Gender Biases and Evaluations: The Moderating Effect of Performance Level

Miranda L. Abild & Jane O’Reilly

University of British Columbia – Vancouver

[Accepted for Poster Presentation at the Society for Industrial and Organizational Psychology Conference, 2010 – Atlanta Georgia]
ABSTRACT

With the undeniable prevalence of workplace stereotyping, we examine the effect of gender biases on the evaluations between mediocre and above average performing individuals. We argue and find evidence demonstrating favor for men over equally qualified women as a function of perceived, though absent, differences in performance and qualifications.

PRESS PARAGRAPH

There are numerous theoretical explanations for discrimination in the workplace and support for the positive experience of white males found in leadership is wide. With Affirmative Action on the minds of many, empirical and objective examination of certain of the possible moderators of this status-based bias seemed appropriate. This research documents the effectiveness of evaluating performance level differences on workplace gender biases and supports that white males experience more positive evaluations than their female counterparts. This study is the first of a research program examining gender and racial discrimination in leadership positions.
It is an unfortunate aspect of society that people are judged not only on their abilities and competences but also based upon their skin color, their gender, their sexual orientation, and other seemingly nominal characteristic (Fiske, 1998). Nowhere is this stereotype-influence bias more prevalent than in the workplace. For decades, scholars and practitioners alike have been interested in understanding the influence of gender and racial biases on organizational interactions and experiences (Bielby, 2000; Agars, 2004; Snizek & Neil, 1992).

Much of this research has established that males, white males in particular, tend to have more positive experiences in organizations than their female or racial minority counterparts. For example, white males are often treated better by their colleagues (Aquino & Bommer, 2003; Fitzgerald, Gelfand & Drasgow, 1995). In another example, Heckman and colleagues found that customers, through customer satisfaction surveys, tend to evaluate male employees more positively than women and minority employees (Heckman, Aquino, Owens, Mitchell, Schilpzand & Leavitt, 2009). Finally, white men tend to be evaluated more positively than their female or racial minority colleagues (e.g., Eagly, Makhijani, & Klonsky, 1992; Ridgeway & Diekema, 1992).

While research has established that white males tend to be evaluated more positively and treated better than their female and racial minority colleagues, less research attention has been devoted to investigating whether performance level will influence this phenomenon. In other words, is the difference in evaluations between a mediocre-performing man and a mediocre performing individual of a minority group the same as the difference in evaluations between a high-performing man and individual of a minority group? In the current research we explore this research question starting with the evaluations of male and female job applicants in a leadership role.
Some scholars have argued that when women employees perform at outstanding levels, the difference in the evaluations between them and their male colleagues is eliminated (see Abramson, Goldberg, Greenberg, & Abramson, 1977). On the other hand, scholars have also argued that the opposite is true; that women employees who are above-average performers are actually devalued compared to their male counterparts because high performance is ‘extraordinarily’ uncharacteristic of a woman’s low-status (Nieva & Gutek, 1980). In the current research, we draw from social characteristics theory (Berger, Conner, & Fisek, 1974; Berger, Fisek, Norman, & Zelditch, 1977) which proposes that people are treated by others based on the social characteristics they possess that symbolize their social status and that high status individuals (e.g., white men) are treated better than low status individuals (e.g., women, racial minorities). Based on social characteristics theory, we argue that male leaders are evaluated more positively than female leaders in general, and that the difference between these evaluations will be greater when comparing the evaluations of high-performing leaders than mediocre-performing leaders.

THEORETICAL BACKGROUND AND HYPOTHESES

Evaluations of Men and Women

Past empirical work has established that men are often evaluated more favorably than women. Research that had investigated the different evaluations of men and women in leadership positions was empirically summarized in a meta-analysis by Eagly and colleagues (Eagly, Makhihani, & Kionsky, 1992). In their meta-analysis, the authors only included research that had methodologically controlled for the objective performance of the men and women evaluation targets. For example, some studies gave participants a vignette of a supervisor in a conflict situation and asked participants to evaluate the supervisor’s behavior (e.g., Rosen & Jerdee,
The only difference across the conditions was the gender of the supervisor in the vignette. In their meta-analysis, Eagly and colleagues (Eagly et al., 1992) found that the positive evaluation bias towards men was even greater in more masculine-oriented contexts. Several theories have been offered to explain the positive bias men are afforded in such situations. Much of this research has argued that gender-role based expectations of women’s behavior and abilities negatively bias evaluations of women compared to men (e.g., Eagly, 1987; Gutek & Morasch, 1982; Nieva & Gutek, 1981). However, these theories do not clearly predict whether performance-level (e.g., whether one performs at a mediocre or above-average level) will moderate perceptions of men and women, particularly in a leadership role. To answer our research question we turn to status characteristics theory (Berger, et al., 1974; 1977). Recent research that has applied status characteristic theory to understanding the more positive experience men enjoy in organizations has found that men are evaluated (Heckman et al., 2009) and treated (Aquino & Bommer, 2003) better across all contexts, not just leadership contexts. Based on this theory, we can also make specific predictions about the nature of the relationships between the performance level of men versus women and others’ evaluations.

*Social Characteristics Theory*

Social characteristics theory proposes that people interact with one another differently depending on their social status (Berger, et al., 1974; 1977). Social status influences how people make inferences about others’ underlying attributes, and people tend to make more favorable inferences about high status individuals. Specifically, compared to low status individuals, the positive behaviors (e.g., high performance, prosocial behavior) of high status individuals are more strongly attributed to underlying individual traits and capabilities. Consistent with social characteristic theory, high status individuals are seen as more intelligent, competent and
attractive than their low status counterparts (Georgeson & Harris, 1998). In other words, people have a strong tendency to perceive high-status individuals as better than low-status individuals and as a result, high status individuals are often treated with more respect and civility (Aquino & Bommer, 2003).

Furthermore, membership in salient social categories, such as gender and ethnicity, can indicate one’s status in society and in organizations (Berger et al., 1974; 1977; Pettigrew & Martin, 1987). While perceptions of the social status of women have gradually improved over the last few decades (Zweigenhaft & Domhoff, 1998), gender is still a salient characteristic of social class and women are generally seen as having less social status than men (Berger et al., 1974; 1977; Ridgeway & Diekema, 1992). Scholars have argued that these perceptions are influenced by the occupational roles women have traditionally held. Historically, women have held positions of little power and influence and with little supervisory responsibilities (Kanter, 1977). The traditional role of women in low-status occupational roles has reinforced the stereotypic perception of women as low status members of society and organizations (Ely, 1995).

Based on the propositions of status characteristics theory, we would expect that positive biases for male leaders will be greater when comparing high-performing male and female leaders than when comparing mediocre-performing male and female leaders. Since people tend to attribute high performance to the underlying traits and capabilities of high status individuals but not for low status individuals, the difference in the inferences people make about high- and low-status individuals’ abilities should become more salient at higher levels of performance. Based on this theory we propose the following hypotheses.

*Hypothesis 1: There will be an interaction between gender of an evaluation target and performance level of an evaluation target such that the relationship between*
performance-level and evaluation will be stronger for men than for women.  

Hypothesis 2: The difference in evaluation between men and women will be greater at a high-performance level than at a mediocre-performance level.

METHOD

Participants

Data were collected at a large Western Canadian university. Participants consisted of undergraduate students. Seventy students participated in this research, however, four were removed from the analysis because they incorrectly recalled the gender of the evaluation target (details below). The final sample included 66 participants, 69.7% were female, the average age was 22.82 (sd = 4.87), 79.7% were of Asian descent, and 18.8% were of Caucasian descent. On average, participants had 4.27 (sd = 4.83) years of work experience.

Study Procedures

Participants were recruited through an online subject pool and received an honorarium in exchange for their participation. Data was collected using surveys. Participants were asked to assume the role of a Human Resource Manager at a privately owned Accounting firm and told they were responsible for hiring a job applicant (evaluation target) for an Accounts Manager position. This occupational role was described to participants as a leadership position that required a candidate to possess effective leadership qualities. They were then given the information about the evaluation target. The study followed a 2 (target performance level: mediocre or above average) X 2 (target gender: male or female) between subjects design. Participants were randomly assigned individually to one of the four conditions. Participants were told that they were to review the information for 3 minutes and then answer a series of questions about their perceptions of the evaluation target.
The information package provided to the participants included a performance evaluation form and photograph of the evaluation target. These photographs were pretested in a pilot study (N = 55) to ensure that the individuals in the photos were perceived to be relatively similar in terms of age and attractiveness. Performance level was manipulated through the performance evaluation form. These performance evaluation forms were based on forms used in a real-life organizational setting. They contained information on eight “performance areas” (e.g., Leadership, Relationship with Others, Communication Skills, etc.) in the form of percentile scores and supervisor comments. In the mediocre performance condition the evaluation target received scores ranging from the 50th to 62nd percentile (which averaged to a percentile score of 58.25 across the different performance areas). In the above-average performance condition the evaluation target received scores ranging from 75th to 87th percentile (which averaged to a percentile score of 83.25). All supervisor comments on the evaluation forms were identical across the two conditions. After 3 minutes the experimenter took the performance evaluation forms and photos of the evaluation targets away and gave participants a questionnaire containing questions about their hiring recommendations and personal demographics.

Measures

Manipulation Checks. We assessed the effectiveness of our performance level manipulation by asking participants two questions: “This applicant’s performance, in general, is above average” and “This applicant’s performance, in general, is mediocre.” These items were measured on a 5-pt Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). To ensure that participants had followed the study instructions and paid attention to the experimental stimuli we also asked participants to recall the gender of the evaluation target at the end of the
questionnaire. Four participants did not correctly recall the evaluation target’s gender and were subsequently removed from the analyses.

*Recommendation.* To assess participants’ evaluation of the targets, we asked two questions. First, we asked participants whether they would recommend that their company hire the evaluation target. This item was measured on a dichotomous scale (0 = no; 1 = yes). We then asked participants the extent to which they were confident in their recommendation, ranging from 1 (*I am extremely UNSURE about my recommendation*) to 10 (*I am extremely SURE about my recommendation*). We then multiplied these two items to create our measure of recommendation.

*Control variables.* We controlled for participants’ gender. Past research has established that the gender of the perceiver is likely to influence the perceptions of suitability of the applicant for hire in a leadership position. Specifically, men tend to construe effective leaders in more masculine terms and thus generally prefer male leaders over women leaders (Eagly & Karau, 2002). Additionally, women, in general, give more positive reviews of other women compared to men (Winquist, Mohr, & Kenny, 1998). We also controlled for participants’ perceptions of the attractiveness of the evaluation target. Past research has indicated that, in general, people evaluate others they find more attractive more positively (Dion, Berscheid, & Walster, 1972; Surawski & Ossoff, 2009). We measured perceived attractiveness by asking participants to rate the evaluation target’s attractiveness on a 9-pt Likert scale, ranging from 1 (*Unattractive*) to 9 (*Attractive*).

*Results*

We first tested whether our manipulation of performance level was effective. We conducted an ANCOVA and controlled for evaluation target gender. Participants rated the
performance of evaluation targets in the mediocre performance condition as less “above average” 
\(M = 3.56, sd = .11\) than the performance of the evaluation targets in the above average conditions 
\(M = 3.94, sd = .12, F(1,63) = 5.22, p < .05\). Participants also rated the performance of evaluation targets in the mediocre performance condition as more mediocre \(M = 3.24, sd = .15\) than the performance of the evaluation targets in the above average conditions \(M = 2.40, sd = .15, F(1,63) = 15.34, p < .01\) Based on these results, we concluded that the manipulations produced their intended effects. The means, standard deviations, correlations and reliabilities are provided in Table 1.

Hypothesis 1 predicted an interaction effect between evaluation target gender and performance. We tested for this interaction effect using hierarchical linear regression (Aiken & West, 1991). We entered the control variables in the first step of this analysis, followed by the conditions (performance level and gender of evaluation target), and the interaction term. The results of this analysis supported a significant interaction between evaluation target performance level and gender \(\beta = .45, p < .05\). The results of this analysis are presented in Table 2. To investigate the nature of this interaction we looked at the relationship between performance level and recommendation separately for male evaluation targets and female evaluation targets. Based on this analysis there was a significant and positive relationship between performance and recommendation for male targets \(\beta = .41, p < .05\) but an insignificant relationship for female targets \(\beta = -.10, ns\). Thus, hypothesis 1 was supported. Figure 1 illustrates the nature of this interaction.

Hypothesis 2 predicted that the degree of difference between the evaluations for male and female evaluation targets would be greater at high levels of performance compared to mediocre levels of performance. We conducted an ANCOVA follow-up analysis to assess the degree of
mean difference between the recommendation scores for male and female targets at mediocre and high levels of performance. There was no significant difference between male \((M = .62, sd = .10)\) and female \((M = .80, sd = .11, F(1,27) = 1.55, ns)\) target evaluations in the mediocre condition. However, in the above average performance condition, males received a significantly better evaluation \((M = 1.08, sd = .05)\) than female evaluation targets \((M = .68, sd = .06, F(1,25) = 22.44, p < .01)\). Thus, hypothesis 2 was also supported.

**DISCUSSION**

Past research has established that men are often evaluated more positively than women (e.g., Eagly et al., 1992; Heckman et al., 2009). Drawing from social characteristics theory (Berger et al., 1972; 1977) we argued that this form of gender discrimination is likely to be greater at higher levels of performance. This discrimination is due to the different inferences that people make about the nature of high status, versus low status, individuals’ positive behavior (e.g., performance) which is more salient when those behaviors are more positive (e.g., above-average versus mediocre performance). We also predicted that the relationship between performance level and evaluation would be more predominant for men than for women in a leadership position.

In an experimental study we found support for our hypotheses. Participants had more positive evaluations of male job applicants than female job applicants at high levels of performance. There was no statistically significant difference in participants’ evaluations between men and women at mediocre levels of performance. Furthermore, high levels of performance did not seem to increase participants’ evaluations of women job applicants as there was no statistically significant positive relationship between job performance and evaluations for the women job applicants.
The results of this study offer several theoretical contributions to the current literature. First, consistent with other recent and past research, this research underscores the more positive experience men have in organizational settings compared to their female counterparts. This research is consistent with Heckman and colleagues’ (2009) recent findings that men receive better evaluations from customers as the quality of their work increases while this same positive relationship between quality of work and evaluation does not occur for women (or ethnic minorities). Our results, along with the results of the Heckman et al (2009) research, provide replicated support for the predictions of status characteristics theory on the discriminatory evaluations of women.

Another significant theoretical implication of this research is that it calls into question the theoretical completeness of the gender-role based theories that attempt to explain why gender discrimination occurs (e.g., Eagly, 1987; Eagly & Karau, 2002). These theories argue that women are discriminated against because the expectations people have of a woman’s performance and behavior are based upon the traditional societal or cultural roles of women, and do not fit with their general representation of what makes an ‘effective’ leader. These theories argue that a positive evaluation bias towards men will be greater in occupational roles that are generally described in more masculine terms (e.g., leadership or supervisory occupational roles) and make no specific predictions about the influence of performance level on gender discrimination. However, in the current study, we find that there is no positive evaluation bias afforded to men, compared to women, at mediocre levels of performance for a leadership role, supporting our status-based explanation for gender-bias evaluations of men and women leaders. Our research does not necessarily reject the propositions of the gender-role based explanations of gender bias, but it does suggest that, at the very least, both status- and gender-role based
explanations play roles in the differences observed. Future research should be devoted to understanding the interactive effects of status and gender roles on gender discrimination in workplace evaluations.

Finally, this research also has practical implications. These results indicate that gender bias towards women leaders is still alive and well in today’s society and this bias may contribute to the lack of female representation in leadership positions. For example, currently worldwide 17.3% of Upper House or Senate positions are filled by women (IPU, 2009). Furthermore, 28 out of 1000 of the Fortune 1000 leaders are women (Fortune Magazine, 2009). Finally, there is a large under-representation of women and racial minorities among the ranks of highly paid managers and professionals within American corporations (e.g., Baldi & McBrier, 1997; Eagly & Karau, 2002; Wilson, Sakura-Lemessy & West, 1999). We explain these biases through theoretical characteristics to suggest possible causes of gender bias though they also suggest potential corrective measures. By redefining organization positions and utilizing neutral terms, Gupta, Turban and Bhawe (2008) find the nullification of stereotypes possible. From a managerial point of view, our findings cast doubt on the efficiency of a sole evaluator in making accurate hiring recommendations. Although alternatives may appear unfavourable, our findings present empirical evidence for the white male bias and suggest that alternatives are necessary.

The current research paradigm contained several methodological strengths that helped to maintain the validity of our results and interpretations. Most importantly, we used an experimental paradigm that controlled for objective levels of performance so that we could examine the evaluation biases against women when they perform the exact same behaviors as their male counterparts. However, like all research, the current research is subject to limitations. Most notably, we used an undergraduate sample. While the participants we used in our sample
had a significant degree of past work experience, it is possible that our sample is predisposed to certain biases that are not present in a more mature sample. Secondly, our sample contained a disproportionate degree of ethnicities which may have also lead to certain unaccounted for biases. Finally, while the current study replicates the results found in Heckman et al (2009), future work is still needed to further replicate these results across a variety of samples and methodologies in order to verify the reliability of our conclusions.

This study is the first in a planned research program that will look further into the status-based biases that influence evaluations of leaders. To address the limitations stated above we plan on conducting further field and experimental studies using working-adult samples in order to replicate our findings using more mature and experienced participants. Also, we plan on applying this status-based approach to discrimination to study the potential biased evaluations people have of racial minorities in a leadership position. We plan on presenting the results of the current study along with the results of these future studies at SIOP 2010.
References


Table 1

*Means, Standard Deviations and Correlations*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender (of participant)</td>
<td>.30</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Attractiveness</td>
<td>5.82</td>
<td>1.83</td>
<td>-.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Performance Level</td>
<td>.48</td>
<td>.50</td>
<td>-.11</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Target Gender</td>
<td>.55</td>
<td>.50</td>
<td>-.06</td>
<td>-.23</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td>5. Recommendation</td>
<td>6.05</td>
<td>3.23</td>
<td>-.15</td>
<td>-.38**</td>
<td>.28*</td>
<td>-.03</td>
</tr>
</tbody>
</table>

*Note: * p < .05, ** p < .01. N = 66. Performance level coded such that mediocre (0) and above average (1). Target gender and participant gender are coded such that female (0) and male (1).*
Table 2

Regression Analysis Results Predicting Recommendation

<table>
<thead>
<tr>
<th>Predictor Variables</th>
<th>Step 1</th>
<th></th>
<th>Step 2</th>
<th></th>
<th>Step 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>t</td>
<td>β</td>
<td>t</td>
<td>β</td>
<td>t</td>
</tr>
<tr>
<td>Gender (of participant)</td>
<td>-.10</td>
<td>-.82</td>
<td>-.07</td>
<td>-.59</td>
<td>-.00</td>
<td>-.03</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>.37</td>
<td>3.17**</td>
<td>.36</td>
<td>3.00**</td>
<td>.39</td>
<td>3.35**</td>
</tr>
<tr>
<td>Performance Level (PL)</td>
<td>.22</td>
<td>1.94</td>
<td>-.05</td>
<td>-.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target Gender (TG)</td>
<td>.06</td>
<td>.47</td>
<td>-.18</td>
<td>-1.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL x TG</td>
<td></td>
<td></td>
<td>.45</td>
<td>2.20*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model R²                     | .16**  |          | .21**  |          | .27**  |
Adj. R²                      | .13    |          | .16    |          | .21    |
Δ R²                         | .05    |          | .06*   |          |

*Note: * p < .05, ** p < .01, Performance level coded such that mediocre (0) and above average (1). Target gender and participant gender are coded such that female (0) and male (1).
Figure 1

The Nature of the Interaction between Evaluation Target Performance and Gender Predicting Recommendation