley 1996). This bias has been attributed to law students' disappointment in having a female professor instead of the prototypical, masculine "Professor Kingsfield" of *Paper Chase* fame (Chamallas 2005). Although anti-female bias has been found in academia, researchers have also found pro-female bias in ratings of college professors (e.g., Rowden and Carlson 1996; Tatro 1995). Basow has demonstrated that the sex bias in teaching evaluations is often contextualized (Basow and Montgomery 2005). Male students, for example, tend to rate female instructors lower than do female students, and they rate female instructors lower than they rate male instructors (Basow 1995; Basow 2000; Basow and Silberg 1987). Female students often rate female faculty higher than do male students, and students in general tend to rate female faculty higher than male faculty on communication and interpersonal interaction criteria (Centra and Gaubatz 2000). Basow and Montgomery concluded that although many studies find no sex bias in teaching evaluations, when sex bias is found, it is usually against women.

1.2 Theoretical explanations

Substantial theory exists in the performance evaluation literature that explains sex bias and the conditions under which sex bias is more or less likely to occur. Status characteristics theory (Berger et al. 1972; Ridgeway 2001) states that perceptions of competence are driven by presumptions of status. Because men are typically accorded higher status than women, they are presumed to be more competent. Conditions that make gender less salient or that make a specific status more salient may mitigate pro-male bias. For example, female pre-school teachers may be rated as more competent than male pre-school teachers because the specific characteristic of pre-school teachers contextualizes perceptions of competence. Similarly, Heilman's lack of fit model (1983) of sex bias in performance evaluations predicts bias against women when the stereotype of the job does not match the stereotype of the person. Women tend to be rated lower than men in jobs that are perceived to be masculine, such as upper-level management and other male-dominated occupations (Heilman 1983, 2001). This bias diminishes when contextualized, job-relevant information is provided to raters (Heilman 2001). Finally, Eagly and Karau's role congruity theory (2002) provides a robust explanation for sex bias across many circumstances. Accordingly, sex bias occurs when role expectations conflict. In general, people expect women to enact communal role behaviors, such as being caring, nurturing, communicative, and understanding, and they expect men to enact agentic role behaviors, such as being self-reliant, intelligent, assertive, and commanding (Eagly and Carli 2007; Hale and Hewitt 1998).

When work role expectations are also agentic, such as in management, women are not perceived to be as competent as men because they are not perceived to possess the requisite role characteristics.

Role congruity theory (Eagly and Karau 2002) also explains why competent women in masculine roles tend to be categorically disliked. Such women are perceived to have adopted the requisite, agentic role characteristics, but in doing so they have violated expectations that they should be feminine and communal. Abundant research on perceptions of female managers and leaders confirms that women who enact masculine, agentic behavior as required by the work role are perceived to be competent but unlikable (Fiske et al. 1999; Gill 2004; Heilman 2001; Rudman and Glick 1999, 2001).

Whereas management is often regarded as agentic in its role expectations (Schein and Mueller 1992; Schein et al. 1989, 1996), it is less well known whether the characteristics associated with competent university teaching are consistent with agentic or communal traits. Bachen et al. (1999) illustrated the dichotomy of agentic and communal characteristics in students' perceptions of their instructors. Students rated male professors as being more professional, assertive, and controlling than female professors. However, female professors were seen as being more caring, flexible, and expressive than male professors. Bachen et al. (1999) further found that students rated their ideal professor as enthusiastic, intelligent, caring, flexible, organized, professional, and assertive, and that a caring and expressive teaching style was favored the most. Similarly, Sprague and Massoni (2005) found that students' perceptions of their best instructors were those who were caring, understanding, intelligent, helpful, interesting, and approachable. Therefore, it tentatively appears from these limited samples that students hold a blend of both agentic (intelligent, assertive) and communal (caring and expressive) role expectations for college professors. Nonetheless, there is evidence that students utilize gender stereotypes to criticize female instructors more than they do male instructors. Sandler (1991), Sandler et al. (1996) and Winkler (2000) provided evidence from student evaluations that suggested that female faculty members who enacted traditionally feminine role characteristics were viewed as *too feminine*, whereas women who possessed agentic traits prescribed to college professors were viewed as *too masculine*.

1.3 Bias in first impressions

The research examining sex bias in teaching evaluations has been predominately based on end-of-term teaching evaluations and surveys assessing teacher traits (Bachen et al. 1999; Basow 2000; Sprague and Massoni 2005). However, there are a handful of studies that have used a Goldberg paradigm (Goldberg 1968; Eagly and Carli 2003) in which students evaluate equivalent samples of teaching behavior attributed randomly to either a male or female professor (e.g., Abel and Meltzer 2007; Arbuckle and Williams 2003). Both Abel and Meltzer (2007) and Arbuckle and Williams (2003) found a significant pro-male bias in their respective samples of undergraduate students. It is possible, therefore, that gender differences in actual teaching evaluations may be partly a function of pre-existing or stereotype-driven biases against female professors. In order to assess the extent to which students are biased for or against female professors, regardless of any sample of teaching behavior, it is important to

examine students' initial impressions of male and female professors. Researchers who have studied first impressions have demonstrated that evaluators can fairly accurately judge teaching competence (Ambady and Rosenthal 1993; Babad et al. 2004) from viewing "thin slices" of nonverbal teaching performance. Thin slices detail obtaining ratings of 6–9 s of taped and muted teaching behavior correlate significantly with end-of-term teaching evaluations. Buchert et al. (2008) also found that teaching evaluations assessed after only 2 weeks of class predicted end-of-term teaching evaluations. Moreover, these "first impressions" were stronger predictors of performance than evaluations that were guided solely on students' knowledge of the instructor based only on their reputation (e.g., from other students or other sources of information about instructors, such as "ratemyprofessor.com"). Clayson and Sheffet (2006) found that end-of-term evaluations significantly correlated with evaluations taken after only 5 min of student-instructor interaction. These students had no knowledge of the instructor's pedagogical practice or grading style, but based evaluations primarily on first impressions. Nonetheless, this research has relied on actual samples of behavior, albeit brief, in their assessments of teaching performance and did not address issues regarding sex bias.

Our research examines a "thinner slice" than did Clayson and Sheffet (2006) and Buchert et al. (2008). We examined whether students have pre-existing bias for or against female professors based only on the picture of the professor without any sample of behavior. Based on theories of sex bias in performance evaluation (e.g., status characteristics theory, lack-of-fit model, and role congruity theory), we predicted that students would rate the pictures of male professors more favorably (i.e., as more competent and more desirable as a college instructor) than pictures of female professors. This prediction is based specifically on the proposition from status characteristics theory (Berger et al. 1972; Ridgeway 2001) that "male gender" is a high master status variable that pervades person perception unless specifically contextualized. According to Heilman's lack-of-fit model (1983), as well as Eagly and Karau's (2002) role congruity theory, individuals who are perceived to not possess the requisite personal characteristics for a specific role are perceived as less competent than those who are stereotypically perceived to fit the role.

Due to the fact that the legal profession is more male-dominated among student bodies, faculties, and practicing professionals, it is likely that the descriptive role expectations for law professors are more agentic and masculine than for the psychology profession where the student body, faculty, and practicing professionals are more female dominated. Various agencies have reported employment and labor statistics that offer support for this assertion. The Association of American Law Schools (Association of American Law Schools 2009) provided labor statistics that demonstrate the disparity between males and females in faculty positions in law schools. The AALS reported that 63 % of all law faculty in 2007 were men. The Bureau of Labor Statistics (2009) reported that in 2008, men held nearly 66 % of the professional law jobs. Although not as dramatic as the disparity in faculty and professional positions, the American Bar Association (2009) reported that of the 141,031 law students enrolled in JD programs in 2007, 53 % were men.

By comparison, Fino and Kohout (2009) reported that in 2007, 76 % of students who had recently completed doctoral degrees in psychology were women and that 77 % of

bachelor degree recipients in 2006 were women. Similarly, Finno and Kohout noted that 72 % of “Early Career” psychologists were women. Conversely, Finno and Kohout reported that in 2006, only 53 % of the total workforce in Psychology were women. However, this represents a greater percentage of women working as professional psychologists than professional lawyers. Therefore, given the stronger male-dominated and agentic nature of law professionals and faculty compared to psychology professionals and faculty, we predicted that pro-male bias would be stronger in evaluations of pictures of potential law school professors than of pictures of potential psychology professors.

Despite these predictions and the large body of research that supports these underlying theories, we are reminded of Copus’s (2005) and Landy’s (2008) claim that studies of sex bias in performance evaluation have very little bearing in the “real world.” Copus (2005) questioned the validity of sex bias research based on the use of non-work samples and the use of stranger-to-stranger experimental designs (e.g., college students rating fictitious employees). Landy (2008) further argued that in the “real world,” performance evaluators are typically well-trained supervisors who have extensive knowledge and experience with the employees they are rating, and thus are unlikely to rely on stereotypes. Fiske and Neuberg’s (1990) continuum model suggests that the reliance on most stereotypes reduce as individualizing information is obtained regarding a target. However, repeated exposure to information regarding a target’s social group membership can, in the absence of confounding relationships and friendship, result in a reduction in the reliance on individualizing information and increases the use of stereotypes (Smith et al. 2006).

We address Copus and Landy’s concern in two ways. First, although our research participants (raters) were college students, they were working in a context in which they had experience with the rating task (professor evaluations) and are considered legitimate evaluators of performance. Second, we manipulated whether the pictures of the professors were those with whom they were familiar (i.e., their own professors), or those with whom they were not familiar (pictures of professors from other universities). When rating familiar professors, students should rely less on stereotypes and more on their knowledge of the professor’s teaching reputation. However, when rating unfamiliar professors, student should rely more on gender stereotypes. Thus, we hypothesized that the pro-male bias in ratings of college professors would be stronger when rating unfamiliar professors than when rating familiar professors.

Our hypotheses are summarized below.

H1: Students will rate pictures of male professors more favorably than female professors in terms of competence and desirability as an instructor.

H2: Familiarity with the professor will interact with sex of the professor, such that there will be stronger pro-male bias in ratings of unfamiliar professors than in ratings of familiar professors.

H3a: This pro-male bias will be stronger among ratings of law professor pictures than among ratings of psychology professor pictures.

H3b: The Familiarity by Sex of Professor interaction will be stronger among ratings of law professors than among ratings of psychology professors.

2 Study 1

2.1 Methods

2.1.1 Participants

Participants for the main study consisted of 139 college students enrolled in psychology courses at a Midwestern university. The mean age of the participants was 23 years old with a range from 18 to 46. The sample consisted of 27 % men and 73 % women; 77 % European Americans, 17 % African Americans 2 % Asian Americans and 4 % Latino Americans (Hispanics); and there were no freshmen, 17 % sophomores, 37 % juniors, and 46 % seniors.

2.1.2 Materials and measures

Stimuli selection and pre-testing. An initial set of 18 photographs of Psychology Department faculty (9 men and 9 women) from the participants' institution were selected from the Department's website, with approval of the selected faculty. Selected photographs were of faculty who taught a high percentage of sophomore and junior level psychology courses to increase the probability that the faculty would be familiar to undergraduate college students. These photographs comprised the initial set of stimuli for the familiar condition. To create stimuli for the unfamiliar condition, photographs of 36 college professors were selected from psychology department web sites from other colleges and universities within the United States. Therefore, it is unlikely students would have any familiarity with these "unfamiliar" faculty. Care was taken to try to select at least two photographs of unfamiliar faculty members who were similar in age, race, and general attractiveness as each of the familiar faculty photographs. For example, if one of the familiar faculty members was in her mid-30s and had long straight hair, we tried to select photographs of two unfamiliar female faculty who appeared to be in their mid-mid-30s with long straight hair. Because the selected photographs of unfamiliar faculty were selected from publicly available websites which were not copyrighted and because these faculty members were not instructors at the participants' own institution, we did not seek permission to use these photographs in our study. Each photograph was a $2\frac{1}{2}$ " (W) \times 3" (L) color portrait. All of the faculty who were selected as stimuli were White/European American. This was done to avoid any potential bias effects that race/ethnicity may have had.

Three research assistants rated each photo for attractiveness and age. From these original ratings, the final set of stimuli consisted of 8 familiar faculty members (4 men and 4 women), matched with two unfamiliar faculty photographs per one familiar faculty photograph. Essentially, the unfamiliar faculty (i.e., faculty from different universities) were matched to familiar faculty (i.e., faculty from the affiliated university) based upon similar attractiveness and age ratings, in addition to general appearance perceptions. All 8 familiar faculty and their unfamiliar matches were rated as closely matching by all three assistants on attractiveness and age. Thus, the pre-testing sample of stimuli consisted of 8 familiar and 16 unfamiliar faculty photographs, with an equal number of male and female photographs in each set.

Table 1 Means of age and attractiveness by categories

	Age		Attractiveness	
	Mean	SD	Mean	SD
Unfamiliar men	36.37	2.86	4.27	1.27
Familiar men	35.51	2.92	4.46	1.08
Unfamiliar women	36.76	3.72	4.44	0.81
Familiar women	38.21	8.36	4.30	0.92

To pre-test these stimuli for perceived attractiveness and age, 105 college students (71 % female) selected from Psychology classes were given an informed consent form and a packet that contained all 8 familiar and 16 unfamiliar photographs. Participants rated all of the photographs presented sequentially allowing each familiar faculty member to then be compared to the two unfamiliar faculty members who had been pre-screened to be their matches. The attractiveness of each photograph was rated on 1 (*not at all attractive for a college professor*) to 7 (*very attractive for a college professor*) scale and participants were also asked to estimate the age of the person in the photograph.

Attractiveness ratings and perceived age estimates were aggregated in each condition [represented by the 2 (familiarity) \times 2 (professor sex) design] and then re-organized as a one-way (category) repeated measures MANOVA to examine aggregated attractiveness and age ratings. The data were aggregated because we were not concerned about individual differences within conditions, but wanted to make sure that on average, the photographs were similar. The means and standard deviations for the age and attractiveness ratings are provided in Table 1. There were no significant differences between the four categories on aggregated perceptions of age, $F(3, 102) = .83$, $p = .10$, and attractiveness, $F(3, 102) = .94$, $p = .11$.

Measures. For the main study, each photo was rated on five questions, the first three of which served as manipulation checks. First, familiarity was measured on a scale ranging with anchors 1 (*not familiar*), 2 (*somewhat familiar*), and 3 (*very familiar*). Second, participants were asked whether the target was a professor at the university (*yes* or *no*). Third, participants were asked to provide the target's name. Perceived competence was rated on a scale ranging from 1 (*very incompetent*) to 7 (*very competent*), and desirability of taking a class was measured on a scale ranging from 1 (*strongly disagree with wanting to take a class with the person*) to 7 (*very strongly agree with wanting to take a class with the person*). Familiarity, competence, and class desirability ratings of professors in each condition were averaged for each rating type (e.g., the competence ratings of the 4 familiar women were averaged to create a composite score). The demographic questionnaire assessed participants' age, gender, ethnicity, and year in school.

The decision to use single item measures of perceived competence and desirability of taking a class with the instructor was based on two considerations. First, we asked participants to make assessments of several professors, therefore the assessments needed to be brief in order to reduce survey fatigue. Second, there is a substantial body of research demonstrating that certain kinds of single-item measures

can be as valid as multi-item measures. Single item measures have been found to be both reliable and valid measures of teaching effectiveness (Wanous and Hudy 2001). Bergkvist and Rossiter (2007) found strong support for single-item measures performing equally well compared to multiple-item measures for singular concrete constructs such as products, brands, and advertisements. Even single-item measures of complex constructs such as self-esteem have been found to be highly related to multiple-item scales (Robins et al. 2001). Wanous et al. (1997) found support for the validity of using a single item measure for assessing constructs such as job satisfaction. Because our ratings were straightforward constructs, we concluded that single-item measures would suffice.

2.1.3 Procedures

Participants received extra credit in their psychology course for participating in this study, as did participants in the aforementioned pretest. Psychology courses with a high percentage of juniors and seniors were chosen to heighten the likelihood that the participants would be familiar with the psychology faculty photos used. To reduce survey fatigue and order effects, each participant received only half of the total pictures, creating a “packet” condition. Each packet contained 12 photographs consisting of two familiar male faculty and two familiar female faculty. For each familiar faculty photograph, the packet also contained photographs of the two unfamiliar faculty members who had been “matched” with them. The packets were created by randomly selecting two familiar female photographs and their respective unfamiliar faculty matches and randomly selecting two familiar male faculty and their matches to be in one or the other packet. Picture order was randomized in every packet. Participants were given a survey packet containing demographic questions, color photographs of male and female professors, and questions concerning each photo.

2.2 Results

Before testing the main hypotheses, manipulation checks and analyses testing for participant gender and packet effects were conducted.

2.2.1 Preliminary analyses

Pictures of faculty from the participants’ university were rated as more familiar ($M = 1.22$, $SD = 0.35$) than pictures of faculty who were not from the participants’ university ($M = 1.08$, $SD = 0.18$), $t(137) = 5.85$, $p < .001$, Cohen’s $d = 0.46$. Sex of the professor did not have a significant effect or interaction with familiarity. Only 12% of the familiar faculty were identified as instructors at the sample’s university compared to none of the unfamiliar faculty so identified. Thus, while participants perceived that the faculty stimuli from their own institution were more familiar than faculty stimuli from other institutions, they tended not to state that the familiar faculty member was from their own institution.

Although we did not hypothesize effects of participant gender or of packet on the competence and class desirability ratings, we checked for such effects. A 2 (participant gender) \times 2 (packet) \times 2 (familiarity) \times 2 (professor sex) mixed model ANOVA was conducted on competence ratings and on course desirability ratings. Familiarity and professor sex were repeated measures. For this analysis, each type of rating was averaged within condition. For example, the competence ratings were averaged for the two familiar male faculty stimuli in packet 1.

With regard to competence ratings, there was no main effect of participant gender, $F(1, 134) = 1.81, p = .18$; no significant participant gender \times familiarity interaction $F(1, 134) = .02, p = .90$; no significant participant gender \times professor sex interaction; and no participant gender \times familiarity \times professor sex interaction, $F(1, 134) = .29, p = .59$. In addition, there was no main effect of packet, $F(1, 134) = .15, p = .70$; no packet \times familiarity interaction, $F(1, 134) = 2.36, p = .13$; no packet \times professor sex interaction, $F(1, 134) = .07, p = .79$; and no packet \times familiarity \times professor sex interaction, $F(1, 134) = 1.28, p = .26$.

Similarly potential effects of participant gender and packet were examined. For class desirability ratings, there was no main effect of participant gender, $F(1, 134) = .36, p = .55$; no participant gender \times familiarity interaction, $F(1, 134) = .02, p = .88$; no participant gender \times professor sex interaction, $F(1, 134) = .18, p = .67$; and no participant gender \times familiarity \times professor sex interaction, $F(1, 134) = .58, p = .45$. Additionally, there was no main effect of packet on class desirability ratings, $F(1, 134) = .01, p = .93$; no packet \times familiarity interaction, $F(1, 134) = .21, p = .65$; no packet \times professor sex interaction, $F(1, 134) = .22, p = .64$; and no packet \times familiarity \times professor sex interaction, $F(1, 134) = .52, p = .47$.

In the remaining analyses, the data were aggregated over participant gender and packet, and both competence ratings and the class desirability ratings within a condition (for example: male, familiar professors) were averaged across the four photographs (from the two different packets) pertaining to that condition. Hypotheses 1 and 2 were tested using 2 (sex of professor) \times 2 (familiarity of professor: familiar or unfamiliar) repeated measures ANOVAs on perceptions of competence and class desirability. Alpha was adjusted to .025 using a Bonferroni correction to account for the two parallel analyses.

2.2.2 Competence

There was a significant main effect for familiarity on competence ratings, $F(1, 134) = 4.32, p = .04$, partial $\eta^2 = .04$. Familiar professors were rated lower ($M = 4.77, SD = .07$) than unfamiliar professors ($M = 4.87, SD = .07$). There was also a significant main effect for professor sex, $F(1, 104) = 5.80, p = .02$, partial $\eta^2 = .05$. Male professors were rated as more competent ($M = 4.89, SD = .07$) than female professors ($M = 4.75, SD = .07$).

These main effects were qualified, however, by a significant professor sex \times familiarity interaction, $F(1, 104) = 30.49, p < .001$, partial $\eta^2 = .23$. Tukey's post hoc tests were used to examine the interaction of professor sex on the familiar and unfamiliar conditions (see Fig. 1). Among unfamiliar professors, female pro-

Fig. 1 Study 1: familiar versus unfamiliar professors and male versus female professors on ratings of competence

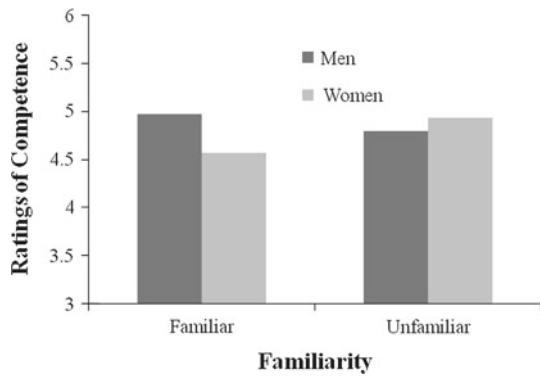
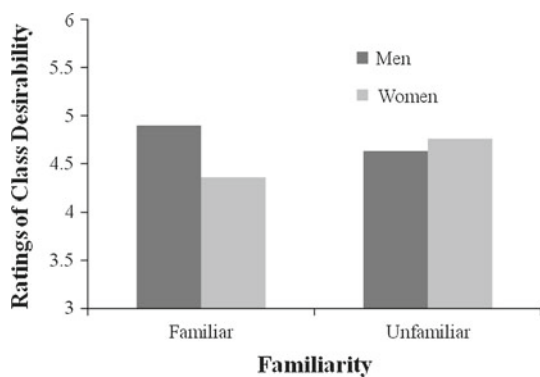


Fig. 2 Study 1: familiar versus unfamiliar professors and male versus female professors on ratings of class desirability



professors were rated as significantly more competent than male professors ($M_{\text{male}} = 4.80$, $SD = .67$ vs. $M_{\text{female}} = 4.93$, $SD = .78$). Among familiar professors, however, male professors were rated as significantly more competent than female professors, ($M_{\text{male}} = 4.97$, $SD = .91$ vs. $M_{\text{female}} = 4.57$, $SD = .78$).

2.2.3 Class desirability

The main effect of familiarity on class desirability ratings was not significant, $F(1, 104) = 1.04$, $p = .31$; however there was a significant main effect of professor sex, $F(1, 104) = 14.16$, $p < .001$, partial $\eta^2 = .12$. Participants desired to take classes more from male professors than female professors ($M_{\text{male}} = 4.76$, $SD = .87$ vs. $M_{\text{female}} = 4.56$, $SD = .71$). This main effect was qualified by a significant familiarity \times professor sex interaction, $F(1, 104) = 38.60$, $p < .001$, partial $\eta^2 = .27$.

Tukey's post hoc tests were used to examine the interaction of professor sex and familiar and unfamiliar conditions (see Fig. 2). In the unfamiliar condition, there was no significant difference in desire to take a course from either male or female professors, ($M_{\text{male}} = 4.63$, $SD = .73$; $M_{\text{female}} = 4.76$, $SD = .72$). In the familiar condition, participants significantly desired to take a course from familiar male faculty more than

familiar female faculty ($M_{\text{male}} = 4.90, SD = .94$ vs. $M_{\text{female}} = 4.36, SD = .93$). Men were rated as having more desirable classes when familiar than when unfamiliar. Conversely, women were rated as having more desirable classes when unfamiliar than when familiar.

2.2.4 Familiarity ratings as a predictor

Because the familiarity manipulation was not particularly strong (only 7% of familiar faculty were identified by name), we re-tested our central hypotheses by using the familiarity rating (originally used as a manipulation check) as a predictor of competence and class desirability ratings. We also examined the familiarity rating \times sex interaction on these outcome variables. To conduct these analyses, we ran Analyses of Covariance on competence and class desirability ratings separately entering the familiarity rating as a covariate and professor sex as a within subjects factor. For both outcome variables, the “main effect” of familiarity rating was not significant (competence ratings: $F(1, 69) = 1.66, p = .20$; class desirability ratings: $F(1, 69) = 2.34, p = .13$). Additionally, the familiarity rating \times professor sex interactions were not significant for either outcome (competence ratings: $F(1, 69) = .17, p = .69$; class desirability ratings: $F(1, 69) = .03, p = .87$).

2.3 Discussion

Psychology students rated male professors as more competent than female professors and desired to take classes from male professors than from female professors. However, this effect was modified by familiarity. The differences between male and female professors on ratings of competence and class desirability were smaller in the unfamiliar condition and favored women. These differences reversed in the familiar condition. Familiar male professors were rated significantly more competent and their classes as more desirable compared to familiar female professors. The results of Study 1 indicated that there was a small pro-female bias with unfamiliar professors, but a larger pro-male bias once students became familiar with professors. We failed to replicate these results, however, with participants' ratings of familiarity of the target faculty. Therefore, it is possible that the manipulated effects of familiarity were either weak or operating at an implicit level, such that they may not have been confident of recognizing familiar target faculty, but nonetheless knew something about them.

Due to the fact that Psychology, as a major and as a profession, is slightly female-dominated, participants who viewed unfamiliar faculty appeared to have relied on “lack of fit” criteria (Heilman 1983) in evaluating the competence of male and female professors. Specifically, participants may have assumed that female Psychology professors should be better instructors than male professors without knowing anything about them. However, when viewing familiar faculty, the bias did not disappear as expected; it actually reversed such that participants had more favorable reactions to their familiar male faculty than to their familiar female faculty. While there may have been individual differences among the set of familiar faculty used as stimuli in this study, such that male faculty were objectively better than female faculty, discussions

among the authors (who know these faculty members well) and with the Chair of the Psychology department where these faculty members are employed did not reveal any striking differences in competence or likability. Furthermore, aggregating ratings across three photographs within each condition should have reduced any idiosyncratic effects of any particular faculty member.

To further test the role of familiarity on reactions to male and female professors, we examined law school students' perceptions of law faculty—a more male-dominated major and profession. To the extent that students rely on lack of fit stereotypes in judging the competence and course desirability of law professors, we expected that students would rate unfamiliar male law professors higher than unknown female law professors. If such stereotyping is reduced when evaluating known professors, we expected sex of professor differences in the familiar condition to be less than in the unknown condition.

3 Study 2

3.1 Methods

3.1.1 Participants

Participants consisted of 58 professional law school students (66 % men) enrolled in required second and third year law courses at a Midwestern university. The mean age of the participants was 25 years old with a range from 22 to 40. There were 50 (86 %) European Americans, 3 (5 %) Asian Americans, 1 (2 %) Hispanic American, and 2 (3 %) who did not report their ethnicity.

3.1.2 Measures and procedures

The same process of pre-testing photographs matching familiar and unfamiliar professors on ratings of attractiveness and age was used from Study 1. An initial sample of 16 familiar professors (8 men and 8 women) along with a sample of 32 unfamiliar professors (16 men and 16 women) drawn from national University web sites were coded by a team of three raters. This resulted in a set of photographs determined to be matched in ratings of age and attractiveness. Of the final set of selected photos, 18 law school professors from law school web sites throughout the Midwest were selected, including 6 (3 male and 3 female) from the participants' own law school and 12 (6 male and 6 female) law faculty from other law programs. A sample of 121 first year law students was used to further pretest the 18 photographs. The familiar professors taught first year courses taken by the majority of law students. The same outcome questions and demographic questions described in Study 1 were used in Study 2. Although we allowed attractiveness and age of the stimuli to vary within a particular condition (e.g., among familiar male faculty), we wanted to be sure that the average level of attractiveness and perceived age was equivalent among the four categories (familiar men, familiar women, unfamiliar men, and unfamiliar women). Therefore, we averaged ratings of attractiveness and perceived age within conditions, re-organized the conditions [2 (sex

Table 2 Means of age and attractiveness by categories

	Age		Attractiveness	
	Mean	SD	Mean	SD
Unfamiliar men	44.14	3.86	4.14	0.77
Familiar men	46.79	4.05	4.23	0.99
Unfamiliar women	45.00	3.93	4.02	0.86
Familiar women	45.94	3.40	4.37	3.34

of professor) \times 2 (familiarity)] into a one-way design (category), and then conducted a one-way within subjects MANOVA on attractiveness and age ratings. The means and standard deviations are provided in Table 2. There was a significant effect of category on age perceptions $F(3, 108) = 10.76, p < .001, \text{partial}\eta^2 = .16$. However, the differences in ages covered a range of less than 3 years; therefore, this difference was not considered to be meaningful. Perceptions of attractiveness did not significantly differ across conditions, $F(3, 114) = .96, p = .19$.

Procedures were the same as Study 1 with the following exception. A research assistant announced the study to students in classes and distributed survey packets to students to complete in their classrooms. All participants were volunteers who did not receive any credit or reward for their participation. Additionally, participants rated all of the pictures in a single presentation packet.

3.2 Results

Preliminary analyses were conducted to examine the manipulation checks and effects of participant gender.

3.2.1 Preliminary analyses

The main effect of familiarity on familiarity ratings was significant, $t(57) = 24.20, p < .001, \text{Cohen's } d = 2.31$. Faculty from participants' own institution (familiar professors) were rated as more familiar ($M = 1.45, SD = 0.15$) than those from other institutions (unfamiliar professors) ($M = 1.10, SD = 0.15$). There was no main effect of professor sex on familiarity ratings. Familiar faculty members were correctly identified as participant's own instructors 63 % of the time compared to a 3 % rate of mis-identification for the unfamiliar faculty.

A 2 (participant gender) \times 2 (familiarity) \times 2 (professor sex) mixed model ANOVA on competence ratings and course desirability ratings was conducted to examine effects of participant gender. Familiarity and professor sex were repeated measures. There was no main effect of participant gender on competence ratings, $F(1, 55) = 1.89, p = .18$; no participant gender \times familiarity interaction, $F(1, 55) = .32, p = .58$; no participant gender \times professor sex interaction, $F(1, 55) = .32, p = .58$; and no participant gender \times familiarity \times professor sex interaction, $F(1, 55) = .32, p = .57$.

For class desirability ratings, the main effect of participant gender approached significance, $F(1, 55) = 4.89, p = .05, \text{partial } \eta^2 = .07$. Women tended to give

higher ratings ($M = 4.84$, $SD = 0.88$) than men ($M = 4.53$, $SD = 0.66$). There was no significant participant gender \times familiarity interaction, $F(1, 58) = .26$, $p = .61$; no participant gender \times professor sex interaction, $F(1, 55) = .05$, $p = .83$; and no participant gender \times familiarity \times professor sex interaction, $F(1, 55) = .08$, $p = .79$. Because the only significant effect (participant gender) was negligible, the remaining analyses were conducted to compare gender differences on ratings of competence and class desirability.

Hypotheses 1 and 2 were tested with 2 (Sex of Professor) \times 2 (Familiarity: Familiar professors vs. Unfamiliar professors) repeated ANOVAs with ratings of competence and class desirability as dependent variables. Alpha was adjusted to .025 to account for the analyses on two dependent variables.

3.2.2 Competence

The main effect of professor sex was not significant, $F(1, 56) = .14$, $p = .71$. The main effect of familiarity was significant, $F(1, 56) = 31.30$, $p < .001$, partial $\eta^2 = .34$, such that familiar professors were rated as more competent ($M = 5.11$, $SD = 0.11$) than unfamiliar professors ($M = 4.72$, $SD = 0.10$). This main effect, however, was qualified by a significant familiarity \times professor sex interaction, $F(1, 56) = 5.81$, $p = .02$, partial $\eta^2 = .09$.

Tukey's post hoc tests were used to examine the interaction between ratings of male and female professors in the familiar and unfamiliar conditions (see Fig. 3). In the unfamiliar condition, male professors were not significantly rated higher on competence than female professors ($M_{\text{male}} = 4.79$, $SD = .81$; $M_{\text{female}} = 4.64$, $SD = .78$). In the familiar condition, female professors were not significantly rated higher on competence than male professors ($M_{\text{female}} = 5.16$, $SD = .97$; $M = 5.05$, $SD = .91$). Competence ratings of familiar female professors were compared to unfamiliar female professors and likewise for familiar and unfamiliar male professors. There was a stronger effect of familiarity between female faculty compared to male faculty. Familiar female professors were rated significantly higher ($M = 5.16$, $SD = .97$) than unfamiliar females ($M = 4.64$, $SD = .78$). By comparison, familiar male professors ($M = 5.05$, $SD = .91$) were also significantly rated higher than unfamiliar male professors ($M = 4.80$, $SD = .82$), but the effect of familiarity was not as strong for ratings of male faculty as they were for ratings of female faculty.

3.2.3 Class desirability

The main effect of professor sex on ratings of class desirability was not significant, $F(1, 56) = .05$, $p = .83$. There was a significant main effect of familiarity on class desirability ratings, $F(1, 56) = 30.55$, $p < .001$, partial $\eta^2 = .35$. Familiar professors' classes were rated as more desirable ($M = 4.83$, $SD = .09$) than unfamiliar professors' classes ($M = 4.42$, $SD = .08$). This main effect was qualified by a significant familiarity \times professor sex interaction, $F(1, 56) = 10.79$, $p = .002$, partial $\eta^2 = .16$.

Tukey's post hoc tests were used to examine the interaction comparing class desirability ratings between male and female professors in the familiar and unfa-

Fig. 3 Study 2: men raters versus women raters and familiar versus unfamiliar professors on ratings of competence

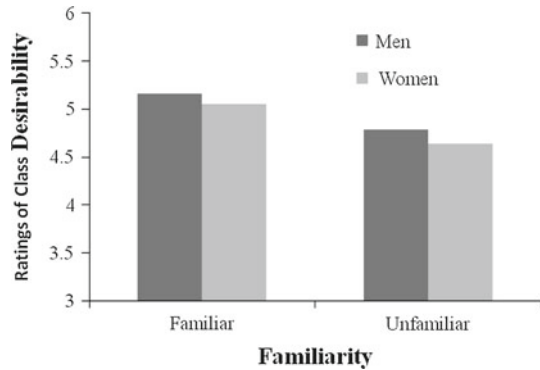
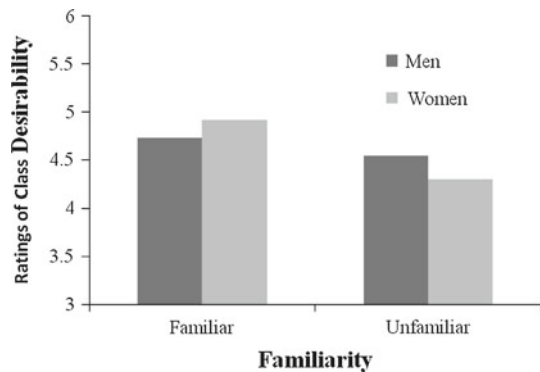


Fig. 4 Study 2: male versus female professors and familiar versus unfamiliar professors on ratings of class desirability



miliar conditions (see Fig. 4). In the unfamiliar condition, class desirability ratings were significantly higher for male professors ($M_{\text{male}} = 4.53$, $SD = .73$) than for female professors ($M_{\text{female}} = 4.30$, $SD = .72$). In the familiar condition, there was no significant difference for class desirability ratings for male or female professors ($M_{\text{male}} = 4.73$, $SD = .81$, $M_{\text{female}} = 4.91$, $SD = .70$). Familiar female professors were compared to unfamiliar female professors and likewise familiar male professors were compared to unfamiliar male professors. Among ratings of female professors, ratings of class desirability were significantly higher for familiar professors than for unfamiliar professors. Among ratings of male professors, the effect of familiarity was not significant with the trend for familiar male professors to receive higher ratings than unfamiliar male professors (see means above).

3.2.4 Comparisons between ratings of psychology and law professors

Hypotheses 3a and 3b respectively stated that the effects of sex of professor and the sex of professor by familiarity interaction would be stronger among law students rating law faculty than among psychology students rating psychology faculty. We did not test this hypothesis directly because the data came from two different studies with two different sets of stimuli. However, comparing the effect sizes the sex of professor and sex of professor \times familiarity interactions provides a quasi-assessment for this hypothesis.

The partial η^2 for the sex of professor main effects on competence (partial $\eta^2 = .05$) and class desirability (partial $\eta^2 = .12$) were stronger in the psychology sample as compared to the law school sample where these effects were not significant. The sex of professor \times familiar effect was stronger in the psychology study (Study 1) than in the law study (Study 2). In the psychology study, partial η^2 's were .23 (competence) and .27 (class desirability); in the law study, partial η^2 's were .09 (competence) and .16 (class desirability). The results were opposite to our stated predictions and thus, hypotheses 3a and 3b were not supported.

3.3 Discussion

In examining ratings of law professors, our central hypotheses were partially supported. A pro-male bias on ratings of class desirability was observed when participants rated unfamiliar professors, but this bias disappeared when participants rated familiar professors. For the competence ratings, the differences were not as dramatic, but a slight pro-male bias was found in the unfamiliar condition and a slight pro-female bias was found in the familiar condition. Moreover, familiarity had a stronger effect on increasing both competence and class desirability ratings for female professors than for male professors.

Heilman's lack-of-fit model (1983) appears to account for these findings. Because law is a male-dominated profession, without individuating information, participants may have relied on stereotypes to judge unfamiliar male law professors as more competent and for their classes to be more desirable than unfamiliar female law professors. When professors are known to participants, however, they should rely less on stereotypes and more on individuating information. Indeed, the results fit this pattern. Although it is possible that ratings of familiar faculty may have reflected job performance differences, care was taken to select photos of familiar law faculty who did not differ dramatically in job performance (as judged by one of the authors who completed a degree from the school of law from which photos of the familiar faculty were drawn and collected qualitative student comments regarding the professors). Also by aggregating competence and class desirability ratings across three different photographs within each condition, we reduced the impact of idiosyncratic differences among the professors in the familiar condition.

4 General discussion

Together our studies give a mixed review of the argument that knowing the target of one's impression reduces reliance on stereotypes and mitigates sex bias. Among psychology students, familiarity increased sex bias against female faculty, whereas among law students familiarity decreased sex bias. These biases were detected on ratings of mere photographs of the targeted faculty members, and research on "thin slices" demonstrates that initial impressions strongly predict end-of-term teaching evaluations (Ambady and Rosenthal 1993; Babad et al. 2004; Buchert et al. 2008; Clayson and Sheffet 2006). Thus, bias in initial impressions is not without consequence.

Nonetheless, familiarity with the target faculty did mitigate sex bias among law students rating law faculty, as predicted by those who espouse the importance of individuating information (Copus 2005; Tosi and Einbender 1985). There are several possible explanations for why sex bias was not abated when psychology students rated familiar faculty. First, psychology undergraduate students, compared to law students, are likely to not yet fully understand the illegality of sex discrimination and the importance of keeping one's biases attenuated. Second, psychology students at a fairly large public university have few interactions with psychology faculty and their interactions with faculty in general are dispersed across different departments across different locations around campus. By comparison, law students at this institution take all their courses by law faculty; all courses are taught in one building, and that building is located on the edge of campus. Law students rarely leave the building and therefore interact with the faculty throughout the day. Thus, law students were much more familiar with law faculty at this institution than were psychology students familiar with psychology faculty. Indeed, our manipulation checks showed that although psychology students rated faculty from their own institution as more familiar than faculty at other institutions, they could barely remember their names, nor could they reliably confirm that a particular person in a photograph taught at their institution. Therefore, the role of familiarity in mitigating sex bias was not sufficiently manipulated in Study 1.

We have proposed that Heilman's lack-of-fit model (1983) accounts for the forms of sex bias found in the unfamiliar conditions of both study 1 and study 2. Because psychology, as a profession and as a major, is more female dominated, without prior knowledge of particular faculty members, we would expect female faculty to be rated more favorably than male faculty because of their presumed better fit with the occupation. Similarly, without prior knowledge, law students would likely presume that male faculty are more competent in the male-dominated profession of law and legal education. Findings from both studies are consistent with this interpretation. To some extent, Eagly and Karau's role characteristics theory (2002) also accounts for the findings in the unfamiliar condition for the same reasons articulated above for the lack-fit-model. However, role characteristics theory goes further to predict negative reactions to people who violate the injunctive norms for a particular role. In this regard, we would have expected that familiar female faculty in the law school may have violated injunctive norms for the role of "woman" to the extent they may have enacted more masculine, agentic norms for the law professor role. In so doing, they may have evoked particularly negative reactions (see also Borgida and Rudman's discussion of violations of prescriptive stereotypes, Borgida et al. 2005; Rudman and Glick 1999). However, our findings did not support this proposition from role characteristic theory. Future research should continue to examine the conditions under which familiarity, performance appraisal experience, and job-relevant information help to mitigate sex bias, as well as the conditions under which violations of injunctive norms or prescriptive stereotyping maintains or increases sex bias.

4.1 Limitations and future research directions

In an effort to address criticisms that stereotype research has little validity for understanding the complexities of personnel evaluation, we attempted to manipulate a key

element of that complexity, familiarity, by capitalizing on a personnel evaluation task with which college students are very experienced: rating college instructors. Familiarity, however, is not an easy factor to manipulate. In order to minimize idiosyncrasies among the familiar faculty stimuli, we needed multiple stimuli within conditions (e.g., multiple familiar female faculty from a single discipline) as well as multiple control stimuli (unfamiliar faculty) who did not considerably vary in attractiveness from the experimental stimuli. Moreover, our research participants needed to be adequately familiar with the experimental stimuli in order for the familiarity manipulation to be meaningful; and, finally, we had to find settings (academic departments) where there were a sufficient number of male and female faculty to serve as potential stimuli and an adequate number of students within the department who could participate in the study. We were fortunate to locate and gain permission from two academic departments—one male dominated and the other female dominated—that ostensibly met these criteria. Nonetheless, our ability to adequately manipulate familiarity in the psychology department was not entirely successful as noted above. Therefore we cannot confidently assume that our findings from Study 1 are reliable.

Also, although our participating departments had fairly large numbers of students who could serve in our study, the samples sizes were still relatively small. Nonetheless, the effect sizes of our significant effects were fairly large. Therefore, the hypothesized effects appear to be fairly robust despite the fact that “familiarity” is a factor that by its nature is inherently fuzzy.

Finally, our study does not fully address the complexities of stereotyping research as it applies to the domain of real world personnel evaluations. Although college students do routinely evaluate faculty and those evaluations are consequential, they are not completely analogous to supervisory evaluations. Supervisors who engage in personnel evaluation are often trained and motivated to engage in accurate assessments (Landy 2008). Furthermore, they have much longer and more in-depth interactions with the subjects of their evaluation than do college students with their instructors (typically). These are important considerations for understanding the interplay between experience, training, motivation, familiarity and stereotypes as they affect personnel evaluation. However, to this end, the results of our study are dramatic, especially in Study 2 where we have more confidence in our findings. Law students’ familiarity with their instructors was sufficient to mitigate their tendency to view female faculty, compared to male faculty, as less competent and their courses to be less desirable.

We view our research as building a bridge between the tight control of laboratory research and the generalizability of field research. An analogous study might examine supervisors’ instant perceptions of their own employees compared to employees from other organizations. Similarly, comparisons could be made over time where supervisors rate the competence of newly hired employees, with whom they have little experience, and rate them again a year or so later after their familiarity with the employees has increased. We also encourage researchers to examine the underlying psychological processes that may be affected by familiarity and experience. We presumed that students who are familiar with their instructors relied less on gender stereotypes when making their competence and class desirability ratings than when they were rating unfamiliar instructors. Such processes could be measured directly in future research. Finally, we encourage researchers to examine boundary conditions for

familiarity and rating experience. In particular, familiarity may arouse other biases. Fiske and colleagues have shown, for example, that information that increases perceptions of women's competence can simultaneously reduce perceptions of women's warmth (and vice versa) (Fiske et al. 2002), yet both competence and warmth are considered to be important for selection and advancement in management positions (Heilman 2001).

4.2 Implications and conclusion

Our findings have direct implications for the context of student evaluations. In male-dominated disciplines it is important for students to be exposed to female instructors in order to reduce pre-existing biases against such instructors. In our study, the discipline of law was the male-dominated context. There is considerable attention being paid to women faculty in science, technology, engineering and math (STEM) disciplines, as evidenced by the National Science Foundation's significant funding of initiatives to increase the hiring and retention of female faculty in STEM disciplines (National Science Foundation 2009). Student evaluations of female faculty in STEM disciplines are an important component for their retention. Our findings also have important implications in the more general realm of personnel evaluation of women in nontraditional occupations. Factors that can increase familiarity with female employees should help to decrease bias against them.

In conclusion, we hope our study helps to stimulate efforts to understand when stereotypes do and when they do not impact performance evaluations, either in an academic or workplace context. The critiques leveled against stereotyping research merit serious consideration, but so too does the large body of well conducted research on stereotyping and bias against women. We look forward to a healthy flow of research that addresses these debates.

References

- Abel, M., & Meltzer, A. (2007). Student ratings of a male and female professors' lecture on sex discrimination in the workforce. *Sex Roles, 57*, 173–180.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441.
- American Bar Association. (2009). Lawyer demographics. Accessed at http://new.abanet.org/marketresearch/PublicDocuments/Lawyer_Demographics.pdf.
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*, 507–516. doi:10.1023/A:1025832707002.
- Association of American Law Schools. (2009). 2007–2008 AALS statistical report on law faculty. Accessed at <http://www.aals.org/statistics/2008dlt/titles.html>.
- Babad, E., Avni-Babad, D., & Rosenthal, R. (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. *Social Psychology of Education, 7*, 3–33.
- Bachen, C., McLoughlin, M., & Garcia, S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education, 48*, 193–210.
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology, 87*, 656–665.
- Basow, S. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 43*, 407–417.

- Basow, S. A., & Montgomery, S. (2005). Student ratings and professor self ratings of college teaching: Effects of gender and divisional affiliation. *Journal of Personnel Evaluation in Education*, 18, 91–106. doi:10.1007/s11092-006-9001-8.
- Basow, S., & Silberg, N. (1987). Student evaluations of college professors: Are male and female professors rated differently. *Journal of Educational Psychology*, 79, 308–314.
- Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 37, 241–255.
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44, 175–184. doi:10.1509/jmkr.44.2.175.
- Borgida, E., Hunt, C., & Kim, A. (2005). On the use of gender stereotyping research in sex discrimination litigation. *Journal of Law and Policy*, 13, 613–628.
- Buchert, S., Laws, E. L., Apperson, J. M., & Bregman, N. J. (2008). First impressions and professor reputation: Influence on student evaluations of instruction. *Social Psychology of Education*, 11, 397–408.
- Bureau of Labor Statistics. (2009). Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity. Accessed at <http://new.abanet.org/marketresearch/PublicDocuments/cpsaat11.pdf>.
- Burns-Glover, A., & Veith, D. (1995). Revisiting gender and teaching evaluation: Sex still makes a difference. *Journal of Social Behavior and Personality*, 10(6), 69–80.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 71, 17–33.
- Chamallas, M. (2005). The shadow of Professor Kingsfield: Contemporary dilemmas facing women law professors. *William and Mary Journal of Women and Law*, 11, 195–208.
- Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, 28, 149–160. doi:10.1177/0273475306288402.
- Copus, D. (2005). A lawyer's view: Avoiding junk science. In F. J. Landy (Ed.), *Employment discrimination litigation: Behavior, quantitative and legal perspectives* (pp. 450–462). San Francisco: Jossey-Bass.
- Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly*, 14, 807–834.
- Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Boston: Harvard Business School Press.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598.
- Farley, C. H. (1996). Confronting expectations: Women in the legal academy. *Yale Journal of Law and Feminism*, 8, 333–358.
- Fino, A. A., & Kohout, J. (2009). The future of the psychology workforce—statistics and trends. In *117th annual APA convention (conference)*. Toronto, Canada. August 07, 2009. Accessed January 12, 2010 at: <http://www.apa.org/workforce/presentations/2009-future-psychology-workforce.pdf>.
- Fiske, S. T., Cuddy, A. J. C., 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, 878–902.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York: Academic Press.
- Fiske, S., Xu, J., Cuddy, A., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, 55(3), 473–489.
- Gill, M. (2004). When information does not deter stereotyping: Prescriptive stereotypes can foster bias under conditions that deter descriptive stereotyping. *Journal of Experimental Social Psychology*, 40, 619–632.
- Goldberg, P. (1968). Are women prejudiced against women? *Transaction*, 5, 316–322.
- Hale, N., & Hewitt, J. (1998). Negative aging stereotypes and the agentic-communal dichotomy. *Perceptual and Motor Skills*, 87, 915–923.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, 5, 269–298.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organization ladder. *Journal of Social Issues*, 57, 657–674.

- Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology, 1*, 379–392. doi:10.1111/j.1754-9434.2008.00071.x.
- National Center for Education Statistics. (2009). Digest of education statistics, 2009. Accessed August 3, 2010 at http://nces.ed.gov/programs/digest/2009menu_tables.asp.
- National Science Foundation. (2009). ADVANCE: Increasing the participation and advancement of women in academic science and engineering careers (ADVANCE). Accessed August 3, 2010 at http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5383.
- Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues, 57*, 637–655.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin, 27*, 151–161. doi:10.1177/0146167201272002.
- Rowden, G. V., & Carlson, R. E. (1996). Gender issues and students' perceptions of instructors' immediacy and evaluation of teaching and course. *Psychological Reports, 78*, 835–839.
- Rudman, L., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology, 77*, 1004–1010.
- Rudman, L., & Glick, P. (2001). Prescriptive gender stereotypes and backlash towards agentic women. *Journal of Social Issues, 57*, 743–762.
- Sandler, B. R. (1991). Women faculty at work in the classroom, or why it still hurts to be a woman in labor. *Communication Education, 40*, 6–15.
- Sandler, B. R., Silverberg, L. A., & Hall, R. M. (1996). *The chilly classroom climate: A guide to improve the education of women*. Washington, D.C.: The National Association for Women in Education.
- Schein, V., & Mueller, R. (1992). The relationship between sex role stereotypes and requisite management characteristics: A cross-cultural look. *Journal of Organizational Behavior, 13*, 439–447.
- Schein, V., Mueller, R., & Jacobson, C. (1989). The relationship between sex role stereotypes and requisite management characteristics and college students. *Sex Roles, 20*, 103–110.
- Schein, V., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager—think male: A global phenomenon? *Journal of Organization Behavior, 17*, 33–41.
- Seldin, P. (1999). Current practices—good and bad—nationally. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 1–24). Boston, MA: Anker.
- Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., & Mackie, D. M. (2006). Familiarity can increase stereotyping. *Journal of Experimental Social Psychology, 42*, 471–478.
- Sprague, J., & Massoni, K. (2005). Student evaluations and gender expectations: What we can't count can hurt us. *Sex Roles, 53*, 779–793.
- Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education, 28*, 73–106.
- Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *The Academy of Management Journal, 28*, 712–723.
- Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. *Organizational Research Methods, 4*, 361–375. doi:10.1177/109442810144003.
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*, 247–252.
- West, M. S. (1995). Women faculty: Frozen in time. *Academia, 81*, 26–29.
- Winkler, J. A. (2000). Faculty reappointment, tenure, and promotion: Barriers for women. *The Professional Geographer, 52*, 737–751.

Author Biographies

Joel T. Nadler Ph.D., is an Assistant Professor of Industrial Organizational Psychology at Southern Illinois University Edwardsville. His research interests include stereotype based biases in selection and performance appraisal, sexual harassment, and social cognition of gender.

Seth A. Berry Ph.D., is a Survey Methodology Consultant in the Center for Social Research at the University of Notre Dame. His research interests include workplace sexual harassment, racial bias within organizations, and web-based survey methodology.

Margaret S. Stockdale (Ph.D. 1990, Kansas State University) is professor and chair of psychology at Indiana University-Purdue University at Indianapolis. Her research interests concern gender and diversity issues in the workplace and is particularly interested in sexual harassment and the experiences of women in non-traditional occupations, such as STEM, law, and the military.