Performance evaluations of women and men in stereotypically gender-congruent and gender-incongruent occupations: a consideration of both situational and individual difference variables.

Marlo C. Henderson
University of Massachusetts Amherst
PERFORMANCE EVALUATIONS OF WOMEN AND MEN IN STEREOTYPICALLY GENDER-CONGRUENT AND GENDER-INCONGRUENT OCCUPATIONS: A CONSIDERATION OF BOTH SITUATIONAL AND INDIVIDUAL DIFFERENCE VARIABLES

A Thesis Presented

by

MARLO C. HENDERSON

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

September 2001

Clinical Psychology
PERFORMANCE EVALUATIONS OF WOMEN AND MEN IN STEREOTYPICALLY GENDER-CONGRUENT AND GENDER-INCONGRUENT OCCUPATIONS: A CONSIDERATION OF BOTH SITUATIONAL AND INDIVIDUAL DIFFERENCE VARIABLES

A Thesis Presented

by

MARLO C. HENDERSON

Approved as to style and content by:

Bonnie R. Strickland, Chair

William F. Chaplin, Member

Norman Simonson, Member

Melinda A. Novak, Chair

Department of Psychology
## CONTENTS

| LIST OF TABLES                                | vi  |
| LIST OF FIGURES                              | vii |

### CHAPTER

1. PERFORMANCE EVALUATION AND BIAS...

   - Introduction........................................................................................................... 1
   - Revisions to Traditional Approaches of Studying Bias and Prejudice................. 1
   - The Advent of Social Cognition.................................................................. 3
   - Rationale for the Present Study................................................................ 6
   - Hypotheses...................................................................................................... 8

2. METHOD.................................................................................................................... 9

   - Overview of the Experimental Design................................................................. 9
   - Participants........................................................................................................ 10
   - Measures............................................................................................................ 10
     - Experimental Evaluational Styles Questionnaire (EESQ)................................. 10
     - Manipulation Check....................................................................................... 11
     - Ambivalent Sexism Inventory (ASI)................................................................ 12
     - Personal Attributes Questionnaire (PAQ)....................................................... 12
     - Male Role Norms Scale (MRNS)....................................................................... 12
   - Procedure.......................................................................................................... 13

3. RESULTS.................................................................................................................. 14

   - Manipulation Check............................................................................................ 14
   - Rating Agreement............................................................................................... 14
   - Derivation and Psychometric Evaluation of the Dependent Variables............ 15
     - The effect of the standards on evaluations.................................................... 15
     - The structure of the EESQ............................................................................ 15
     - Derivation of the EESQ scales....................................................................... 16
Psychometric Evaluation of the Individual Difference Variables ........................................... 18

The Influence of Either the Situational or Individual Difference Variables on Evaluations ........................................... 18

The effect of the situational variables on evaluations ......................................................... 19

The relation between the individual difference measures and evaluations ................................. 19

Combining Individual Difference and Situational Variables to Predict Evaluations ......................... 20

Do men and women generally evaluate the performance of men and women in traditional and nontraditional jobs differently? ................................................................. 21

Do men and women who differ on Sexism, Masculine Role Norms, or Masculinity-Femininity show biased evaluations toward male and/or female targets in masculine or feminine jobs? ................................................................. 22

Graphing interactions of continuous x categorical variables .................................................. 23

Benevolent Sexism as a moderator ................................................................. 24

Male Toughness as a moderator ................................................................. 26

Male Status as a moderator ................................................................. 31

Masculinity as a moderator ................................................................. 33

4. DISCUSSION ................................................................................. 36

Abbreviated Summary of Results ................................................................................. 36

Findings of Particular Interest ................................................................................. 36

General Differences in Participants’ Evaluations of Male and Female Targets ................................................. 36

Positive and Negative Bias, Moderated by Male Toughness, Status, and “Masculinity,” Toward Targets Working in Traditional and Nontraditional Occupations ................................................. 38

Negative bias of traditional participants toward targets working in nontraditional occupations ................................................. 38

Positive bias of nontraditional women toward targets working in nontraditional occupations ................................................. 39

Positive bias of traditional women toward targets working in nontraditional occupations ................................................. 40

Negative bias of nontraditional women toward female targets in nontraditional occupations ................................................. 41
Positive and negative bias of nontraditional men toward female targets in traditional occupations and male targets in nontraditional occupations, respectively. .......................... 41

Implications and Conclusions ................................................. 44

APPENDICES

A. MEASURES ........................................................................ 49
   EESQ .............................................................................. 49
   Manipulation Check ......................................................... 57
   ASI .................................................................................. 58
   MRNS ............................................................................. 59
   PAQ ................................................................................. 60

B. TABLES ............................................................................ 61

C. FIGURES .......................................................................... 67

BIBLIOGRAPHY ..................................................................... 79
LIST OF TABLES

Table                                                                                       Page
1. Number of Participants included in the Final Sample................................................61
2. Means and Standard Deviations for the 27 Evaluative Items in Descending Order..................62
3. Means, Standard Deviations, Coefficient Alphas, and Correlations Among the EESQ General Scale and the Valence Scales....................................................63
4. Means, Standard Deviations, and Coefficient Alphas for the Subscales from the Ambivalent Sexism Inventory (ASI), Masculine Role Norms Scale (MRNS), and Personal Attributes Questionnaire (PAQ).................................................................64
5. Means and Standard Deviations for the General and Negative Evaluational Bias Scores for each of the Target x Job Conditions......................................................65
6. Correlations Between the Individual Difference Variables and the Five Bias Measures.....................66
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Plot of the Cell Means from the $2 \times 2 \times 2$ (Gender $\times$ Target $\times$ Job) Analysis of Variance of the Evaluative Ratings Based on the Negative Evaluative Information</td>
<td>67</td>
</tr>
<tr>
<td>2. Plot of the Predicted Values from the Gender $\times$ Target $\times$ Job $\times$ Benevolent Sexism Hierarchical Regression Analysis of Evaluative Ratings based on Positive Evaluative Information</td>
<td>68</td>
</tr>
<tr>
<td>3. Plot of the Predicted Values from the Target $\times$ Job $\times$ Benevolent Sexism Hierarchical Regression Analysis of the Men’s Evaluative Ratings Based on Positive Evaluative Information</td>
<td>69</td>
</tr>
<tr>
<td>4. Plot of the Predicted Values from the Gender $\times$ Target $\times$ Job $\times$ Toughness Hierarchical Regression Analysis of Evaluative Ratings based on General Evaluative Information</td>
<td>70</td>
</tr>
<tr>
<td>5. Plot of the Predicted Values from the Gender $\times$ Target $\times$ Job $\times$ Toughness Hierarchical Regression Analysis of Evaluative Ratings based on Negative Evaluative Information</td>
<td>71</td>
</tr>
<tr>
<td>6. Plot of the Predicted Values from the Target $\times$ Job $\times$ Toughness Hierarchical Regression Analysis of Men’s Evaluative Ratings Based on Negative Evaluative Information</td>
<td>72</td>
</tr>
<tr>
<td>7. Plot of the Predicted Values from the Gender $\times$ Target $\times$ Job $\times$ Toughness Hierarchical Regression Analysis of Evaluative Ratings based on Mixed Evaluative Information</td>
<td>73</td>
</tr>
<tr>
<td>8. Plot of the Predicted Values from the Target $\times$ Job $\times$ Toughness Hierarchical Regression Analysis of the Women’s Evaluative Ratings Based on No Evaluative Information</td>
<td>74</td>
</tr>
<tr>
<td>9. Plot of the Predicted Values from the Gender $\times$ Target $\times$ Job $\times$ Status Hierarchical Regression Analysis of Evaluative Ratings based on Positive Evaluative Information</td>
<td>75</td>
</tr>
<tr>
<td>10. Plot of the Predicted Values from the Target $\times$ Job $\times$ Status Hierarchical Regression Analysis of the Men’s Evaluative Ratings Based on Positive Evaluative Information</td>
<td>76</td>
</tr>
</tbody>
</table>
11. Plot of the Predicted Values from the Target x Job x Masculinity Hierarchical Regression Analysis of the Women’s Evaluative Ratings Based on General Evaluative Information.................................77

12. Plot of the Predicted Values from the Target x Job x Masculinity Hierarchical Regression Analysis of Women’s Evaluative Ratings Based on Negative Evaluative Information........................................78
CHAPTER 1
PERFORMANCE EVALUATION AND BIAS

Introduction

In a classic study, Goldberg (1968) found that participants evaluated an article more favorably when the author was presented as male (e.g., John T. McKay) rather than female (e.g., Joan T. McKay). However, many studies have failed to replicate this finding, and the conclusions of related research on prejudice against women have generally been mixed (Swim, Borgida, Maruyama, & Myers, 1989). Additionally, there have been significant changes in the social climate over the latter half of the 20th Century, and evidence suggests that egalitarian social beliefs have increasingly become the norm in recent years (Swim, Aiken, Hall, & Hunter, 1995; Judd, Park, Ryan, Brauer, & Kraus, 1995). “Data from national opinion polls suggest that fewer people endorse old-fashioned prejudicial beliefs such as unequal treatment of African-Americans as compared with European-Americans (McConahay, 1986) and suggest that fewer people disapprove of nontraditional roles for women (Myers, 1993)” (Swim et al., 1995). Further, contrary to the almost universal assumption that women are evaluated less favorably than men, recent evidence suggests that women are actually evaluated more favorably than men (Eagly, Mladinic, & Otto, 1991).

However, the suggestion that U.S. society is presently not only egalitarian but also biased favorably toward women should be viewed cautiously. A number of researchers have questioned the optimistic conclusion of researchers who suggest that prejudice and discrimination have been steadily diminishing over time. If prejudice and discrimination are diminishing, why do women and minorities continue to occupy a disadvantaged
position in American society, at least in regard to wages and promotions (Eagly et al., 1991)? Women and minorities have gained access to a number of occupations that were once closed to them, but ample evidence remains suggesting that inequities persist in the contemporary workplace (Maume, 1999; Lyness & Judiesch, 1999; Stuhlmacher & Walters, 1999; Kay & Hagan, 1995; Morrison & von Glinow, 1990). Thus, the mixed results of previous research may be due, at least partially, to differences in the expression of bias and prejudice in recent years compared to the past. Contemporary manifestations of bias may be more subtle than in previous years, possibly due to the presence of strong normative pressures discouraging the endorsement of blatantly prejudicial remarks (McConahay, 1986; Swim et al., 1995).

Relatedly, some studies may have failed to document bias because, concomitant with the changes in popular attitudes during the past 50 years, research participants have become increasingly savvy about the underlying motives of researchers, and the average person is now more reticent to express blatantly prejudicial attitudes than was previously the case. Although the days of routinely hearing blatantly prejudicial remarks have fortunately passed, given the enduring inequities in social status and wealth between European-American men and other groups in the United States, it seems premature indeed to conclude that prejudice and discrimination are problems of the past. It is more likely that prejudice and discrimination continue to persist, albeit in more covert forms. Consequently, researchers of prejudice have gradually refined their research designs from the relatively simple to the increasingly complex.
Revisions to Traditional Approaches of Studying Bias and Prejudice

Until recently, Allport’s (1954) definition of prejudice as “an antipathy based upon a faulty and inflexible generalization” (p. 9) was the prevailing framework for understanding prejudice and bias. However, the utility of using Allport’s model to understand contemporary expressions of bias has been called into question by researchers who note the differences between traditional and contemporary expressions of bias. Consequently, a number of measures designed to be sensitive to the nuances of contemporary prejudice have emerged over the past 15 years (e.g., subtle racism, Pettigrew & Meertens, 1995; symbolic racism, Sears, 1988; modern sexism, Swim et al., 1995; modern racism, McConahay, 1986; ambivalent sexism, Glick & Fiske, 1996; racial ambivalence theories, Katz & Hass, 1988; Katz, Wackenhut, & Hass, 1986). Most of these measures of contemporary prejudice, however, were developed for the study of contemporary racism (specifically the racism of European-Americans toward African-Americans) rather than sexism, although several of these models have recently been extended to examine sexism (see Swim et al., 1995; Harvie, Marshall-McCaskey, & Johnson, 1998; Gilbert, 1997; Swim & Cohen, 1997; Glick & Fiske, 1996). The central feature of these measures is a focus on the denial of continuing discrimination (Eckes & Six-Materna, 1998). Specifically, it is proposed that the denial of discrimination against minority groups and women masks underlying hostility (Glick & Fiske, 1996; Swim et al., 1995; McConahay, 1986).

The Advent of Social Cognition

Glick and Fiske (1996) have argued that “sexism is indeed a prejudice, but a special case of prejudice marked by a deep ambivalence, rather than a uniform antipathy,
toward women” (p. 491). They further suggest that the typical conceptualization of sexism as a reflection of hostility toward women ignores the “benevolent” aspect of sexism (i.e., the subjectively positive feelings toward women that idealize women in traditional female roles) that men often hold as a result of their intimate connections with women (e.g., as wives, mothers, and romantic partners; Glick, Diebold, Bailey-Werner, & Zhu, 1997). These subjectively positive feelings toward women are conceptualized as sexist because, in addition to hostile attitudes toward women, they also justify traditional gender roles and power relations which in turn serve to reinforce and legitimize men’s social control (Glick & Fiske, 1996).

Many of the contemporary models for understanding sexism were developed from racism models due to the clear parallels between sexism and racism, but it is also important to delineate the differences between the two “-isms.” As Glick and Fiske (1996) have noted, “the biology of sex creates a situation that is uniquely different from other in-group—out-group distinctions” because “sexual reproduction lends women ‘dyadic power’ (power that stems from dependencies in 2-person relationships) in that it compels men to rely on women as bearers of children and, generally, for the satisfaction of sexual needs” (p. 492; see also Guttentag & Secord, 1983). Thus, in addition to hostile attitudes, sexist men may have genuinely positive feelings toward women (Glick & Fiske, 1996). This conception of ambivalence is markedly different from other ambivalence theories in that hostile and benevolent sexism, the two components of ambivalent sexism, tend to be positively correlated because they both justify traditional gender roles and power relations (Glick et al., 1997). In contrast, “other ambivalence theorists have assumed (and have found) that beliefs associated with ambivalence are typically conflicting
(and therefore negatively correlated) or, at best, are unrelated” (Glick & Fiske, 1996, p. 494; Cacioppo & Bernston, 1994; Thompson, Zanna, & Griffin, 1995). Further, the “positive” feelings associated with racist ambivalence are thought to stem from guilt over past discrimination or “sympathy for the underdog,” whereas the “positive” side of sexist ambivalence results from feelings of love and reverence for women. Hence, “‘decent Jews’ or ‘good Blacks’ may be liked, but they are not put on a pedestal” (Glick et al., 1997, p. 1333). Additionally, Gaertner and Dovidio (1986) have argued that “sympathy for the underdog” is not a pro-Black attitude. Indeed, Glick and Fiske (1996) note that the conception of racial ambivalence proposed by Katz et al. (1986; 1988) has an element of paternalism analogous to the protective paternalism in Glick and Fiske’s benevolent sexism construct.

Another difference between current expressions of sexism and racism concerns the desire to protect an egalitarian image. Fiske and Stevens (1993) have argued that people are less worried about appearing sexist than they are about appearing racist. Because sexist men have strongly favorable feelings toward some women, they can easily rationalize their behavior and feelings by dividing women into “good” and “bad” subtypes, thereby avoiding a sense of conflict or dissonance with the rationalization that they don’t dislike women in general, only those who “deserve it” (Glick et al. 1997; Glick & Fiske, 1996). In contrast, there is more of a desire to protect an egalitarian image in manifestations of contemporary racism. For example, the theory of aversive racism suggests that people experience a conflict between their feelings and beliefs associated with a sincerely egalitarian belief system and unacknowledged negative feelings toward African-Americans (Gaertner & Dovidio, 1986). Thus, if aversive racists are unable to de-
emphasize racial factors, this conflict will create a sense of discomfort that “has been viewed as a feature of racism that can aid attempts to reduce prejudice” (Glick et al., 1997, p. 1333; see also Devine, Monteith, Zuwerink, & Elliot, 1991). Sexist ambivalence, however, results in no subjective discomfort (Glick & Fiske, 1996). Thus, “ambivalent sexists may experience little of the compunction that ambivalent racists often feel when engaging in hostile behavior (see Fiske & Stevens, 1993), all the while remaining convinced that they are not prejudiced against women” (Glick et al., 1997, p. 1333).

Conceptualizing sexism in this way led Glick et al. (1997) to wonder whether women would be evaluated differently as a function of how traditional they are perceived to be. They defined “traditional” and “nontraditional” women as “homemakers” and “career women,” respectively, hypothesizing that traditional and nontraditional women would be differentially evaluated as a function of the evaluator’s attitudes and beliefs. In support of their hypotheses, Glick et al. (1997) found that women in a nontraditional role were evaluated less favorably by men who held “hostile” attitudes toward women than by men who did not hold such attitudes. Conversely, women in a traditional role were evaluated more favorably by men espousing “benevolently sexist” attitudes than by men who rejected such attitudes (see Glick & Fiske, 1996). Adding these findings to the already impressive body of research produced by Glick and Fiske (1996) provides convincing evidence of the necessity for conceptualizing sexism as “ambivalent” rather than as the uniform antipathy originally proposed by Allport (1954).

Rationale for the Present Study

The present study was designed to address several limitations of previous research on evaluative bias. It extends the work of Glick and Fiske (1996) by assessing not only
participants’ degree of sexism, but also their attitudes about appropriate male roles and the degree to which participants possess instrumental (i.e., desirable “masculine” characteristics) and expressive (i.e., desirable “feminine” characteristics) traits. Male and female participants were asked to evaluate a male or female target in either a stereotypical masculine or feminine occupation. The methodology of the current study differs from previous studies of bias in several ways. First, the current study includes three distinct evaluative standards, each of which may be emphasized under different circumstances, whereas previous methods of measuring evaluative bias have typically relied on only normative evaluative standards. Secondly, the present study measures both positive and negative bias, rather than negative bias alone, as there may be certain instances in which targets are positively evaluated relative to other groups rather than negatively evaluated. Additionally, the job performance of targets in the present study varied from good to poor to neutral, providing a benchmark against which to interpret evaluations, because bias may be more apparent when targets’ job performance is unequivocally good or bad, rather than unknown by participants. Lastly, because men and women are not homogenous groups and there is actually more within-group than between-group variation among them, the present study includes female participants. Studies of gender bias that include only male participants likely result from the idea that sexism is a “male” problem requiring a change in men’s attitudes (see also Judd et al., 1995). However, it is certainly possible that women may exhibit bias toward both men and other women. Gender bias is a complex process, and if we are ever going to understand it, we need to know something about the perspectives of both groups.
Hypotheses

I expect to find systematic differences in the patterns of evaluations such that some men and/or women may exhibit positive or negative bias toward some target groups but not others. For example, the in-group—out-group conceptualization of prejudice suggests that men may evaluate male targets more favorably than they evaluate female targets, while women may be more favorable toward female targets than male targets. However, given that the similarities between men and women are greater than the differences, I do not expect to find evidence of in-group—out-group bias. Rather, I expect that evaluations of male and female targets in traditionally masculine or feminine occupations will be moderated by the attitude and personality variables included in the study, providing insight into the complicated nature of evaluative bias. Additionally, I expect that more traditional participants (as indicated both by the degree to which their attitudes and their “masculine” or “feminine” personality traits are stereotypically consistent with their gender) will evaluate targets in traditional gender-typed jobs more favorably than they will evaluate targets in non-traditional gender-typed jobs. Lastly, depending on the target group being evaluated and the participants’ individual characteristics, I expect to find evidence for systematic tendencies of participants to emphasize some evaluative standards over others, resulting in biased evaluations.
Overview of the Experimental Design

This study represents what West, Aiken, and Krull (1996) referred to as an “experimental personality design.” The design combines experimentally manipulated variables with naturalistic variables to assess the direct and interactive influence of both types of variables on a dependent measure. The experimental portion of the design is a $2 \times 2 \times (3 \times 3 \times 3)$ mixed model design with repeated measures on the last three factors. The two between-subjects factors are the target’s gender and the job’s gender stereotype (heavy equipment operator versus receptionist in a law firm). Previous research indicates that these two jobs are at the extreme ends of a masculine—feminine continuum, respectively, and are approximately equivalent in level of prestige (Glick, Wilk, & Perreault, 1995). The three repeated measures factors are the positive, negative, and neutral (absent) versions of the normative, ipsative, and expectation evaluative standards (e.g., Goolsby & Chaplin, 1988; Wilson, Chaplin, & Thorn, 1995). Thus, participants were asked to evaluate either a male or female target performing either a stereotypically masculine or stereotypically feminine job using all 27 possible combinations of positive, negative, or neutral (absent) normative, ipsative, and expectation standards. The naturalistic variables in the study are the gender of the participant, which is a categorical variable, and a set of continuous variables assessing sexist attitudes toward women, attitudes about appropriate male roles, and the personality traits of masculinity and femininity.
Participants

In exchange for course credit 261 undergraduate psychology students (132 women and 129 men) from the University of Massachusetts Amherst participated in the experiment. With the exception of participants’ gender, no demographic characteristics were collected; however, given the demographics of the undergraduate population at the University of Massachusetts Amherst, the vast majority of participants were European-American and between 18-21 years of age. Participants were randomly assigned to one of the four conditions: female target in a female occupation, female target in a male occupation, male target in a female occupation, or male target in a male occupation.

Measures

Participants completed five questionnaires: the Experimental Evaluational Styles Questionnaire (EESQ), a manipulation check questionnaire, the Ambivalent Sexism Inventory (ASI), the Masculine Role Norms Scale (MRNS), and the Personal Attributes Questionnaire (PAQ). A copy of each measure is provided in Appendix A.

Experimental Evaluational Styles Questionnaire (EESQ). Each participant completed one of four versions of the EESQ, which represented all possible combinations of target gender and gender stereotype of job (e.g., female target in a stereotypically feminine occupation). Participants read a brief description of the hypothetical target they were asked to evaluate in which the target was described in general terms (see Appendix A). All descriptive information remained constant across the four versions except for the target’s gender and job. Participants were then told that the target was recently evaluated at work and received a job performance rating of 22. This number was purposefully ambiguous so that participants would have to rely on the information provided in each
item to interpret this score. The EESQ is comprised of 27 items, each representing one combination of three different types of standards, each of which was valenced in one of three directions. Specifically, items included differently valenced (positive, neutral, or negative) normative, ipsative, and/or expectancy information. Participants were asked to consider the information given in each item (independent of the other items) and evaluate the target’s performance. Thus, each participant evaluated one of four possible targets 27 different times, basing his or her evaluations on different information regarding the target’s job performance for each rating.

**Manipulation Check.** It was crucial that participants be cognizant of their respective target’s gender and job when making their evaluations. To verify their awareness of the target they were evaluating without making the purpose of the study obvious, this six-item multiple choice measure was developed which asked participants to recall, 1) the target’s age, 2) the target’s gender, 3) the target’s level of education, 4) the target’s job, 5) the status (living or deceased) of the target’s parents, and 6) the target’s socioeconomic status. The manipulation check is provided in Appendix A. The questions were asked in multiple-choice format, and the correct answers to all of the questions had been provided in the original description of the target. Of the six questions, only questions two and four were critical to the study, each addressing the target’s gender and job, respectively. Participants’ data were included in the analyses as long as they correctly identified both the gender and job of their respective targets, regardless of their responses to the irrelevant items. Any participant who incorrectly identified the target’s gender, job, or both was dropped from the study and replaced by another participant.
Ambivalent Sexism Inventory (ASI). The ASI is a 22-item Likert-type scale developed by Glick and Fiske (1996) to assess two correlated dimensions of sexist attitudes toward women. The Hostile Sexism subscale concerns openly negative attitudes toward women and includes items such as “When women lose fairly they claim discrimination,” and, “Women exaggerate problems at work.” The Benevolent Sexism subscale represents the subjectively positive attitudes toward women of sexist men. It includes items that concern “protective paternalism” (e.g., “Men should sacrifice to provide for women), “complementary gender differentiation” (e.g., “Women have a quality of purity that few men possess”), and “heterosexual intimacy” (e.g., “Every man ought to have a woman he adores”). The theory on which the ASI is based and an extensive evaluation of its psychometric virtues are provided in Glick and Fiske (1996).

Personal Attributes Questionnaire (PAQ). The PAQ consists of a set of 24 bipolar adjective scales. It was developed by Spence and Helmreich (1978) to assess instrumental and expressive personality characteristics. The PAQ is one of the most widely used measures of the degree to which individuals possess “masculine” (instrumental) and/or “feminine” (expressive) traits. The PAQ yields three different scores: Masculinity (endorsement of desirable instrumental characteristics), Femininity (endorsement of desirable expressive characteristics), and Masculinity-Femininity (endorsement of traditional masculine characteristics over traditional femininity characteristics).

Masculine Role Norms Scale (MRNS). The MRNS is a 26-item Likert-type scale developed by Thompson and Pleck (1986) to assess a person’s endorsement of traditional masculine ideology. It contains three subscales: “Anti-femininity,” emphasizing that men should avoid anything feminine (e.g., “It is a bit embarrassing for a man to have a job that
is usually filled by a woman”); “Status,” emphasizing the importance of achievement for a man’s status (e.g., “Success in his work has to be a man’s central goal in this life”); and “Toughness,” emphasizing that a man should be rugged and strong (e.g., “When a man is feeling a little pain he should try not to let it show very much”). The psychometric virtues of the MRNS are summarized in Thompson, Pleck, and Ferrera (1993) and Sinn (1997).

Procedure

The data were collected in group testing sessions during the Spring and Fall semesters of 1999. To ensure standardized procedures, participants were given few verbal instructions. The experimenter informed participants that the study concerned how people evaluate other people’s job performance and informed them that there would be a total of five questionnaires. Participants were given an informed consent form, and after providing their informed consent (no participant refused), the experimenter randomly distributed one of the four versions of the EESQ to each participant. Participants were instructed to complete the EESQ and then give it to the experimenter, at which time they would receive a packet of four other questionnaires. The first questionnaire in this packet was the manipulation check, followed by (in random order) the ASI, PAQ, and MRNS. After completing the packet of questionnaires, participants were thanked for their participation and given a debriefing form that explained the actual purpose of the experiment. Participants were invited to contact the experimenter if they had further questions or concerns regarding the study.
CHAPTER 3

RESULTS

Manipulation Check

In total, data were collected from 261 participants, 129 men and 132 women. However, 10 of the 261 participants (six men and four women) were dropped from the study because they failed to correctly identify the target’s gender, job, or both. Thus, after excluding these 10 participants, the data set included 251 participants, 123 men and 128 women.

Rating Agreement

The 251 participants were very consistent in their ratings across the 27 items of the EESQ (coefficient alpha = .998). However, four of the participants exhibited inconsistent ratings relative to the other participants. Three of these participants’ responses contained no variance and yielded a zero correlation with the other participants. Specifically, regardless of the evaluative information given, two of these three participants (both women evaluating a male target in a masculine job) evaluated their respective target’s performance as “neutral,” while the third participant (a man evaluating a female target in a feminine job) evaluated his target’s performance as “slightly poor” on each of the 27 items on the EESQ. The fourth participant (a man evaluating a male target in a masculine job) rated his target so unpredictably that his responses were negatively correlated (-0.7) with the other participants’ responses. These four participants were dropped from the study, so the final sample used for the analysis of data included 247 participants, 121 men and 126 women. See Table 1 for the number of men and women in each condition of the final sample. The coefficient alpha for the final sample of 247 participants was .998.
Derivation and Psychometric Evaluation of the Dependent Variables

The effect of the standards on evaluations. Table 2 shows the mean evaluation rating for each of the 27 items. Consistent with previous research (e.g., Goolsby & Chaplin, 1988), all of the evaluative standards had an effect on the evaluative ratings such that positive versions of the standards resulted in significantly higher evaluations than negative versions of the standards. Also as expected (Goolsby & Chaplin, 1988), the Normative standard had the largest effect on the evaluative ratings, followed by the Ipsative and Expectation standards.

The structure of the EESQ. The foregoing results indicate that our manipulation of the evaluative information had its expected effect on the participants' ratings. However, it is necessary to consider the structure of the EESQ in order to make sense of the participants' general evaluative tendencies. Participants' responses to each item on the EESQ were influenced by various combinations of two factors, style and valence, each with three subtypes. The style factors include the normative, ipsative, and expectation standards manipulated in this study; they are placed under the rubric “style factors” because each offers a distinct way of viewing and evaluating a situation, and theoretically, individuals may develop preferences for one standard over another. The valence factors refer to the positive, negative, or neutral value attached to each standard. For example, referring back to Table 2, the EESQ item “PNO” contains positive normative, negative ipsative, and neutral (or absent) expectation information. It makes intuitive sense that items containing evaluative standards of the same valence will be rated in a consistent manner, and indeed, this intuition was empirically supported. As Table 2 illustrates, items containing only positively valenced standards were evaluated most favorably while items
containing only negatively valenced standards were evaluated least favorably. Table 2 also illustrates that the mean ratings of participants on the 12 “mixed valence” items containing standards of both positive and negative valence tended to cluster toward the middle relative to the other items.

The purpose of this study was to examine the impact of differently valenced (positive, negative, or mixed) evaluative information on ratings of job performance as a function of gender of target, type of job, and various characteristics of the raters (e.g., gender, level of sexism, beliefs about masculine roles, and adherence to traditional gender roles). Such an investigation, however, could result in various patterns of 5-, 6-, 7- (or even higher) way interactions in the $2 \times 2 \times (3 \times 3 \times 3)$ experimental design, in addition to a number of moderator variables. The difficulty of statistically detecting such higher order interactions is well documented (e.g., Cohen & Cohen, 1983). Moreover, such higher order interactions are nearly impossible to interpret and communicate. Thus, we sought to simplify the data generated by our design by aggregating the 27 $(3 \times 3 \times 3)$ items of the EESQ into multi-item scales.

**Derivation of the EESQ scales.** We sought to simplify the data generated by our design by aggregating the 27 $(3 \times 3 \times 3)$ items of the EESQ into multi-item scales. The derivation of these scales not only allows for improved interpretation and communication, but also increases the statistical power and precision of the study by assessing bias more reliably than would be possible using single item measures which, by definition, tend to be unreliable. This process of aggregating the 27 items of the EESQ into independent scales resulted in five dependent variables: the General, Positive, Negative, Mixed, and Neutral
scales. The first four scales are comprised of multiple items, while the last scale is based on the single neutral item of the EESQ.

The General scale is comprised of the entire 27 items on the EESQ, while the other four “valence” scales include the items consistent with the valence of the respective scale’s name. Scores on the General scale indicate that participants differ systematically in their overall evaluative tendencies, regardless of the specific evaluative information contained in each item. Specifically, regardless of the valence(s) of the evaluative information in each item, participants displayed an overall tendency to evaluate people either more positively (used the higher end of the rating scale) or more negatively (used the lower end of the scale).

The four valence scales were derived to provide a measure of how participants interpreted and evaluated their respective targets as a function of the valence of the specific evaluative information contained in the items. Specifically, the Positive scale was derived by taking the average ratings of the seven items on the EESQ containing only positively valenced standards; the Negative scale includes the seven items that are solely negative in valence; and the Mixed scale is comprised of the 12 items on the EESQ that include both positively and negatively valenced evaluative information. Additionally, we used the neutral item on the EESQ to create a single-item valence scale. Although this single-item scale is, by nature, less reliable than the multi-item scales, it can offer valuable information about how people evaluate others in the absence of evaluative information.

Table 3 presents the correlations among the five EESQ scales. Also shown in Table 3 are the means, standard deviations, and coefficient alphas for each of these scales.
Psychometric Evaluation of the Individual Difference Variables

Although the ASI, MRNS, and PAQ are reported to have acceptable internal consistency, we assessed the internal consistency of these measures in our sample. Table 4 presents the means and standard deviations, separately for men and women, for the individual difference scales in our sample. Also shown is the coefficient alpha based on the entire sample.

The Influence of Either the Situational or Individual Difference Variables on Evaluations

In an effort to improve the comprehensibility of our results, we centered each of the EESQ scales around the mean of the total sample. These centered variables are expressed as deviation scores; a positive score on any of these measures indicates a higher rating than the average of the entire sample, and a negative score indicates a lower rating than the overall average.

The effect of the situational variables on evaluations. In the experimental tradition, the first set of analyses were 2 x 2 (Target x Job) between-subjects analyses of variance for each of the five dependent variables. A main effect for Target gender was found for both the General and Negative Scales of the EESQ. Specifically, when participants’ ratings were averaged across all possible combinations of the evaluative information, female targets were evaluated more positively than male targets ($F(1, 243) = 5.60, p = .019$). Thus, female targets were generally evaluated more positively than male targets, but this phenomenon was particularly evident when participants based their evaluations on negative information indicating that their respective target’s job performance was particularly poor ($F(1, 243) = 11.82, p = .001$). Table 5 presents the means and standard deviations for the each condition of the $2 \times 2$ (Target x Job) ANOVA for the General
Scale and Negative Scale separately. These analyses provided no evidence of any other main effects or interaction effects.

The relation between the individual difference measures and evaluations. In the classic individual differences tradition, the next set of analyses were correlations between each of the individual difference variables and the five evaluation measures. The correlation coefficients for these analyses are presented in Table 6. With two exceptions, neither of the attitude scales nor gender were related to participants’ evaluations. However, in the context of negative evaluative information, participants’ ratings were modestly correlated with both their gender and their attitudes regarding masculine-role norms, specifically the need for men to be emotionally and physically “tough.” With regard to gender, women (M = .11, SD = .79) tended to evaluate targets performing poorly on the job more positively than men (M = -.12, SD = .66) evaluated the same targets. The second exception was related to participants’ attitudes. There was an inverse relationship between participants’ evaluations of targets performing poorly on the job and their attitudes concerning the need for men to be emotionally and physically tough. Specifically, participants who endorsed statements indicating that it is important for men to display “toughness” were more likely to negatively evaluate targets performing poorly on the job than were participants who held less traditional gender-role beliefs. These two exceptions notwithstanding, the attitude scales and participant gender were generally not related to participants’ evaluations. It was only when participants were evaluating targets whose job performance was consistently poor that any relationship was present.

Conversely (and also indicated in Table 6), more consistent, albeit modest, relationships were suggested between the personality variables and participants’
evaluations. Generally, higher scores on Femininity (i.e., endorsement of desirable expressive or "feminine" characteristics) were positively correlated with evaluations while higher scores on both Masculinity (i.e., endorsement of desirable instrumental or "masculine" characteristics) and Masculinity-Femininity (i.e., endorsement of traditional masculine characteristics over traditional feminine characteristics) tended to be negatively correlated with evaluations. Specifically, participants scoring higher on Femininity evaluated targets more favorably than participants scoring lower on Femininity both generally and when the evaluative information was mixed or negative. In contrast, participants scoring higher on both Masculinity and Masculinity-Femininity tended to evaluate targets more negatively than participants scoring lower on these two scales. These negative correlations were present both generally and when the evaluative information was both mixed and consistently negative for the Masculinity scale. For the Masculinity-Femininity scale, the negative correlations were evident generally and when evaluative information was consistently negative, but not when the evaluative information was mixed.

Combining Individual Difference and Situational Variables to Predict Evaluations

One of the factors motivating the complex design of this study was the belief that predicting and understanding evaluative bias requires combining both situational and individual difference variables into the same study. Thus, we began combining these variables in a series of ANOVAS by examining the moderating effect of participant gender, a categorical individual difference variable, on the two situational variables of target gender and gender stereotype of job. The subsequent and more complicated analyses involving the higher-order moderating effects of sexism, masculine role norms,
and the personality traits of masculinity and femininity on how men and women evaluate male and female targets in traditionally male or female jobs were examined using hierarchical regression analyses.

Do men and women generally evaluate the performance of men and women in traditional and nontraditional jobs differently? A $2 \times 2 \times 2$ analysis of variance was used to assess the moderating effects of participant gender on the effect of target and job for each of the five evaluation measures. The three-way interaction (Gender x Target x Job) was significant only for the Negative scale of the EESQ ($F(1, 239) = 4.90, p = .028$); see Figure 1. Targets who were clearly performing poorly on the job were evaluated differently by men and women as a function of the respective target’s gender and the type of job. Generally, women evaluated targets whose job performance was consistently poor more favorably than men evaluated those same targets. Additionally, women evaluated targets performing poorly in nontraditional jobs more favorably than men evaluated them. Specifically, women rated female targets in stereotypically masculine jobs and male targets in feminine jobs more favorably than men did when the targets’ job performance was clearly not as good as the performance of other employees, not as good as the target’s previous performance, and not as good as participants had expected the targets to perform. Moreover, men evaluated poorly performing male targets in feminine occupations more negatively than they evaluated poorly performing targets in other conditions. Women, on the other hand, gave men in stereotypically feminine occupations more positive evaluations. The group receiving the most positive evaluations from both women and men were the female targets performing stereotypically masculine jobs badly. Although women evaluated these targets more favorably than men did, men also evaluated
female targets performing poorly in a masculine job more positively than they evaluated other poorly performing targets. Conversely, male targets with poor job performance in traditionally masculine occupations were evaluated negatively by both men and women, whereas female targets whose performance in traditionally feminine occupations was poor were evaluated slightly positively by both men and women.

Additionally, there was some suggestion of a main effect for gender when the dependent variable was the single item rating based on no evaluative information ($F(1, 239) = 3.199, p = .075$). Although not significant by conventional standards, there was some indication that women ($M = .10, SD = .88$) provided slightly more positive ratings of job performance than men ($M = -.10, SD = .83$) when they had no information on which to base their ratings.

Do men and women who differ on Sexism, Masculine Role Norms, or Masculinity-Femininity show biased evaluations toward male and/or female targets in masculine or feminine jobs? Although the moderating effect of gender on evaluations based solely on negative evaluative information is intriguing, the foregoing results do not suggest that men and women generally evaluate the targets in the different conditions differently. This conclusion is not surprising since we have argued that treating men and women as homogenous groups has been one of the limitations of previous research. The rationale for considering individual differences in personality and attitudes in this study was based on the hypothesis that these variables provide critical information for understanding gender bias in performance evaluations. To evaluate this hypothesis, a series of hierarchical regression analyses were conducted, using each of the five evaluation measures separately as the dependent variables. We began by considering the effects-
coded target, job, and participant gender categorical variables with one of the continuous gender role or attitude variables. Next we added the six cross-products ("two-way interactions") involving all possible pairs of each of these variables. Then we added the four triple cross-products (e.g., Target x Job x "Attitude") involving all 3-way combinations of the four variables. Finally we added the four-way cross-product (e.g., Gender x Target x Job x "Attitude"). The semi-partial correlation between the four-way cross-product and the evaluation measure indicates the strength (proportion of variance) of the moderating effect of the individual difference variable on the male and female participants' evaluations of the different targets in the different jobs. This semi-partial can also be tested for significance, and because of the generally low statistical power of tests of higher order interactions (see Aiken & West, 1991), we elected to further explore any four-way interactions with associated probability levels of .10 or lower (two-tailed test), with the further condition that the four-way interaction account for a minimum of 1% of the variance in the participants' evaluations (Chaplin, 1991).

**Graphing interactions of continuous x categorical variables.** The effects reported in the remainder of the results section involve both continuous and categorical variables. To interpret these interactions we used the procedures recommended by Aiken and West (1991) and West, Aiken, and Krull (1996) for graphing continuous x categorical variable interactions. Specifically, we obtained the partial regression coefficients for all the terms and their cross-products in the equation used to test the interactions. We then used these coefficients to predict values for the evaluational measure for all high and low combinations of the variables that made up the interaction by weighting high and low scores (1 and -1 for the effects coded categorical variables, and one standard deviation
above and below zero for the centered continuous variables) by their appropriate partial regression coefficients and summing across the variables. Thus, the figures presented in the remainder of the results section are based on predicted (rather than actual) values.

**Benevolent Sexism as a moderator.** The four-way interaction for Gender x Target x Job x Benevolent Sexism for predicting evaluations based solely on positive evaluative information (i.e., the Positive scale of the EESQ) had a squared semi-partial correlation of .02 ($t(231) = -2.44$, $p = .015$). As can be seen in Figure 2, women who endorsed “benevolently sexist” statements (e.g., “women should be cherished and protected by men” and “many women have a quality of purity that few men possess”) evaluated female equipment operators who performed well on the job much more positively than they evaluated other groups of people. Figure 2 represents several other interactions as well. Generally, male and female participants who scored low on the Benevolent Sexism scale (i.e., did not endorse “benevolently sexist” statements) differed in their evaluational patterns. For example, men scoring low on Benevolent Sexism evaluated male receptionists performing well on the job much more positively than did women scoring low on Benevolent Sexism. A similar pattern was found for the evaluations of female equipment operators, with men low on benevolent sexism evaluating female equipment operators exhibiting good job performance more positively than women scoring low on benevolent sexism evaluated them. However, the pattern was reversed when these participants evaluated both female receptionists and male equipment operators. Specifically, women who scored low on benevolent sexism evaluated both female receptionists and male equipment operators more positively than men who scored low on benevolent sexism did.
As can be seen in Figure 2, there are clearly other, albeit less pronounced, effects apparent in this 4-way interaction. However, the reader can now fully appreciate the difficulty of interpreting such higher order interactions. Given the complexity of Figure 2, we sought to make the effects more interpretable by exploring this effect for men and women separately. When trying to make sense of this interaction, it is important to remember that these evaluations were based solely on positive evaluative information (i.e., it was clear to participants that these targets were all performing their jobs very well). The squared semi-partial correlation for the Benevolent Sexism x Target x Job interaction for men was .11 (t (113) = 3.77, p < .001). However, no effect was found for women, as the squared semi-partial for this 3-way interaction considered for women was minuscule (t = .22, p = .82). As shown in Figure 3, the graph of these predicted values for only the male participants indicates a clear interaction between the evaluations of men scoring high on benevolent sexism and men scoring low on benevolent sexism. Specifically, men who endorsed benevolently sexist statements evaluated male equipment operators more positively than they evaluated any other group, while men who did not endorse benevolently sexist statements evaluated this group (male equipment operators) more negatively than they evaluated other groups. Conversely, female receptionists were evaluated by the benevolently sexist men more negatively than these men evaluated other groups, while female receptionists were evaluated more positively by men who did not endorse benevolently sexist statements than these non-benevolently sexist men evaluated other groups.
Male Toughness as a moderator. Participants’ attitudes regarding the need for men to be physically tough were a robust predictor of evaluative bias. The four-way interaction for Gender x Target x Job x Toughness was statistically significant for three of the dependent variables (General scale, Negative scale, and Mixed scale), and the Neutral scale was marginally significant, warranting further exploration based on the aforementioned criteria. It was only when the evaluative information was solely positive that Toughness had no moderating effect. When evaluations were averaged across all possible combinations of evaluative information (i.e., when the dependent variable was the General scale), the four-way interaction involving Gender x Target x Job x Toughness had a squared semi-partial correlation of .02 ($t (231) = 2.18, p = .030$). When this effect was analyzed for male and female participants separately, however, there were no significant interactions for Target x Job x Toughness for either men ($t (113) = -1.65, p = .10$) or women ($t (118) = 1.41, p = .16$). Although the interpretation is challenging, there are several striking interaction effects in this four-way interaction (see Figure 4). First, male and female participants who endorsed statements regarding the necessity for a man to be tough (e.g., “When a man is feeling a little pain he should try not to let it show very much,” and “Nobody respects a man very much who frequently talks about his worries, fears, and problems”), differed in their evaluative patterns. For example, male participants who believed it is important for men to be tough generally evaluated female equipment operators positively, while women with the same attitudes generally evaluated female equipment operators negatively. In contrast, these male participants generally evaluated both female receptionists and male equipment operators negatively, while the female participants generally evaluated these two groups slightly positively. There was no
difference in the evaluations of male receptionists by men and women endorsing statements regarding the need for men to be tough. Conversely, the evaluative patterns were different for participants who did not hold beliefs regarding the necessity for men to be tough. Most strikingly, women who did not endorse male toughness generally evaluated female equipment operators much more positively than other groups were evaluated by both women and men, irrespective of their attitudes regarding male toughness. Male participants who rejected toughness attitudes were slightly positive in their evaluations of female equipment operators, while women who rejected male toughness evaluated female equipment operators quite positively. Men scoring low on toughness generally evaluated male receptionists quite negatively. Women who did not endorse male toughness, however, generally evaluated male receptionists neutrally. Lastly, male participants scoring low on the Toughness subscale generally evaluated both female receptionists and male equipment operators slightly positively, while women endorsing similar beliefs about male toughness gave these two groups slightly negative evaluations.

Consistent with the general ratings moderated by toughness that were just discussed, participants basing their ratings solely on negative performance information (i.e., the Negative scale) exhibited a similar pattern of evaluations. However, there was more variation in the magnitude of ratings based on negative performance information than there was for the general ratings. The four-way interaction based on negative performance information for the Gender x Target x Job x Toughness interaction had a squared semi-partial correlation of \(0.03 (t (231) = 2.62, p = .009)\) and is represented in Figure 5. As can be seen when comparing Figure 5 to Figure 4, the graphs for participants
who did not endorse statements indicating the need for men to be tough look remarkably
similar except for the difference in scale of the two figures. For example, women who
rejected the notion that men need to be tough evaluated female equipment operators
positively both generally and when their evaluations were based solely on negative
performance information; however, when the performance information was solely
negative, these women gave female equipment operators substantially higher ratings (.55)
than they did generally (.25). Similarly, men who did not endorse male toughness
statements evaluated male receptionists negatively both generally and when their ratings
were based solely on negative performance information, but their evaluations of male
receptionists were much more negative when based solely on negative performance
information (-.85) than on a wide range of positive and negative information (-.27).
Additionally, men scoring low on Toughness generally evaluated female receptionists
positively (.07), but when the evaluative information indicated poor job performance,
these men evaluated female receptionists more positively (.35). When the evaluations
were based on poor job performance, women who did not endorse male toughness
attitudes evaluated male receptionists positively (.20), whereas their evaluations were
neutral when based on general information (0). Participants who did endorse toughness
attitudes also exhibited similar evaluative patterns, whether their ratings were based on
general (Figure 4) or negative (Figure 5) performance information. The main difference
was again an issue of magnitude. For example, when job performance was clearly poor,
men and women evaluated male receptionists most negatively (-.32), followed by male
equipment operators (-.22). However, when job performance was averaged over a
number of good and bad situations, these “high toughness” participants evaluated male receptionists only slightly negatively (-.07).

When the 4-way interaction presented in Figure 5 was considered for men and women separately in order to improve the communicability of the results, the squared semi-partial correlation for men was .04 (t (113) = -2.36, p = .020) and for women was .02 (t (118) = 1.45, p = .15). As can be seen in Figure 6, when evaluations were based solely on negative evaluative information, male participants scoring low on the Toughness scale evaluated male receptionists most negatively (-.81) and female receptionists most positively (.39). These male participants were neutral and slightly positive in their evaluations of male equipment operators and female equipment operators, respectively. There was less, although substantial, variation in the ratings of men who scored high on the Toughness scale. Similar to the men not endorsing toughness attitudes, the men who believed it is important for man to be tough evaluated male receptionists more negatively (-.32) than they evaluated other groups, although not nearly as negatively as the “low toughness” men did (-.81). These men who endorsed toughness attitudes also evaluated male equipment operators (-.29) and female receptionists (-.15) negatively when basing their ratings solely on poor job performance.

When ratings were based on mixed evaluative information (i.e., the Mixed scale comprised of the 12 items on the EESQ containing evaluative information that was both positive and negative), the four-way interaction involving Gender x Target x Job x Toughness (presented in Figure 7) had a squared semi-partial correlation of .02 (t (231) = 2.21, p = .028). Consistent with the other interactions involving Toughness, women who scored low on the Toughness scale evaluated female equipment operators more positively.
than participants evaluated any other group. Additionally, men scoring low on Toughness evaluated male receptionists negatively, as they had both generally and when evaluative information was based solely on poor job performance. The ratings of participants who endorsed male toughness attitudes were also similar when based on mixed job performance compared to the negative and general job performances already presented. Specifically, “high toughness” men evaluated female equipment operators with mixed job performance positively, while “high toughness” women evaluated the same group negatively. Additionally, these “high toughness” men evaluated female receptionists negatively, while the “high toughness” women evaluated them slightly positively.

When the effect presented in Figure 7 was analyzed for men and women separately, the squared semi-partial correlation for men was .03 ($t (113) = -1.81, p = .073$) and for women was .01 ($t (118) = 1.23, p = .222$). These interactions are not presented graphically since neither reached conventional levels of statistical significance.

The final four-way interaction involving Gender x Target x Job x Toughness was based on the single item including no evaluative information and had a squared semi-partial correlation of .01 ($t (231) = 1.63, p = .104$). Although not statistically significant by conventional standards, this effect warranted further exploration. When explored for men and women separately, the squared semi-partial correlation for women was .03 ($t (118) = 2.12, p = .036$) but for men was minuscule in size. Figure 8 presents the 3-way interaction of Target x Job x Toughness for only the female participants when evaluations were based on no evaluative information (i.e., the neutral item of the EESQ). Women who endorsed statements concerning the necessity for men to be tough evaluated male equipment operators quite positively (.57) when provided no information about the
target’s actual job performance, but women who did not endorse male toughness statements evaluated male equipment operators negatively (-.20). A similar pattern was apparent for the evaluations of female receptionists. The “high toughness” women evaluated female receptionists positively (.43), while the “low toughness” women evaluated female receptionists negatively (-.08) when provided no information about the target’s job performance. Likewise, although this pattern was less extreme, male receptionists were evaluated neutrally (.02) by women endorsing male toughness, while women who did not endorse male toughness statements evaluated male receptionists negatively (-.10). In contrast, women evaluated female equipment operators positively regardless of the women’s attitudes concerning male toughness; however, the women who did not endorse male toughness statements were more positive (.40) than the women who did endorse these statements (.20).

**Male Status as a moderator.** The Status subscale of the MRNS measures attitudes concerning the importance of status in men’s lives. When evaluations were based solely on positive evaluative information, the squared semi-partial correlation for the Gender x Target x Job x Status interaction was .02 (t (231) = -2.17, p = .031). As can be seen in Figure 9, women who endorsed statements concerning the importance of status in men’s lives (e.g., “Success in his work has to be a man’s central goal in this life,” and “A man should always try to project an air of confidence even if he really doesn’t feel confident inside.”) evaluated female equipment operators much more positively (.68) than any other group was evaluated when target job performance was clearly good. Conversely, men who endorsed the same statements evaluated female equipment operators negatively (-.12), even when it was clear that the performance of the female equipment operator was
excellent. There were also differences in how men and women with different attitudes regarding male status evaluated both male and female receptionists. Specifically, “high status” males were neutral (0) in their evaluations of female receptionists exhibiting a good record of job performance, but were slightly negative (-0.08) in their evaluations of male receptionists who performed their jobs well. In contrast, women who believed male status is important evaluated male receptionists neutrally (0) and female receptionists negatively (-0.12), even when it was clear that all targets had performed their jobs well. There were no significant differences in “high status” men’s and women’s evaluations of male equipment operators, both of which were slightly positive (0.12). For the “low status” participants, however, a different evaluative pattern emerged. For example, “low status” men positively evaluated both male receptionists and female equipment operators who were performing their jobs well (0.24 and 0.16, respectively) while “low status” women evaluated these groups negatively (-0.12 and -0.08, respectively). Conversely, “low status” men and women both evaluated male equipment operators negatively, although the women were slightly more positive (-0.04) in their evaluations than the men (-0.16). “Low status” men also evaluated female receptionists negatively (-0.20), while “low status” women evaluated female receptionists positively (0.12).

The follow-up analyses of the 4-way interaction presented in Figure 9 for men and women separately yielded a squared semi-partial correlation of 0.05 ($t(113) = 2.49, p = 0.014$) for men, but for women, the squared semi-partial correlation was near zero ($t(118) = -0.69, p = .49$). The graph of the predicted values of the men’s evaluations based solely on positive evaluative information is shown in Figure 10. “Low status” men evaluated female receptionists negatively (-0.20), but they evaluated other groups positively.
receptionists received the highest evaluations (.23), followed by female equipment operators (.17) and male equipment operators (.15). Conversely, "high status" men evaluated only male equipment operators positively (.11); their evaluation of female receptionists was neutral (-.01), while male receptionists (-.09) and female equipment operators (-.11) were evaluated negatively.

Masculinity as a moderator. None of the four-way interactions involving Masculinity as a moderator variable were statistically significant by conventional standards, but two warranted further exploration. The first was based on general evaluative information, that is, the ratings of participants averaged across all combinations of evaluative information. The second was based on negative evaluative information. The first interaction (Gender x Target x Job x Masculinity), based on general evaluative information, had a squared semi-partial correlation of .01 ($t(231) = 1.66, p = .099$).

When this effect was examined for men and women separately, the squared semi-partial correlation for women was .05 ($t(118) = 2.78, p = .006$) but for men was zero ($t(113) = .07, p = .947$). The graph of the predicted values of women's evaluations based on the average of all possible combinations of evaluative information is shown in Figure 11.

Women who were more traditional, in that they indicated they do not possess "masculine" (instrumental) characteristics, evaluated female equipment operators most positively (.25), followed by male receptionists (.15), but they evaluated female receptionists neutrally (.02) and male equipment operators slightly negatively (-.08). Women who were less traditional (i.e., indicated they possess "masculine," or instrumental, characteristics), however, were fairly neutral in their evaluations of female equipment operators (0), female...
receptionists (.02), and male equipment operators (-.02). In contrast, these less traditional (i.e., more masculine) women evaluated male receptionists negatively (-.25).

The second four-way interaction (Gender x Target x Job x Masculinity) that warranted further investigation had a squared semi-partial correlation of .01 ($t (231) = 1.96, p = .051$). This interaction was based solely on negative evaluative information, and when it was examined for men and women separately, the squared semi-partial correlation for women was .04 ($t (118) = 2.52, p = .013$), but the squared semi-partial correlation for men was, again, zero ($t (113) = -.254, p = .800$). The graph of the predicted values for women’s evaluations based solely on negative job performance information is shown in Figure 12. Most strikingly, women’s evaluations of female equipment operators were substantially different as a function of the degree of desirable “masculine” characteristics the female participants indicated that they possess. Women who possessed a number of desirable “masculine” traits evaluated female equipment operators with poor job performance negatively (-.37), but women possessing fewer desirable “masculine” traits evaluated poorly performing female equipment operators quite positively (.68). There was also a marked difference in women’s evaluations of male receptionists, depending on the level of masculinity that the respective female participant indicated she possessed. Women with more desirable “masculine” traits evaluated male receptionists with poor job performance slightly negatively (-.07), while women with fewer desirable “masculine” traits evaluated male receptionists positively (.26). Additionally, when the job performance of targets was poor, there were not significant differences in how women possessing differing degrees of masculine traits evaluated either female receptionists (both groups of women were slightly positive: .05) or male equipment operators (both groups of
women were slightly negative: -.15). However, there were within-group evaluative
differences when targets’ job performance was poor. Specifically, women possessing few
desirable “masculine” characteristics evaluated female equipment operators most
positively ( .68), followed by male receptionists (.26) and female receptionists (.05), with
only the male equipment operators receiving a negative evaluation ( -.15). Conversely,
women possessing a number of desirable “masculine” characteristics evaluated female
equipment operators most negatively (-.37), followed by male equipment operators (-.15)
and male receptionists (-.07), with only female receptionists receiving a positive evaluation
(.05).
CHAPTER 4
DISCUSSION

Abbreviated Summary of Results

As predicted, the results of this study suggest that men’s and women’s evaluations of male and female targets’ job performance in stereotypically gender-congruent and gender-incongruent occupations are influenced by a complex pattern of both individual difference and situational variables. Given the mixed results of previous research (see Pulakos, White, Oppler, & Borman, 1989) suggesting both that participant gender does not influence evaluations (e.g., Hall & Hall, 1976; Peters, O’Connor, Weekley, Pooyan, Frank, & Erenkrantz, 1984; Schmitt & Lappin, 1980) and that participant gender does affect evaluations such that women tend to evaluate targets more positively than men (e.g., Bartol & Butterfield, 1976; London & Poplawski, 1976), the current study included a number of possible moderator variables to further elucidate the complex processes of evaluative bias. Additionally, by manipulating the degree to which performance evaluations were based on good, poor, average, or ambiguous job performance, the results of the current study indicated that bias manifests differently as a function of the quality of employees’ job performance. Specifically, positive or negative bias was particularly evident when participants evaluated targets with good or poor job performance, respectively. The positive or negative direction of the bias was directly related to participants’ attitudes and personality traits.

Findings of Particular Interest

General Differences in Participants’ Evaluations of Male and Female Targets. As expected, there were generally not in-group—out-group differences in men’s and
women’s evaluations of male and female targets. However, there was one exception to this general lack of in-group—out-group differences and it occurred when participants were evaluating targets who were performing their jobs poorly. Specifically, targets who exhibited particularly poor job performance in stereotypically nontraditional occupations were evaluated more favorably by women than by men (see Figure 1). It was clear to participants that these targets’ job performance was worse than that of other employees, worse than it had been on other jobs the targets had held, and worse than the participants expected these targets to perform. Thus, one would expect both men and women to evaluate these targets negatively. Yet this was not the case, as women were actually quite favorable in their evaluations of female equipment operators who were clearly not doing their jobs well. Interestingly, men did not evaluate poorly performing female equipment operators negatively either; instead, they were neutral in their evaluations of the inept female equipment operators.

This exception to the general lack of in-group—out-group differences between men and women was also apparent in men’s and women’s respective evaluations of male receptionists who were clearly performing their jobs badly (see Figure 1). As with the female equipment operators, women were again more favorable than men in their evaluations. Women were slightly positive in their evaluations of male receptionists whose job performance was clearly poor. Men, however, assigned ratings that were quite negative to the inept male receptionists. Given the huge number of women who have entered the work force since the Feminist Movement of the 1960s and 1970s, it is possible that these differences in evaluative styles result from the collective experience of women entering a workforce dominated by men. Perhaps the women in this sample identified with
the male and female targets working in occupations traditionally viewed as unsuitable for them solely because of their gender. It is possible that women generally had more empathy for employees working in nontraditional occupations and were hesitant poorly evaluate these targets.

Positive and Negative Bias, Moderated by Male Toughness, Status, and "Masculinity," Toward Targets Working in Traditional and Nontraditional Occupations.

As expected, there was some tendency for more traditional participants to negatively bias their evaluations of targets working in nontraditional occupations. In contrast, some groups of less traditional participants were actually positively biased toward targets working in nontraditional occupations. Not expected, however, was the finding that in various circumstances, some less traditional participants displayed negative bias toward targets working in nontraditional occupations while other, more traditional participants were actually positively biased toward targets in nontraditional occupations.

Negative bias of traditional participants toward targets working in nontraditional occupations. There was some suggestion that participants who held more traditional beliefs were negatively biased, in certain circumstances, toward targets in nontraditional occupations. This tendency was particularly evident when there was no doubt to participants that targets were doing their jobs well. For example, men who were traditional in that they believed, relative to other men in the sample, that status is an important part of a man's life (i.e., that men need to strive for respect and admiration through achievement; Sinn, 1997), were actually negatively biased in their evaluations of both male receptionists and female equipment operators who displayed records of good job performance (see Figure 10). This tendency cannot be accounted for by factors other
than negative bias against targets in nontraditional occupations because these targets were clearly performing their jobs well. Their job performance was better than that of other employees, better than it had been in the past, and better than the participants expected it possibly could be. Yet, these more traditional men still evaluated them negatively. These men certainly would be hard-pressed to explain their negative evaluations in terms other than gender bias; it seems that they were not concerned about being perceived as sexist. Perhaps Fiske and Stevens (1993) were right when they proposed that there is much less normative pressure on people to deny they are sexist than to deny they are racist.

There was also evidence that more traditional women may exhibit negative bias toward targets in nontraditional occupations. For example, women who more strongly believed, relative to other women, that it is important for a man to be emotionally and physically tough, evaluated both female equipment operators and male receptionists negatively when they had no information on which to base their evaluations (see Figure 8). There was also some indication that these women emphasized the negative job performance information over the positive information when basing their evaluations on conflicting performance information (see Figure 7). In these instances of negative bias, both traditional men and women were biased against male and female targets working in nontraditional occupations. This certainly seems to support the idea that ideology is a better predictor of evaluative bias than group membership.

Positive bias of nontraditional women toward targets working in nontraditional occupations. The data indicate that participants' attitudes regarding the necessity for men to be physically and emotionally tough were a robust predictor of evaluations generally, and evaluations of targets with consistently poor job performance, mixed job performance,
and job performance that was unknown to the raters. Positively biased evaluations were particularly evident when participants evaluated targets who performed their jobs badly. For example, women who were less traditional, as indicated by their disagreement, relative to other women, that men need to be tough, generally evaluated female equipment operators positively, but this tendency was especially evident when the female equipment operators were clearly doing their jobs badly (see Figures 4, 5, & 7). Indeed, positive bias was clear in these less traditional women when they evaluated both female equipment operators and male receptionists whose performance was worse than that of other employees, worse than it had been in previous jobs, and worse than these women expected the targets’ performance could be (compare Figures 4 & 5). These women may have felt it was important to give targets in nontraditional jobs “a break.” Indeed, one of the women who participated in the study commented to the experimenter afterward that she evaluated her target (a female equipment operator) very positively because she felt her target “could use all the help she could get.” This participant stated that it must be difficult for a woman to work as an equipment operator, and she reasoned that her positive evaluation was needed to counter all of the negative evaluations she believed this particular target would receive.

Positive bias of traditional women toward targets working in nontraditional occupations. An unexpected finding was the positive bias evident in women who were traditional in a different way. Women who indicated that they possess few instrumental traits (i.e., desirable “masculine” characteristics), relative to other women, also positively evaluated female equipment operators who were clearly not good at their jobs (see Figure 12). It is possible that these less “masculine,” hence, more traditional, women were
imagining themselves working as female equipment operators and could see themselves struggling to do the job well. This identification may have led them to positively evaluate the poorly performing female equipment operators.

**Negative bias of nontraditional women toward female targets in nontraditional occupations.** Interestingly, women who were less traditional in that they indicated they possessed more instrumental traits (i.e., desirable “masculine” characteristics), relative to other women, evaluated incompetent female equipment operators quite negatively. Why might this be the case? A similar explanation to that of the more traditional women may be applicable. It’s plausible that these less traditional women, who viewed themselves as possessing a number of instrumental traits (e.g., independent, active, competitive, and self-confident), could visualize themselves being successful equipment operators and had little empathy for a female equipment operator who performed the job badly.

**Positive and negative bias of nontraditional men toward female targets in traditional occupations and male targets in nontraditional occupations, respectively.** An unexpected instance of bias was evident in evaluations by men who disagreed with the notion that it is important for men to be tough. Because one would intuitively expect participants to negatively evaluate any target who exhibited poor job performance, the positive bias of these less traditional men toward both female receptionists and female equipment operators with poor job performance (see Figure 6) is difficult to explain. Moreover, these men evaluated the incompetent female receptionists significantly more positively than the incompetent female equipment operators. Why would these ostensibly nontraditional men more positively evaluate a female receptionist over a female equipment operator when both targets were exhibiting poor job performance? Unlike the more
traditional men (in regard to male status rather than toughness) who were negatively
biased against female equipment operators, these less traditional men (in regard to their
attitudes concerning toughness) seemed to be concerned about appearing to be sexist and
may have been overcompensating by positively evaluating all women with poor job
performance. Yet, their positive evaluations of women with poor job performance do not
explain the discrepancy between their significantly more positive evaluations of inept
female receptionists compared to their only slightly positive evaluations of inept female
equipment operators. This discrepancy may be due to a relationship between the less
traditional men’s fear of being sexist and the salience of the job’s gender stereotype. More
specifically, the salient factor when evaluating a female receptionist would be the target’s
gender but not the target’s job because it is not unusual for a woman to work as a
receptionist. Thus, when evaluating the female receptionist, these less traditional men
were likely focused on the target’s gender but not the target’s job, and they were reluctant
to negatively evaluate any female target. In contrast, when these men evaluated the poorly
performing female equipment operators, both the target’s gender and the gender
stereotype of the job would have been triggered, as it is unusual for a woman to work as a
heavy equipment operator. Thus, it is possible that the gender incongruence of the job
held by the female equipment operators was so salient for these men that it caused them to
be more thoughtful in their evaluations of the female equipment operator compared to the
female receptionist. The additional time that these less traditional men took to reconcile
their reluctance to negatively evaluate women with the actual poor job performance of the
female equipment operator may have resulted in a more sober assessment of the target’s
actual job performance.
In contrast to the positive bias these less traditional men exhibited toward the female targets, they were negatively biased toward male receptionists with poor job performance (see Figure 6). Initially, one might attribute this difference to a sort of “reverse double standard” in that these men seemed more accepting of a woman performing her job poorly than a man performing his job poorly. However, these men were neutral in their evaluations of the incompetent male equipment operator, so they were clearly not biased against men with poor job performance in general. Additionally, given that these men were uniformly negative in their evaluations of male receptionists, regardless of job performance (compare Figures 4-7), it is unlikely that they felt it was ever acceptable for a man to hold a job as a receptionist, even if the man performed his job well. Although one would expect targets with poor job performance to be evaluated negatively, the negative evaluation of male receptionists by these less traditional men seems greatly exaggerated in comparison to the other evaluations. It is interesting to note that the more traditional men who believed that it is important for men to be tough were actually more positive (but negative, nonetheless) in their evaluations of inept male receptionists (see Figure 6). This finding is puzzling and seems inconsistent with the other results which suggest that less traditional men would be positively biased toward targets in nontraditional occupations. However, there may be a logical explanation for why less traditional men would be more harsh in their evaluations of incompetent male receptionists than more traditional men were. It makes intuitive sense that any target performing his or her job badly would receive a negative evaluation, so perhaps the more traditional men were not biased at all but were simply evaluating the target based on the information provided regarding the target’s job performance (in this case a male receptionist who
performed his job particularly poorly). However, it’s difficult to reconcile the pattern of evaluations provided by the less traditional men with the intuition which suggests that these men should be more accepting of a man working as a receptionist than more traditional men would be. Perhaps if Shakespeare were alive he would explain the less traditional men’s evaluations by saying, “me thinks they doth protest too much.” Rather than a “reverse double standard,” it is possible that the negative evaluations of male receptionists by less traditional men reflect some type of “reverse reaction formation.”

Given that these men rejected the notion that men need to be physically and emotionally tough, it is unlikely that they considered themselves to be particularly “tough,” and I would expect that they would consciously claim that they approved of a man working as a receptionist. It is conceivable that these men were identifying with the male receptionists but also felt threatened in some way, causing them to distance themselves from the male receptionists by exaggerating their negative evaluations. Despite their contention that men do not need to be tough, their distorted negative evaluations of male receptionists (whom they likely did not picture as physically or emotionally tough when asked to imagine the target they would be evaluating) indicate that they may, at least unconsciously, feel contempt for men who are not tough. Although these men ostensibly believe that toughness is not an important quality for a man to possess, their evaluations clearly provide evidence to the contrary.

Implications and Conclusions

One of the strengths of this study was that it did not rely solely on participants’ gender to explain gender bias. Unlike many previous studies that were limited because they considered only situational variables, the present study included attitude and
personality variables that the experimenters hypothesized would influence evaluations of male and female targets. The results of this study provide valuable insight into the mechanisms of evaluative gender bias that previous studies focusing on in-group—out-group differences between men and women could not. With the exception of participants’ evaluations of targets who performed their jobs poorly, there were generally not significant differences between evaluations by men and women. This is consistent with Maurer and Taylor’s (1994) assertion that gender by itself is of limited value for documenting biased evaluations. However, given that there were not general differences between men’s and women’s evaluations, why were differences present when they evaluated targets with poor job performance? Women generally seemed more lenient than men when evaluating targets who performed their jobs badly. Why? Conceivably, women’s leniency could result from the common phenomenon of being perceived as less competent in the workplace than men. Perhaps women’s experiences of being viewed as less competent led them to discount information indicating that the targets in this study were performing their jobs incompetently. This explanation would be plausible in a real world evaluative situation, but in the artificial laboratory setting of this experiment, it seems unlikely that women would identify with faceless targets who performed their jobs badly to the point of actually disregarding the negative information provided about the targets’ job performance.

It is probable that the more lenient evaluations of women can be more accurately explained by the confounding factor of “femininity” rather than as a between-group difference resulting from gender category. The Femininity subscale of the PAQ was positively correlated with evaluations based on negative performance information. Thus,
participants who scored higher on Femininity tended to evaluate targets with poor job performance more favorably than participants possessing fewer “feminine” personality traits. Because women likely possess more “feminine” characteristics than men (at least in regard to extremity, if not in actual number), this seeming difference between men’s and women’s evaluations may actually be accounted for more accurately by the personality traits of “femininity” rather than gender category. But why would different levels of “femininity” influence evaluations? A review of the Femininity items on the PAQ may help explain the evaluative differences. It seems reasonable that participants who viewed themselves as kind, aware of the feelings of others, understanding of others, and helpful to others (all items on the Femininity subscale) would be more lenient, perhaps even favorable, when evaluating targets with poor job performance than participants who did not view themselves as kind and aware of the feelings of others. Thus, the ostensible between-group differences between men and women found in this study when participants evaluated poorly performing targets are likely due more to level of “femininity” than to gender category. Although the results initially seemed to indicate that gender differences were detected in at least one situation, though not generally, the differences in men’s and women’s evaluations found in the sole context of poor job performance were confounded by Femininity. Thus, this study may help explain the mixed results of previous studies, and it clearly provides further evidence that participant gender does not account for differential evaluations of male and female targets.

Rather than mistakenly attribute the differences in evaluations to gender differences, the gender bias of participants in the present study is more accurately accounted for by the complex relationships between the situational variables and the
attitude and personality variables included in the design. For example, both women and men who were more traditional in terms of gender-role ideology were negatively biased against targets working in nontraditional occupations. This finding is particularly concerning because of the clear implications it has for hiring and promotion decisions. For instance, one can see how “the old boy network” perpetuates itself because “the old boys” in power may tend to be negatively biased toward outsiders who do not share, and may even threaten, their ideology. Thus, if “the old boy network” does not approve of certain individuals, those individuals will probably not be hired, but even if they are hired, it is unlikely that they would be promoted.

One of the limitations of this study is that it is not clear whither these traditional individuals were cognizant of their negative biases. Designing and implementing effective interventions for reducing gender bias, requires that researchers be clear of the reasoning behind the evaluations of biased individuals. For example, as much of the research in social cognition has documented (e.g., Crosby, Bromley, & Saxe, 1980; Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995; Gaertner & McLaughlin, 1983; Greenwald & Banaji, 1995; Wittenbrink, Judd, & Park, 1997; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Vanman, Paul, Ito, & Miller, 1997), it is common for individuals to exhibit bias without being consciously aware of their bias. Thus, one of the strategies for reducing gender bias in individuals who are not aware that they are biased could be some form of consciousness raising. Different strategies would be required, however, to reduce the gender bias of individuals who are aware of their biases and may even feel justified in their biased evaluations.
This study represents a new direction in gender bias research, and future researchers should heed Maurer and Taylor’s (1994) warning that gender by itself is of limited value for documenting biased evaluations. The findings presented here offer compelling evidence that evaluative gender bias is influenced by various attitudes and personality traits of the evaluators rather than by the gender of the evaluators. Although the generalizability of these results is unknown at this time, this study warrants further investigation, especially in real-world evaluative contexts. The analyses in this study are challenging, but if we are ever going to have a clear understanding of the factors contributing to gender bias, researchers will have to move beyond simple experimental designs that lack real-world validity and design studies capable of explaining the complicated nature of gender bias as it exists in reality. Much more research is needed, but this study marks an excellent beginning.
EESQ

(Note: This version of the EESQ features a male target in a masculine job. Each version is identical with the exception of the description of the person (below) that participants were asked to evaluate. The only information that varied across the four versions was the target’s gender (male or female) and job (heavy equipment operator or receptionist in a law firm.)

PLEASE PLEASE!! DO NOT WRITE ON THIS QUESTIONNAIRE. ANSWER BY MARKING THE APPROPRIATE ITEM ON THE GREEN ANSWER SHEET LABELED “MM” PLEASE INDICATE YOUR GENDER (MALE OR FEMALE) ON THE ANSWER SHEET.

Evaluating and Explaining Another Person’s Job Performance

We are interested in how people evaluate other people’s job performances. On the questionnaire that follows, you will be given a person’s performance ratings on a job and some specific information about those ratings that we want you to use to evaluate that person’s performance.

To help you imagine an actual person in an actual job, we have provided a brief descriptive paragraph of the person we want you to evaluate and the job that they have. Please read this description carefully and take a few moments to imagine a person like the one described below. To ensure that you have read the information carefully, we will be asking you a few questions about the person and their job after you have completed evaluating the person’s performance.

After you have read the description and thought about the person, turn to the next page and read about how we want you to explain and evaluate that performance.

This is the person whose performance we want you to evaluate:

This person is 25 years old. This person is not a college graduate but has completed two years of college study. This person enjoys going to movies and listening to music. This person is male and works full time as a heavy equipment (e.g., bulldozer) operator. This person has a brother and a sister, and both of the person’s parents are living. This person’s family is best described as middle class.

This is the person whose job performance we want you to evaluate. Take a few moments and try to imagine a person like the one we described doing the job we have indicated. Then turn the page for more instructions.

REMEMBER: It is the performance of this person in this job that you are to evaluate.)
Evaluating Job Performance

This questionnaire contains 27 items that provide different types of information that might be useful to you for evaluating the person’s job performance ratings given below. This person and job were described on the first page of this questionnaire. If you need to refresh your memory about this person or the job now or anytime while completing the questionnaire, please turn back and reread the description.

The person received a job performance rating of 22 after completing one year of work at the company.

We are purposefully not telling you anything about the rating scale, so you do not know if 22 is a high or low rating by itself. In fact, we do not want you to assume anything about the scale.

Instead, in each of the items we will provide you with information about the person’s job performance rating relative to some standard. Specifically, you will be told how the person performed:

1) relative to other people,
2) relative to how you expected the person to perform,
and/or
3) relative to that person’s performance ratings from other jobs.

Thus, each item will contain either one, two, or three types of relative performance information.

Please think about all of the information provided in an item before responding to that item.

Although each of the items involves the same performance rating, each item will contain different information about the event and you should treat each item separately.

On each of the following pages, read the information in each of the items carefully, taking a moment to imagine the person described on the previous page, and thinking about the information in the item. Then rate how you would evaluate that person’s performance by choosing a number on the following scale:

Based on all the information in this item I would evaluate this person’s performance as:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Poor</td>
<td>Quite Poor</td>
<td>Slightly Poor</td>
<td>Neutral</td>
<td>Slightly Good</td>
<td>Quite Good</td>
<td>Extremely Good</td>
</tr>
</tbody>
</table>

Remember, although they concern the same event, each item is independent of the others. Therefore, your response to each item should be made without any influence from the other items. To emphasize that they are separate, each item is separated by a bold line and only a few items appear on each page.
(Note: The event and rating scale are provided for each item. However, in an effort to conserve space, only items 1-4 are presented in their entirety, exactly as they appear on the EESQ that participants used. Because the only information that varies across items is the information provided to interpret the rating, items 5-27 listed in Appendix A include only the evaluative information participants used to interpret and evaluate the event.)

Event: This person received a performance rating of 22 after completing one year of work at this company.

1. **This rating is much lower than the performance rating you expected this person to receive.**

   Based on all this information I would evaluate the person’s performance as......

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Quite</td>
<td>Slightly</td>
<td>Neutral</td>
<td>Slightly</td>
<td>Quite</td>
<td>Extremely</td>
</tr>
<tr>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

Event: This person received a performance rating of 22 after completing one year of work at this company.

2. **This rating is much lower than performance ratings this person has received on any other job.**

   Based on all this information I would evaluate the person’s performance as......

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Quite</td>
<td>Slightly</td>
<td>Neutral</td>
<td>Slightly</td>
<td>Quite</td>
<td>Extremely</td>
</tr>
<tr>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

Event: This person received a performance rating of 22 after completing one year of work at this company.

3. **This rating is much higher than most of the performance ratings given to other employees.**

   Based on all this information I would evaluate the person’s job performance as......

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Quite</td>
<td>Slightly</td>
<td>Neutral</td>
<td>Slightly</td>
<td>Quite</td>
<td>Extremely</td>
</tr>
<tr>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>
Event: This person received a performance rating of 22 after completing one year of work at this company.

4. This rating is much lower than performance ratings this person has ever received on any other job.

and

This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is much higher than most of the performance ratings given to other employees.

Based on all this information I would evaluate the person’s performance as.....

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Poor</td>
<td>Quite Poor</td>
<td>Slightly Poor</td>
<td>Neutral</td>
<td>Slightly Good</td>
<td>Quite Good</td>
<td>Extremely Good</td>
</tr>
</tbody>
</table>

5. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is much lower than most of the performance ratings given to other employees.

6. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is much higher than most of the performance ratings given to other employees.

and

This rating is the highest performance rating that this person has ever received on any job.
7. This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is much lower than most of the performance ratings given to other employees.

8. This rating is the highest performance rating that this person has ever received on any job.

9. This rating is much lower than performance ratings this person has ever received on any job.

and

This rating is much higher than most of the performance ratings given to other employees.

and

This rating is much lower than the performance rating you expected this person to receive.

10. This rating is much higher than most of the performance ratings given to other employees.

and

This rating is much lower than performance ratings this person has ever received on any job.

11. This rating is much lower than most of the performance ratings given to other employees.

and

This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is much lower than performance ratings this person has ever received on any job.
12. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is much higher than most of the performance ratings given to other employees.

13. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is much lower than performance ratings this person has received on any other job.

14. This rating is much lower than most of the performance ratings given to other employees.

15. This rating is much lower than performance ratings this person has ever received on any job.

and

This rating is much lower than the performance rating you expected this person to receive.

and

This rating is much lower than most of the performance ratings given to other employees.

16. This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is the highest performance rating that this person has ever received on any job.

17. This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is much higher than most of the performance ratings given to other employees.
18. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is the highest performance rating that this person has ever received on any job.

and

This rating is much lower than most of the performance ratings given to other employees.

19. This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is lower than performance ratings this person has received on any other job.

20. This rating is much higher than most of the performance ratings given to other employees.

and

This rating is the highest performance rating that this person has ever received on any job.

21. This rating is higher than the highest performance rating you thought this person could possibly receive.

and

This rating is much lower than most of the performance ratings given to other employees.

and

This rating is the highest performance rating that this person has ever received on any job.
22. This rating is lower than performance ratings this person has received on any other job.

and

This rating is much lower than most of the performance ratings given to other employees.

23. This rating is higher than the highest performance rating you thought this person could possibly receive.

24. This rating is much lower than the performance rating you expected this person to receive.

and

This rating is the highest performance rating that this person has ever received on any job.

25. This rating is much lower than most of the performance ratings given to other employees.

and

This rating is the highest performance rating that this person has ever received on any job.

26. This rating is the highest performance rating that this person has ever received on any job.

and

This rating is much higher than most of the performance ratings given to other employees.

and

This rating is higher than the highest performance rating you thought this person could possibly receive.

27. (no information given)
Manipulation Check

What are the Characteristics of the Person
Whose Job Performance You Just Evaluated?

Please answer each of the questions by circling the answer you believe is correct. If you are unsure of the answer, make your best guess.

1) How old is this person?
   a) 20
   b) 25
   c) 18
   d) Was not told

2) What is this person’s gender?
   a) Male
   b) Female
   c) Was not told

3) Does the person have a college degree?
   a) Yes
   b) No
   c) Was not told

4) What is the person’s job?
   a) Day Care Worker
   b) Heavy Equipment Operator
   c) Repairing Telephone Lines
   d) Receptionist

5) Are both of the person’s parent’s living?
   a) No
   b) Yes
   c) Was not told

6) This person’s family is
   a) Lower Middle Class
   b) Upper Middle Class
   c) Middle Class
   d) Was not told
Beliefs About Women

Below are a series of statements concerning women and their role in contemporary society. Please indicate the degree to which you agree or disagree with each statement using the following scale:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>disagree</td>
<td>disagree</td>
<td>disagree</td>
<td>neutral</td>
<td>agree</td>
<td>agree</td>
<td>agree</td>
</tr>
<tr>
<td>strongly</td>
<td>somewhat</td>
<td>slightly</td>
<td></td>
<td>slightly</td>
<td>somewhat</td>
<td>strongly</td>
</tr>
</tbody>
</table>

1. No matter how accomplished he is, a man is not complete unless he has the love of a woman.
2. Many women seek special favors, such as hiring policies that favor them over men, by using the excuse of asking for “equality.”
3. In a disaster, women should not necessarily be rescued before men.
4. Most women interpret innocent remarks or acts as being sexist.
5. Women are too easily offended.
6. People are often truly happy in life without being romantically involved with a member of the opposite sex.
7. Feminists are not asking that women have more power than men.
8. Many women have a quality of purity that few men possess.
9. Women should be cherished and protected by men.
10. Most women fail to appreciate fully all that men do for them.
11. Women seek to gain power by getting control over men.
12. Every man ought to have a woman whom he adores.
13. Men are complete without women.
14. Women exaggerate problems they have at work.
15. Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.
16. When women lose to men in a fair competition, they typically complain about being discriminated against.
17. A good woman should be set on a pedestal by her man.
18. There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing men’s advances.
19. Women, compared to men, tend to have a superior morals.
20. Men should be willing to sacrifice their own well being to provide financially for the women in their lives.
21. Feminists are making entirely reasonable demands of men.
22. Women, as compared to men, tend to have a more refined sense of culture and good taste.
Beliefs about Men

Below are a series of statements concerning men and their roles in contemporary society. Please indicate the degree to which you agree or disagree with each statement using the following scale:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>disagree</td>
<td>disagree</td>
<td>disagree</td>
<td>neutral</td>
<td>agree</td>
<td>agree</td>
<td>agree</td>
</tr>
<tr>
<td>strongly</td>
<td>somewhat</td>
<td>slightly</td>
<td></td>
<td>slightly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Success in his work has to be a man’s central goal in life.
2. When a man is feeling a little pain he should try not to let it show very much.
3. It bothers me when a man does something I consider feminine.
4. The best way for a young man to get the respect of other people is to get a job, take it seriously, and do well.
5. People do not have much respect for a man who frequently talks about his worries, fears, and problems.
6. I probably would not like a man whose hobbies are cooking, sewing, and going to the ballet.
7. A man owes it to his family to work at the best paying job he can get.
8. A good motto for a man would be “When the going gets tough, the tough get going.”
9. A man should generally work overtime to make more money whenever he has the chance.
10. It is a bit embarrassing for a man to have a job that is usually filled by a woman.
11. A man always deserves the respect of his wife and children.
12. I think a young man should try to become physically tough, even if he is not big.
13. It is essential for a man to always have the respect and admiration of everyone who knows him.
14. Unless he was really desperate, I would probably advise a man to keep looking rather than accept a job as a secretary.
15. A man should never back down in the face of trouble.
16. I always like a man who is totally sure of himself.
17. Fists are sometimes the only way out of a bad situation.
18. If I heard about a man who was a hairdresser and a gourmet cook, I might wonder how masculine he was.
19. A man should always think everything out coolly and logically, and have a rational reason for everything he does.
20. A real man enjoys a bit of danger now and then.
21. I think it is extremely good for a boy to be taught to cook, sew, clean house, and take care of younger children.
22. A man should always try to project an air of confidence even if he really doesn’t feel confident inside.
23. In some situations a man should be ready to use his fists, even if his wife or girlfriend would object.
24. A man must stand on his own two feet and never depend on other people to help him do things.
25. I might find it a little silly or embarrassing if a male friend of mine cried over a sad love scene in a movie.
26. A man should always refuse to get into a fight, even if there seems to be no way to avoid it.
PAQ

Personal Attributes Questionnaire

The items below concern what kind of person you think you are. Each item consists of a pair of characteristst, with the numbers 1-5 in between. For example:

Not at all artistic  1...2....3....4....5  Very artistic

Each pair describes contradictory characteristics—that is, you cannot be both at the same time, such as very artistic and not at all artistic.

The numbers form a scale between the two extremes. You are to choose a number that describes where you fall on the scale. For example, if you think you have no artistic ability, you would choose 1. If you think you are fairly artistic, you might choose 4. If you are neither particularly artistic nor unartistic, you might choose 3, and so forth.

1. Not at all aggressive..........................1...2...3...4...5......................Very aggressive
2. Not at all independent..........................1...2...3...4...5......................Very independent
3. Not at all emotional.............................1...2...3...4...5......................Very emotional
4. Very submissive..................................1...2...3...4...5......................Very dominant
5. Not at all excitable in a major crisis...1...2...3...4...5......................Very excitable in a major crisis
6. Very passive ...................................1...2...3...4...5......................Very active
7. Not at all able to devote self completely to others..........................1...2...3...4...5...................... Able to devote self completely to others
8. Very rough......................................1...2...3...4...5......................Very gentle
9. Not at all helpful to others..........................1...2...3...4...5......................Very helpful to others
10. Not at all competitive..........................1...2...3...4...5......................Very competitive
11. Very home oriented............................1...2...3...4...5......................Very worldly
12. Not at all kind..................................1...2...3...4...5......................Very kind
13. Indifferent to others’ approval................1...2...3...4...5......................Highly needful of others’ approval
14. Feelings not easily hurt..........................1...2...3...4...5......................Feelings easily hurt
15. Not at all aware of feelings of others......1...2...3...4...5......................Very aware of feelings of others
16. Can easily make decisions.....................1...2...3...4...5......................Have difficulty making decisions
17. Gives up very easily............................1...2...3...4...5......................Never gives up
18. Never cries......................................1...2...3...4...5......................Cries very easily
19. Not at all self-confident.......................1...2...3...4...5......................Very self-confident
20. Feels very inferior.............................1...2...3...4...5......................Feels very superior
21. Not at all understanding of others...........1...2...3...4...5......................Very understanding of others
22. Very cold in relations with others............1...2...3...4...5......................Very warm in relations with others
23. Very little need for security...............1...2...3...4...5......................Very strong need for security
24. Goes to pieces under pressure...............1...2...3...4...5......................Stands up well under pressure
APPENDIX B

TABLES

Table 1

Number of Participants included in the Final Sample.

<table>
<thead>
<tr>
<th>Job</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target Gender</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Masculine</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Feminine</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Totals</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>126</td>
</tr>
</tbody>
</table>
### Table 2

Means and Standard Deviations for the 27 Evaluative Items in Descending Order

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPP</td>
<td>6.58</td>
<td>0.81</td>
</tr>
<tr>
<td>POP</td>
<td>6.34</td>
<td>0.77</td>
</tr>
<tr>
<td>PPO</td>
<td>6.26</td>
<td>0.81</td>
</tr>
<tr>
<td>OPP</td>
<td>6.03</td>
<td>1.04</td>
</tr>
<tr>
<td>POO</td>
<td>5.75</td>
<td>0.98</td>
</tr>
<tr>
<td>OPO</td>
<td>5.64</td>
<td>1.07</td>
</tr>
<tr>
<td>PNP</td>
<td>5.64</td>
<td>1.02</td>
</tr>
<tr>
<td>PPN</td>
<td>5.57</td>
<td>1.04</td>
</tr>
<tr>
<td>OOP</td>
<td>5.49</td>
<td>1.13</td>
</tr>
<tr>
<td>PNO</td>
<td>4.87</td>
<td>1.04</td>
</tr>
<tr>
<td>PON</td>
<td>4.87</td>
<td>1.11</td>
</tr>
<tr>
<td>PNN</td>
<td>4.54</td>
<td>1.13</td>
</tr>
<tr>
<td>NPP</td>
<td>4.47</td>
<td>1.25</td>
</tr>
<tr>
<td>OPN</td>
<td>4.33</td>
<td>1.06</td>
</tr>
<tr>
<td>OOO</td>
<td>4.31</td>
<td>0.86</td>
</tr>
<tr>
<td>ONP</td>
<td>4.28</td>
<td>1.10</td>
</tr>
<tr>
<td>NOP</td>
<td>3.90</td>
<td>1.21</td>
</tr>
<tr>
<td>NPO</td>
<td>3.86</td>
<td>1.20</td>
</tr>
<tr>
<td>NNP</td>
<td>3.52</td>
<td>1.35</td>
</tr>
<tr>
<td>NPN</td>
<td>3.40</td>
<td>1.12</td>
</tr>
<tr>
<td>OON</td>
<td>3.25</td>
<td>1.32</td>
</tr>
<tr>
<td>ONO</td>
<td>2.99</td>
<td>1.14</td>
</tr>
<tr>
<td>ONN</td>
<td>2.56</td>
<td>1.19</td>
</tr>
<tr>
<td>NOO</td>
<td>2.39</td>
<td>0.86</td>
</tr>
<tr>
<td>NON</td>
<td>2.39</td>
<td>1.12</td>
</tr>
<tr>
<td>NNO</td>
<td>2.23</td>
<td>1.04</td>
</tr>
<tr>
<td>NNN</td>
<td>1.67</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Note. N = 247. Evaluations are based on a 7-point scale, with 1 being “extremely poor” and 7 being “extremely good.”

How to read Table 1 abbreviations
Standards are placed in the following order: normative, ipsative, expectations; P = positive, O = neutral (i.e., absent), N = negative.

Thus, PPP represents the item containing positive normative, positive ipsative, and positive expectation information; PON represents the item containing positive normative, neutral (or absent) ipsative, and negative expectation information, and ONP represents the item containing neutral (or absent) normative, negative ipsative, and positive expectation information.
<table>
<thead>
<tr>
<th>Scale</th>
<th>General</th>
<th>Positive</th>
<th>Negative</th>
<th>Mixed</th>
<th>Neutral Item</th>
<th>Mean</th>
<th>SD</th>
<th>Coefficient Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>.44</td>
<td>.55</td>
<td>-.35</td>
<td>----</td>
<td>4.34</td>
<td>6.01</td>
<td>.60</td>
<td>.73</td>
</tr>
<tr>
<td>Neutral</td>
<td>.89</td>
<td>.34</td>
<td>.30</td>
<td>----</td>
<td>4.44</td>
<td>2.50</td>
<td>.74</td>
<td>.79</td>
</tr>
<tr>
<td>Item</td>
<td>.45</td>
<td>.16</td>
<td>.25</td>
<td>.34</td>
<td>4.31</td>
<td>----</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.  N = 247
Evaluative ratings were made on a 7-point scale ranging from 1 = Extremely Poor Performance to 7 = Extremely Good Performance.
Table 4

Means, Standard Deviations, and Coefficient Alphas for the Subscales from the Ambivalent Sexism Inventory (ASI), Masculine Role Norms Scale (MRNS), and Personal Attributes Questionnaire (PAQ)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Men</th>
<th>Women</th>
<th>Coefficient Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostile Sexism</td>
<td>2.85 (1.00)</td>
<td>2.36 (1.03)</td>
<td>.84</td>
</tr>
<tr>
<td>Benevolent Sexism</td>
<td>3.07 (0.91)</td>
<td>2.59 (0.95)</td>
<td>.76</td>
</tr>
<tr>
<td><strong>MRNS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toughness</td>
<td>3.08 (0.93)</td>
<td>2.12 (1.03)</td>
<td>.77</td>
</tr>
<tr>
<td>Anti-Femininity</td>
<td>2.48 (1.18)</td>
<td>1.75 (1.12)</td>
<td>.81</td>
</tr>
<tr>
<td>Status</td>
<td>2.89 (1.05)</td>
<td>2.39 (1.00)</td>
<td>.82</td>
</tr>
<tr>
<td><strong>PAQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculinity</td>
<td>3.78 (0.49)</td>
<td>3.40 (0.54)</td>
<td>.71</td>
</tr>
<tr>
<td>Femininity</td>
<td>3.87 (0.46)</td>
<td>4.27 (0.40)</td>
<td>.70</td>
</tr>
<tr>
<td>Masculinity-Femininity</td>
<td>3.09 (0.48)</td>
<td>2.54 (0.53)</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note. N = 121 men and 126 women. Standard deviations are in parentheses. Coefficient alpha is based on the total sample (n = 247). For the ASI and the MRNS, ratings were on a 7-point scale ranging from 0 to 6. For the PAQ, ratings were on a 5-point scale ranging from 1 to 5.
Table 5

Means and Standard Deviations for the General and Negative Evaluational Bias Scores for each of the Target x Job Conditions.

### General Evaluational Bias

<table>
<thead>
<tr>
<th>Target*</th>
<th>Male</th>
<th>Female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-.05 (.41)</td>
<td>.13 (.41)</td>
<td>.04 (.42)</td>
</tr>
<tr>
<td>Job</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-.06 (.37)</td>
<td>-.02 (.31)</td>
<td>-.04 (.34)</td>
</tr>
<tr>
<td></td>
<td>-.06 (.39)*</td>
<td>.06 (.37)*</td>
<td></td>
</tr>
</tbody>
</table>

### Negative Evaluational Bias

<table>
<thead>
<tr>
<th>Target**</th>
<th>Male</th>
<th>Female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-.19 (.59)</td>
<td>.27 (.82)</td>
<td>.05 (.75)</td>
</tr>
<tr>
<td>Job</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-.13 (.78)</td>
<td>.04 (.66)</td>
<td>-.04 (.73)</td>
</tr>
<tr>
<td></td>
<td>-.16 (.70)**</td>
<td>.16 (.75)**</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 247. Standard deviations are in parentheses. Positive Means indicate that the target was evaluated more positively, on the average, in that condition, whereas negative Means indicate that the target was evaluated more negatively. There was a significant main effect for Target for both the General and Negative scales, at the .05* and .01** levels, respectively.
Table 6

Correlations Between the Individual Difference Variables and the Five Bias Measures

<table>
<thead>
<tr>
<th>Individual Differences</th>
<th>General</th>
<th>Positive</th>
<th>Negative</th>
<th>Mixed</th>
<th>Neutral Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>.10</td>
<td>.01</td>
<td><strong>.16</strong></td>
<td>.02</td>
<td>.12</td>
</tr>
<tr>
<td>Ambivalent Sexism Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostile Sexism</td>
<td>.01</td>
<td>.02</td>
<td>-.09</td>
<td>.07</td>
<td>.06</td>
</tr>
<tr>
<td>Benevolent Sexism</td>
<td>-.05</td>
<td>-.03</td>
<td>-.12</td>
<td>.02</td>
<td>.05</td>
</tr>
<tr>
<td>Masculine Role Norms Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toughness</td>
<td>-.07</td>
<td>.02</td>
<td><strong>-.17</strong></td>
<td>-.01</td>
<td>.04</td>
</tr>
<tr>
<td>Anti-Femininity</td>
<td>-.11</td>
<td>-.03</td>
<td>-.12</td>
<td>-.06</td>
<td>.01</td>
</tr>
<tr>
<td>Status</td>
<td>-.04</td>
<td>.01</td>
<td>-.08</td>
<td>-.02</td>
<td>.02</td>
</tr>
<tr>
<td>Personal Attributes Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculinity</td>
<td><strong>-.19</strong></td>
<td>-.00</td>
<td><strong>-.19</strong></td>
<td><strong>-.16</strong></td>
<td>-.03</td>
</tr>
<tr>
<td>Femininity</td>
<td><strong>.23</strong></td>
<td>.10</td>
<td><strong>.17</strong></td>
<td><strong>.18</strong></td>
<td>.05</td>
</tr>
<tr>
<td>Masculinity-Femininity</td>
<td><strong>-.14</strong></td>
<td>-.01</td>
<td><strong>-.14</strong></td>
<td>-.11</td>
<td>-.12</td>
</tr>
</tbody>
</table>

Note. N = 247.
Correlations in bold-face are significant at the .05* or the .01** level, two-tailed test.
APPENDIX C

FIGURES

Figure 1

Plot of the Cell Means from the $2 \times 2 \times 2$ (Gender x Target x Job) Analysis of Variance of the Evaluative Ratings Based on the Negative Evaluative Information
Figure 2

Plot of the Predicted Values from the Gender x Target x Job x Benevolent Sexism Hierarchical Regression Analysis of Evaluative Ratings based on Positive Evaluative Information
Plot of the Predicted Values from the Target x Job x Benevolent Sexism Hierarchical Regression Analysis of the Men's Evaluative Ratings Based on Positive Evaluative Information

Men Only

Figure 3

Centered Evaluative Ratings

High Benevolent Sexism Low Benevolent Sexism

Female Receptionist
Female Equipment Operator
Male Receptionist
Male Equipment Operator
Figure 4

Plot of the Predicted Values from the Gender x Target x Job x Toughness Hierarchical Regression Analysis of Evaluative Ratings based on General Evaluative Information.
Figure 5

Plot of the Predicted Values from the Gender x Target x Job x Toughness Hierarchical Regression Analysis of Evaluative Ratings based on Negative Evaluative Information
Figure 6

Plot of the Predicted Values from the Target x Job x Toughness Hierarchical Regression Analysis of Men's Evaluative Ratings Based on Negative Evaluative Information

Men Only

Centered Evaluative Ratings

High Toughness Low Toughness

& Female Receptionist

Female Equipment Operator

# Male Equipment Operator

@ Male Receptionist
Plot of the Predicted Values from the Gender x Target x Job x Toughness Hierarchical Regression Analysis of Evaluative Ratings based on Mixed Evaluative Information

Figure 7
Figure 8

Plot of the Predicted Values from the Target x Job x Toughness Hierarchical Regression Analysis of the Women's Evaluative Ratings Based on No Evaluative Information

Women Only

Centered Evaluator Ratings

High Toughness Low Toughness

Female Equipment Operator

Female Receptionist

Male Receptionist

Male Equipment Operator
Figure 9

Plot of the Predicted Values from the Gender x Target x Job x Status Hierarchical Regression Analysis of Evaluative Ratings based on Positive Evaluative Information
Figure 10

Plot of the Predicted Values from the Target x Job x Status Hierarchical Regression Analysis of the Men’s Evaluative Ratings Based on Positive Evaluative Information

Men Only

Centered Evaluative Ratings

High Status Low Status

Male Receptionist
Female Equipment Op.
Male Equipment Operator
Female Receptionist

-25 -21 -17 -13 -09 -05 -01 .03 .07 .11 .15 .19 .23 .27 .31
Figure 11

Plot of the Predicted Values from the Target x Job x Masculinity Hierarchical Regression Analysis of the Women's Evaluative Ratings Based on General Evaluative Information

Women Only

- Female Equipment Operator
- Male Recept.
- Female Recept.
- Male Equipment Operator

Centered Evalutative Ratings
High Masculinity Low Masculinity
Figure 12

Plot of the Predicted Values from the Target x Job x Masculinity Hierarchical Regression Analysis of Women’s Evaluative Ratings Based on Negative Evaluative Information


