ORGANIZATIONAL CITIZENSHIP BEHAVIOR, GENDER, AND PERFORMANCE RATINGS: DOES THE RATER MATTER?

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Psychology

with a Concentration in

Applied Psychology

by

Bahareh Soltani

Summer 2013
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the
Thesis of Bahareh Soltani:

Organizational Citizenship Behavior, Gender, and Performance Ratings:

Does the Rater Matter?

Mark Ehrhart, Chair
Department of Psychology

Lisa M. Kath
Department of Psychology

Amy Randel
Management Department

5/13/13
Approval Date
Copyright © 2013
by
Bahareh Soltani
All Rights Reserved
DEDICATION

I dedicate my thesis to my family and friends, who have never faltered in their support of my many endeavors. My most sincere gratitude goes to my father, Mohammad Reza Soltani, for continually encouraging me to become the best version of myself. I would also like to express my gratitude to my mother, Zari Soltani, my sisters, Taraneh, Sanaz, and Sara Soltani, and my many friends who have supported me throughout my graduate work and who kept me grounded in my most stressful moments.

I also dedicate this thesis to Dr. Lisa Kath, who has continually championed for me throughout my professional and personal development. Your mentorship has meant the world to me.
ABSTRACT OF THE THESIS

Organizational Citizenship Behavior, Gender, and Performance Ratings: Does the Rater Matter?
by
Bahareh Soltani
Master of Science in Psychology with a Concentration in Applied Psychology
San Diego State University, 2013

Although many strides have been made with regard to gender differences over the past few decades, there are still discrepancies in treatment and expectations of women and men in the workplace that warrant further research. One such difference is evident in perceptions of organizational citizenship behaviors, or OCBs, which are organizationally relevant actions taken by employees that go above and beyond their formal job requirements. Following the logic of gender role theory, several studies have examined how cultural expectations regarding gender roles affect the way individuals’ performance of different various OCBs are evaluated and rewarded. One potential connection that has not received much attention in the literature thus far focuses not on the employee performing OCBs, but the person who is using that information to form a performance rating of the employee. In light of mixed findings regarding the role for rater gender, this thesis hypothesized that the raters’ gender personal identity, gender social identity, and modern sexist beliefs will affect their ratings of the overall performance of females and males exhibiting varying levels of helping behavior (a stereotypically feminine OCB) and voice (a stereotypically masculine OCB). Two models come to fruition here. In one, the stronger a rater’s gender personal identity and/or modern sexist beliefs, the more likely she or he will be to make judgments with traditional gender norms in mind, thus providing lower ratings for women who perform stereotypically masculine OCBs. Alternatively, the stronger a rater’s gender social identity and/or gender identity salience, the more likely she or he will be to make judgments with the ratee’s position as part of their gender’s in- or out-group in mind, thus providing lower ratings to ratees in the out-group.

The hypothesized models were investigated in two studies using a sample of currently employed college students (Study 1: n=234, Study 2: n=175). The first was an experimental policy capturing study, in which participants read scenarios describing the behavior of hypothetical employees and rated their overall performance. In the second study, participants rated the OCBs and overall performance of one female and one male coworker. Both studies measured the gender personal identity, gender social identity, gender identity salience, and modern sexism of the performance rater. Hierarchical linear modeling was conducted to analyze the hypotheses. The results supported the hypotheses that OCBs would significantly affect overall performance ratings. Additionally and contrary to expectations, the results suggested that male targets are punished more than female targets when they are low in either voice behaviors or task performance. No support was found for the remaining hypothesized interactions.
Although there were few significant findings in this thesis, there is still an abundance of research that suggests that OCBs are important because they affect decisions that influence career advancement. Considering the influence of these behaviors, it is vital to continue investigating how OCBs contribute to overall performance ratings in the workplace. Doing so may shed light on how performance ratings are formed, as well as how those ratings may eventually impact gender gaps in career advancement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INTRODUCTION</td>
<td>Gender and Organizational Citizenship Behavior</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Organizational Citizenship Behavior and Performance Ratings</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Predictors of Gender Stereotyping</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Rater Gender</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Rater Gender Personal Identity</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Rater Gender Social Identity</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Rater Gender Identity Salience</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Rater Modern Sexism</td>
<td>18</td>
</tr>
<tr>
<td>2 METHOD</td>
<td>Participants</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Study 1 Procedure</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Study 1 Materials</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Policy Capturing Scenarios</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Performance Ratings</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Gender Personal Identity</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Gender Social Identity</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Gender Identity Salience</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Modern Sexism</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>OCB Expectations by Gender</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Study 1 Statistical Analyses</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Study 2 Procedure</td>
<td>27</td>
</tr>
</tbody>
</table>
Study 2 Materials .................................................................29
  Overall Performance ......................................................29
  Task Performance ..........................................................29
  Self-Ratings of OCB .......................................................30
  Other-Ratings of OCB .....................................................30
  Job Interdependence ......................................................30
Study 2 Statistical Analyses ..................................................31

3 RESULTS ........................................................................32
  Initial Analyses ..............................................................32
  Study 1 ........................................................................33
  Study 2 ........................................................................39

4 DISCUSSION ..................................................................45
  Initial Analyses ..............................................................45
  Studies 1 and 2 ..............................................................46
  Limitations ....................................................................50
  Future Directions ..........................................................52
  Conclusion ......................................................................53

REFERENCES ..................................................................55

APPENDIX
  A POLICY CAPTURING SCENARIOS AND PERFORMANCE RATINGS ....61
  B GENDER PERSONAL IDENTITY ........................................66
  C GENDER SOCIAL IDENTITY ............................................68
  D GENDER IDENTITY SALIENCE .....................................70
  E MODERN SEXISM SCALE .............................................72
  F OCB EXPECTATIONS BY GENDER – SINGLE ITEM MEASURE ......74
  G OCB EXPECTATIONS BY GENDER – MULTIPLE ITEM MEASURE ....76
  H OVERALL PERFORMANCE: SOLTANI ................................78
  I OVERALL PERFORMANCE: WILLIAMS & ANDERSON ..............80
  J SELF-RATINGS OF OCB ...............................................82
  K OTHER-RATINGS OF OCB ............................................84
  L JOB INTERDEPENDENCE .............................................86
LIST OF TABLES

Table 1. Means, Standard Deviations, and Intercorrelations among between-Subjects Variables ......................................................................................................................33

Table 2. Level 1 Analysis: The Effects of Scenario Conditions on Overall Performance Ratings ..................................................................................................................35

Table 3. Level 1 Analysis: The Effect of Coworker Gender and OCBs on Overall Performance Ratings ..................................................................................................................40
LIST OF FIGURES

Figure 1. Effect of ratee gender and voice behavior on performance ratings (Hypothesis 2b). ...................................................................................................................................................36

Figure 2. Effect of ratee gender and task performance on performance ratings. ..................36
ACKNOWLEDGEMENTS

I would like to give special acknowledgment to my advisor, Dr. Mark Ehrhart, for his continual guidance throughout the process of completing my thesis. Dr. Ehrhart’s support and expertise allowed me to conduct research in an area that I am passionate about, and I am eternally grateful for that.

I would also like to extend my gratitude to the members of my thesis committee, Drs. Lisa Kath and Amy Randel, who have generously shared their time and insights for the benefit of my thesis. Their contributions and overall positive spirit are greatly appreciated.

Many thanks also go out to those individuals who assisted with the processes of collecting data and running statistical analyses for my thesis. I would specifically like to thank the undergraduate research assistants in Dr. Mark Ehrhart’s ROLE Lab (Jaime Cummings, Erika Maccani, Sandra Martinez, Lauren Ostroski, and Nicholas Tran), who ran numerous data collection sessions, as well as Abdifatah Ali and Dr. Scott Roesch, who helped me navigate through the virtual world of Mplus.
CHAPTER 1

INTRODUCTION

According to the U.S. Bureau of Labor (2010), women make up about 47% of all employed individuals in the nation. Although many strides have been made with regard to gender differences over the past few decades (for example, women earned 62% of what men earned in 1979 and 80% of what men earned in 2009; U.S. Bureau of Labor, 2010), there are still discrepancies in treatment and expectations of women and men in the workplace that warrant further research. One such difference is evident in perceptions of engagement in organizational citizenship behaviors, or OCBs, which are defined as organizationally relevant actions taken by employees that go above and beyond the formal requirements of their jobs (Smith, Organ, & Near, 1983).

Although there is not complete consensus regarding the dimensionality of OCB, a review by P. M. Podsakoff, MacKenzie, Paine, and Bachrach (2000) identified seven common dimensions. Helping behavior includes volunteering to help others to solve or inhibit the incidence of work-related problems. Individuals who engage in sportsmanship behavior do not complain in the face of organizational inconveniences. Organizational loyalty consists of promoting and defending the organization. Organizational compliance relates to adhering to the organization’s rules, regulations, and procedures. Individual initiative refers to going above and beyond the formal requirements for job tasks. Civic virtue comprises being committed to the organization as a whole, and includes behaviors such as attending organization-wide meetings and taking part in committees. Finally, self-development consists of voluntarily engaging in behavior that will improve one’s knowledge, skills, and behavior. This thesis focused on the dimensions of helping behavior and voice, a commonly studied sub-dimension of civic virtue, which is defined as “promotive behavior that emphasizes expression of constructive challenge intended to improve rather than merely criticize” (Van Dyne & LePine, 1998, p. 109).

In the review by P. M. Podsakoff and his colleagues (2000), the authors reported that there was no evidence at the time to suggest that there were gender differences in OCB.
However, shortly thereafter Kidder and McLean Parks (2001) drew from gender role theory to suggest that some dimensions of OCB may be in-role for women, while others may be more in-role for men. Gender role theory posits that individuals conform to cultural expectations about their gender roles because of social pressures that favor this behavior (Eagly, Karau, & Makajhani, 1995). The culture in American society suggests that the role of the woman is to show concern and selflessness, both of which are traits that are consistent with altruism and courtesy OCBs (Kidder & McLean Parks, 2001). In contrast, the role of the man is to show aggression and competitiveness (Schein & Mueller, 1992), which are more consistent with sportsmanship and civic virtue OCBs (Ehrhart & Godfrey, 2003). Following this logic then, Kidder and McLean Parks (2001) proposed that we would expect that altruism and courtesy OCBs (which are similar to P. M. Podsakoff et al.’s, 2000, dimension of helping behavior) would be viewed as more in-role for women and extra-role for men, while sportsmanship and civic virtue OCBs would be viewed as more in-role for men and extra-role of women.

P. M. Podsakoff and his colleagues (2000) suggested that additional data exploring the presence of gender differences in OCB be collected before this issue is resolved conclusively. Since then, research has found some evidence in support of these differences (e.g., Allen & Rush, 2001; Ehrhart & Godfrey, 2003; Ehrhart, Godfrey, & Morales, 2005; Farrell & Finkelstein, 2007; Heilman & Chen, 2005; Kidder, 2002). And because evidence suggests that there is a relationship between OCBs and job performance ratings (e.g., Borman, White, & Dorsey, 1995; Johnson, 2001; Rotundo & Sackett, 2002; N. P. Podsakoff, Blume, Whiting, & Podsakoff, 2009; Van Dyne & LePine, 1998; Whiting, Podsakoff, & Pierce, 2008), researchers have also investigated and found evidence for the moderating role of ratee gender in the relationship between these two variables (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005). In one such study, the moderating role of ratee gender was not a significant (Allen & Rush, 2001), whereas other studies found a significant positive relationship for male but not female ratees who perform OCBs (Allen, 2006; Heilman & Chen, 2005), as well as a significant negative relationship for female but not male ratees who do not perform OCBs (Heilman & Chen, 2005).

One potential connection that has not received much attention in the literature thus far focuses not on the employee engaging (or not engaging) in OCB, but the person who is using
that information to form a performance rating of the employee. Thus, this paper went beyond the current literature relating to gender, OCB, and performance ratings, and further investigated these constructs by focusing on the person making the performance rating (herein referred to as the “rater”). Specifically, I hypothesized that the raters’ gender personal identity, gender social identity, gender identity salience, and their modern sexist beliefs would affect the way they rated the overall performance of female and male ratees exhibiting varying levels of helping behavior (which is considered more typical of women to perform) and voice (an often regarded sub-dimension of civic virtue considered more typical of men to perform; Kidder & McLean Parks, 2001). Two models come to fruition here. In one, the stronger a rater’s gender personal identity and/or modern sexist beliefs, the more likely she or he will be to make judgments with traditional gender norms in mind (Kidder & McLean Parks, 2001; Swim, Aiken, Hall, & Hunter, 1995), thus providing lower ratings for women who perform behaviors that are more typical of men than women. Alternatively, the stronger a rater’s gender social identity and/or gender identity salience, the more likely she or he will be to make judgments with the ratee’s position as part of their gender’s in-group or out-group in mind (e.g., Linville, 1982; Tajfel & Turner, 1986), thus providing lower ratings to ratees in the out-group.

In what follows, I will review the literature in these areas, and provide relevant hypotheses along the way. First, the literature exploring the relationship between gender and OCB will be discussed. Then, research that suggests that OCBs contribute significantly to the formation of performance ratings will be laid out. Next, research that has considered the moderating role of ratee gender in the relationship between OCB and performance ratings will be described, as well as subsequent research that has considered the additional moderating role of rater gender. Finally, in response to inadequate evidence for the role of rater gender alone, rater gender personal identity, rater gender social identity, rater gender identity salience, and rater modern sexism are presented as moderators in the relationships among OCB, gender, and performance ratings, in order to further explore the role of the rater.
GENDER AND ORGANIZATIONAL CITIZENSHIP BEHAVIOR

Over the past ten years, evidence supporting the presence of a relationship between gender and OCB has been found (e.g., Ehrhart & Godfrey, 2003; Farrell & Finkelstein, 2007; Heilman & Chen, 2005), contrary to previous research that found little evidence for such a relationship (P. M. Podsakoff et al., 2000). This research has even begun to go beyond investigating the simple presence of gender differences, and has started to consider gender differences in OCBs relating to gendered jobs as well (e.g., Allen & Rush, 2001; Ehrhart et al., 2005; Farrell & Finkelstein, 2007; Kidder, 2002). In this section, this literature will be reviewed, and then its contribution to the research proposal will be described.

In 2003, Ehrhart and Godfrey examined the relationship between gender and OCB schema content in a study of undergraduate students at a university on the west coast. They found that, in general, OCBs were considered more in-role for women than they were for men, suggesting that these behaviors are less expected and, thus, more noticeable when men perform them compared to when women do so. This may account for why previous research has supported the presence of a perception that women and men perform similar amounts of OCB. This finding, however, only applied to the dimensions of helping behavior, self-development, and organizational compliance, while the dimensions of sportsmanship, civic virtue, and individual initiative were considered more in-role for men. These results provided an impetus for further research in this area to help clarify the roles of OCB as it relates to gender in an organizational context.

Among other investigations, Heilman and Chen’s (2005) article explored whether certain dimensions of OCB are considered more of a requirement than a discretionary behavior for women or men. In line with the study’s hypotheses, altruistic citizenship behaviors (most similar to the helping behavior dimension here; P. M. Podsakoff et al., 2000) were rated as being more required of women than men (or, alternatively, more of an optional behavior for men than women). Additionally, the study found that raters considered civic virtue and individual initiative behaviors as more required of men than women. These ratings did not vary by the gender of the participant. These findings help provide evidence that expectations regarding performance of OCB vary according to the dimension of OCB and gender of the ratee.
In 2007, Farrell and Finkelstein conducted a study with three sets of hypotheses relating to OCB (helping behavior and civic virtue), ratee gender, stereotypical job type, and attributions made regarding impression management motives for OCB. Overall, this study found that women were generally expected to engage in more helping behavior and civic virtue than men, and that participants were more likely to associate men’s OCBs with impression management motives than the OCBs of women. These findings offer additional evidence for gender differences in the different dimensions of OCB, and also provide an explanation of how individuals view men and women who engage in OCB.

In addition to research that addresses gender differences in expectations for OCB, there has been some evidence to suggest that gender differences in OCB may vary across a variety of gender-typed jobs (e.g., Allen & Rush, 2001; Ehrhart et al., 2005; Farrell & Finkelstein, 2007; Kidder, 2002). Allen and Rush (2001) found that, compared to men, women were perceived as performing OCB more often in gender-neutral and male-typed jobs. Interestingly, Kidder (2002) found no gender differences in reports of altruism, but did find that women were significantly less likely than men to report exhibiting civic virtue behaviors. Ehrhart et al. (2005) found that, compared to child psychologists (a stereotypically female job), engineers (a stereotypically male job) were more associated with stereotypically male OCBs, including sportsmanship, civic involvement (which is similar to civic virtue; P. M. Podsakoff et al., 2000), and taking initiative (which is the same as individual initiative; P. M. Podsakoff et al., 2000), and compared to engineers, child psychologists were more associated with stereotypically female OCBs, including helping, conscientiousness, and self-development. Additionally, they found that compared to electricians (a stereotypically male job), dieticians (a stereotypically female job) were more associated with helping and self-development. Finally, Farrell and Finkelstein (2007) found that civic virtue was more expected of men in a masculine stereotyped job. After increasing the “voice” aspect of their civic virtue items, they also found that in the context of a gender neutral job position, men were more expected than women to engage in civic virtue. Although this literature provides an intriguing new direction for OCB research, this thesis controlled for the variable of gender-typed job by focusing on a job type that is considered gender neutral. This was done in favor of a more parsimonious research design, so that the focus can be kept on clarifying the precise role of the rater on OCB. However, it is suggested that future research continue to
explore this avenue so as to provide additional evidence for the role of gender-typed job in the relationship between gender and OCB.

These articles provide evidence for the position that there are varying expectations for OCBs for women and men (e.g., Ehrhart & Godfrey, 2003; Farrell & Finkelstein, 2007; Heilman & Chen, 2005). This thesis sought to expand on this growing literature, focusing on the dimensions of helping and voice behaviors because there are previous findings to support these choices (e.g., Ehrhart & Godfrey, 2003; Farrell & Finkelstein, 2007; Heilman & Chen, 2005; Kidder, 2002), as well as reasoning for why traditional gender roles and stereotypes would suggest gender differences in expectations for these particular dimensions of OCBs (Kidder & McLean Parks, 2001; Van Dyne & LePine, 1998). In particular, Van Dyne and LePine (1998) note that the helping behavior dimension includes interpersonal and cooperative behavior that is other-oriented, which is associated with an understanding of the needs of others and is illustrated by feminine stereotypes. So, expectations should be greater for women than men to engage in the helping behavior dimension of OCB. On the other hand, “voice” behavior takes on a change-orientation because it emphasizes discussing ideas and organizational issues (Van Dyne & LePine, 1998). The change orientation of voice behavior can damage relationships as opposed to nurturing them, and aligns with masculine stereotypes describing men as more assertive and task-oriented (as opposed to relationship-oriented). So, expectations should be greater for men than women to engage in voice behavior.

**Organizational Citizenship Behavior and Performance Ratings**

One reason the relationship between OCB and gender is of interest to organizational researchers relates to OCB’s link to workplace outcomes. Specifically, past research has begun to establish the presence of a relationship between OCB and performance ratings (e.g., Borman et al., 1995; Johnson, 2001; N. P. Podsakoff et al., 2009; Rotundo & Sackett, 2002; Van Dyne & LePine, 1998; Whiting et al., 2008). In addition, the moderating role of gender has been considered (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005). In this section, this literature will be reviewed, and then its contribution to this thesis will be discussed.
In 1995, Borman and colleagues examined the effects of various rater-ratee relationship and ratee-characteristic variables on job performance ratings of both supervisors and peers in a sample of first-tour United States Army soldiers. Among other things, this study found that contextual performance (also referred to as citizenship performance; Coleman & Borman, 2000; Organ, 1997) is weighted just as highly as task performance by supervisors when making overall performance ratings. This study provided foundational evidence for the proposition that OCB is given substantial weight in employee performance ratings.

Van Dyne and LePine (1998) considered how task performance, helping behavior, and voice behavior relate to overall ratings of job performance in a longitudinal field study. Specifically, the authors collected self, peer, and supervisor ratings of task performance, helping behavior, and voice behavior (collected at times 1 and 2), and correlated those ratings with supervisor ratings of job performance (collected at time 2). The study found that voice behaviors positively related to performance ratings for all three rating groups, and that both helping and voice behaviors explained unique variance in supervisors’ performance ratings. However, Van Dyne and LePine (1998) noted that it could not be clear whether each individual OCB dimension or a combination of the two contributed to higher overall performance ratings because the correlations between helping and voice were relatively high (the average of the correlations was 0.74, with a range of 0.63 to 0.81). This thesis sought to provide further evidence of the role of helping and voice behaviors in the formation of overall performance ratings using different methods from this particular study.

A study by Johnson (2001) considered the relative contribution of specific dimensions of contextual performance (including organizational citizenship performance, interpersonal citizenship performance, and job-task conscientiousness) on overall performance ratings by considering the way raters weigh and combine the various dimensions of performance. Overall, the study found that each contextual performance dimension made a unique contribution to overall performance ratings, and that the contextual performance dimensions were frequently even perceived as more important than specific components of task proficiency. This study provides further evidence of the significant role that OCB plays in the formation of overall performance ratings.
Subsequent research by Rotundo and Sackett (2002) focused on the relative importance of three types of performance—task, citizenship, and counterproductive—on overall job performance ratings using a policy-capturing approach. This study found that the three types of performance played a significant role in the formation of overall performance ratings, and that, on average, task performance and counterproductive performance had more weight on individuals’ overall performance ratings. Furthermore, this study found partial support for the hypothesis that task performance moderates the relationship between citizenship performance and overall performance ratings, in that citizenship performance would be weighted more heavily for high task performers compared to low task performers. However, the authors expressed a wariness to give this finding much weight, as most relationships tend to attain a conventional level of significance when there is a large sample size (because the authors pooled the data across raters when running the analysis for this hypothesis, there were a total of 16,128 data points). Nevertheless, this study provides further evidence for the significant role of OCB in overall performance ratings.

In an experimental design manipulating hypothetical critical incidents of secretaries, Whiting and colleagues (2008) investigated the effects of task performance, helping behavior, voice behavior, and organizational loyalty on performance evaluations. The study found that each of these types of behaviors had a significant, independent effect on performance evaluations. Additionally, they found that when task performance and helping behavior were both low, raters gave fewer rewards for performing high levels of voice behavior. Taking the suggestion of this article, the current thesis sought to expand the literature focusing on voice behavior as an important dimension of OCB aimed at the organization, as well as contribute to the growing the body of literature that explores the causal effect of helping behavior on performance ratings.

In 2009, N. P. Podsakoff and colleagues published a meta-analysis that summarized research that has considered the relationship between OCB and various outcomes. They found that ratings of different types of OCBs, as well as ratings of task performance, are distinguishable from each other. Among other individual-level outcomes, this meta-analysis found a significant relationship between OCB and job performance ratings, and showed that the strength of this relationship is somewhat greater than that between task performance and job performance ratings. This meta-analysis contributed to this thesis because it adds to
evidence for the proposition that OCBs are distinct behaviors that may have differential effects on performance ratings.

These studies verify the presence of evidence for a relationship between OCB and overall performance ratings (Borman et al., 1995; Johnson, 2001; N. P. Podsakoff et al., 2009; Rotundo & Sackett, 2002; Van Dyne & LePine, 1998; Whiting et al., 2008). Taking this research into account, this thesis hypothesized the following:

- **Hypothesis 1a:** Helping behavior will positively predict overall performance ratings.
- **Hypothesis 1b:** Voice behavior will positively predict overall performance ratings.

Taking advice given by N. P. Podsakoff et al. (2009), this thesis further explored why OCBs have the effect that they do on outcomes such as performance ratings, specifically with regard to gender. This issue is further discussed in the following sections.

Extending research that explores the relationship between OCB and performance ratings, there has been research that considers the moderating role of ratee gender on that relationship (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005). These findings have varied from finding no significant moderating relationship, to finding a notably more positive relationship for men compared to women. As a result, this relationship needs further exploration so as to clarify the role of ratee gender in the relationship between OCB and performance ratings. In this section, articles that have focused on this area, as well as their contribution to this thesis, will be discussed.

Besides making propositions regarding types of OCBs within stereotypically feminine and masculine jobs, Kidder and McLean Parks (2001) also suggested that individuals are less likely to be rewarded (in their performance ratings) for gender-congruent OCBs compared to individuals of the opposite sex who exhibit the same OCBs. Specifically, they proposed that women would be less likely to be rewarded for feminine OCBs compared to men, and men would be less likely to be rewarded for masculine OCBs compared to women. The authors argued that this is because gender-congruent OCBs are considered in-role, so individuals engaging in gender-congruent OCBs would not be recognized as going above and beyond their job requirements, and might even get penalized if they do not perform these behaviors. This proposition and its rationalization provide an impetus for considering the role of ratee gender on the relationship between OCB and performance ratings.
In addition to considering the moderating role of gender-typed job in the relationship between gender and perceived frequency of OCB (as discussed in a previous section of this paper) in their 2001 article, Allen and Rush examined whether ratee gender serves as a moderator in the relationship between OCB and overall performance rating or reward recommendations. Surprisingly, the effect of this moderator was not found to be significant. The authors suggest that this may because of a base-rate issue, wherein women are expected to perform more OCB, so its presence attracts less attention because it is considered more commonplace. Because of the empirically shown importance of OCB in overall performance ratings, the authors suggest that this proposed relationship among OCB, ratee gender, and performance ratings be further investigated in future research.

Beyond investigating whether certain dimensions of OCB are considered more of a requirement than a discretionary behavior for women or men, Heilman and Chen (2005) also considered the effects of prescriptive stereotypes about gender roles on reactions to altruistic citizenship behaviors of women and men. They found that the same act of altruistic citizenship behavior prompted different ratings and recommendations for women than for men. Specifically, this study found that performing altruistic citizenship behavior improved men’s (but not women’s) ratings and recommendations, while not performing these behaviors worsened women’s (but not men’s) ratings and recommendations. This provides further evidence for the hypothesis that there are gender differences in expectations for performance of OCB, as well as for the proposition that these expectations influence subsequent ratings and recommendations for employees.

In 2006, Allen considered the moderating role of ratee gender in the relationship between OCB and two organizational rewards, salary and promotion. Allen’s (2006) perspective on OCB differed from the other studies discussed in this paper in that it focused on OCBI and OCBO—OCB directed toward individuals and the organization, respectively. Allen (2006) characterized OCBI (which includes behaviors such as helping others with heavy workloads and sharing information with coworkers) as feminine OCB, and OCBO (which includes behaviors such as having above-average work attendance and offering innovative ideas to improve unit quality) as masculine OCB. Consistent with her hypotheses, she found that the relationship between both OCBI and OCBO and promotions was significant for men but not for women. As Allen (2006) pointed out, these findings are
partially inconsistent with Kidder and McLean Parks’ (2001) proposition regarding OCB and performance ratings because they suggest that OCBO has a weaker relationship with promotions for women compared to men. Nonetheless, this article provides evidence for the moderating role of ratee gender in the relationship between OCB and performance evaluations, and shows that more research needs to be done in this area to help clarify just how this relationship works.

These articles show that there is some relationship among ratee gender, OCB, and performance evaluations, although the detail regarding the role of each variable has not been consistent. This thesis sought to examine this relationship considering the evidence from the articles discussed above (Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005), in hopes that this relationship would be clarified by the proposed model. Specifically, it was argued that women will not be rewarded as greatly as men in their performance ratings when performing OCBs that are considered congruent with their gender (i.e., helping behavior) because those OCBs are considered more in-role for them.

- **Hypothesis 2a**: Men, but not women, will will receive higher overall performance ratings for performing more helping behavior.

Contrary to the proposal made by Kidder and McLean Parks (2001), the research in this area does not suggest that this relationship holds true for men who exhibit gender-congruent OCBs (e.g., Allen, 2006). Instead, it suggests that men are rewarded more than women for performing masculine OCBs, but women aren’t rewarded more than men for performing feminine OCBs (e.g., Allen, 2006). This may relate to evidence that women are perceived as warm but incompetent (Fiske, Cuddy, Glick, & Xu, 2002), whereas men, because of their higher status, are perceived as high in competence (e.g., Fiske et al., 2002; Ridgeway, 1991), which may promote distinct types of behaviors from raters. Specifically, raters may engage in patronizing behavior, whereby women receive more condescending flattery (e.g., pats on the back, flattery) and less valued resources (e.g., promotions, raises) than men (Biernat & Vescio, 2002). So, it is plausible that raters perceive voice behaviors performed by women as not as useful because women’s perceived incompetence, as well as generally unbecoming of their stereotypically warm nature. As a result, this behavior is insufficient grounds for high performance ratings (which are often tied to valued resources), while they perceive voice behaviors performed by men as a reflection of competence that is
deserving of valued organizational rewards. Within the context of this thesis, this would suggest that raters perceive the performance of voice behaviors by men as a positive contribution that deserves greater rewards than a woman would receive for performing the same voice behavior.

- **Hypothesis 2b**: Men, but not women, will receive higher overall performance ratings for performing more voice behavior.

**PREDICTORS OF GENDER STEREOTYPING**

Another approach to examining the relationship between OCB and performance ratings focuses on the rater as opposed to the ratee. Some research has addressed this by testing the moderating effect of rater gender, but with mixed results (e.g., Ehrhart & Godfrey, 2003; Ehrhart et al., 2005; Heilman & Chen, 2005). The next step in clarifying the inconsistencies in rater gender effects is to assess individuals’ identification with their gender, as opposed to assuming it (e.g., Mannix & Neale, 2005; Riordan, 2000). So, as an alternative to considering the role of rater gender, this thesis proposed to instead focus on how individuals see themselves in terms of their gender, how they see others in their gender group, and how they see others who are not in their gender group. In this section, previous research that has investigated the role of rater gender will be discussed, and then rater gender personal identity, rater gender social identity, rater gender identity salience, and rater modern sexism will be proposed as additional moderators, in hopes that this will clarify the role of the rater in the relationship between OCB, ratee gender, and performance ratings.

**Rater Gender**

In their 2003 study with undergraduate students, Ehrhart and Godfrey found that female raters perceived that women are more associated with dimensions of OCB than men, while male raters did not associate men with any particular type of behavior significantly differently than female raters. The authors also found a slight in-group bias, in that raters perceived their own gender as being more associated with OCB and task-related dimensions, but warned that these findings should not be given much consideration without further research evidence, as most of these differences were not statistically significant.

Another study that considered the moderating effect of rater gender in the relationship between OCB and performance ratings is that of Ehrhart and colleagues (2005). In a sample
of undergraduate students, they found that male raters associated engineers (a male-stereotyped job) significantly more than child psychologists (a female-stereotyped job) with civic involvement and taking initiative, while female raters associated neither child psychologists nor engineers significantly more with the above two dimensions and sportsmanship. They also found that both female and male raters associated child psychologists and dieticians (another female-stereotyped job) significantly more than engineers and electricians (another male-stereotyped job) with helping and self-development (female raters also associated child psychologists and dieticians significantly more with conscientiousness). In other words, to a certain extent, male raters associated male-typed jobs more with stereotypically-male OCBs, while both female and male raters associated female-typed jobs more with stereotypically-female OCBs. This study provides additional, although only partial, support for rater gender differences.

Heilman and Chen’s (2005) research with both undergraduate and MBA students showed rater gender findings that differed greatly from the other two studies discussed in this section. Specifically, their study found that rater gender did not have any significant main effects, nor did it interact with altruistic citizenship behavior in participants’ formation of overall performance and reward recommendation ratings. There was a significant main effect between rater gender and interpersonal civility, in that female raters evaluated ratees significantly more favorably on interpersonal civility compared to male raters, but there were not any significant interactions with this variable. Thus, based on the limited literature that exists on this topic to date, there are mixed findings regarding the role of rater gender in the relationship between OCB, ratee gender, and performance ratings.

Rater Gender Personal Identity

In response to previous research that has found mixed results regarding rater gender, it was proposed that rater gender personal identity be investigated as a moderator, in order to clarify the role of the rater in the relationship between OCB, ratee gender, and performance ratings. This proposal was adapted from the logic and proposition given by Kidder and McLean Parks in their 2001 article. In this section, gender personal identity will be defined, theoretical background tying this construct to OCB, ratee gender, and performance ratings will be given, and formal hypotheses will be described.
Generally speaking, gender identity refers to the extent to which an individual identifies with her or his gender (Randel & Jaussi, 2008). One type of gender identity is gender personal identity, which is the extent to which an individual places importance on the personal characteristics of sex as a part of her or his self-concept (e.g., Rosenberg, 1979). This part of the self-concept is oriented toward the individual, and so does not depend on the presence of others to reflect the personal importance one places on gender (Brewer & Gardner, 1996; Tajfel & Turner, 1986). Individuals with a strong gender personal identity would be expected to think about gender often, and to behave in ways that reflect the importance they place on gender (Randel & Jaussi, 2008). Additionally, when an individual’s gender is important to them (as would be expected of a person who has a strong gender personal identity), she or he will consequently be more sensitive to that particular aspect of other people (Kidder & McLean Parks, 2001).

This thesis followed Kidder and McLean Parks’ (2001) logic that status characteristics theory helps explain how an individual’s gender personal identity relates to her or his judgments of others based on their gender. As proposed by Ridgeway (1991), status characteristics theory suggests that men are seen as more competent in Western societies, and thus have higher social status. This theory posits that females who perform masculine behaviors may be perceived by males as threatening their male self-identity, leading them to have a negative perception of these females. Evidence suggests that females have also internalized this status difference, and have been shown to hold negative perceptions of females performing masculine behaviors as well (e.g., Eagly, Makhijani, & Klonsky, 1992; Heilman, Block, Martell, & Simon, 1989; Lewis & Fagenson-Eland, 1998; Rojahn & Willemsen, 1994).

Following the logic of status characteristics theory, it was proposed that both men and women who have a strong gender personal identity will be more likely to hold negative perceptions of women who perform OCBs that are considered gender-incongruent (Kidder & McLean Parks, 2001). However, this hypothesis focuses only on OCBs that are considered gender-incongruent for women (e.g., voice behavior), as evidence exists that suggests that men (but not women) are rewarded for performing behaviors that are outside the norms for their group. For example, Heilman and Chen (2005) found that men, but not women, were rewarded for performing altruistic citizenship behavior, providing evidence that men are not
punished for performing gender-incongruent OCBs. This may occur because having higher status in a group (however that status be obtained), allows members of that group more freedom from maintaining norms without punishment (Hollander, 1958, 1985). This is referred to as idiosyncrasy credits, where high-status members are allowed to deviate more from norms with less punishment than low-status members. In light of this evidence, the hypothesis related to gender personal identity focused only on the negative effect that performing gender-incongruent OCBs may have for women, not for men (as this effect was not predicted to be negative, see Hypothesis 2b).

- **Hypothesis 3:** Women who are rated by a rater with lower gender personal identity will receive higher performance ratings when they perform more voice behavior, whereas women who are rated by a rater with higher gender personal identity will receive lower performance ratings when they perform more voice behavior. Men will receive higher ratings when they perform more voice behavior across all levels of rater gender personal identity.

Status characteristics theory does not directly discuss what occurs when individuals perform behaviors that are consistent with the norms and stereotypes prescribed to their gender, but following the evidence of previous research in the field (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005) as well as the theoretical background on the perceived levels of warmth and competence of women and men (Fiske et al., 2002) and on patronizing behavior (Biernat & Vescio, 2002), it was expected that, across levels of rater gender identity, raters would give men higher performance ratings than women when they perform helping behavior (see Hypothesis 2a).

**Rater Gender Social Identity**

Another way of assessing the extent to which an individual identifies with her or his gender is to focus on their gender social identity. Instead of focusing on the self, as gender personal identity does, this construct focuses on identifying oneself as part of a group (Brewer & Gardner, 1996). This construct may help provide clarification on the relationship between OCB, ratee gender, and performance ratings by focusing on how individuals perceive and judge others in relation to their placing as part of their gender group or not part of their gender group (e.g., Tajfel & Turner, 1986). In this section, gender social identity will be defined, its theoretical background will be described, and formal hypotheses regarding its role in the current study will be given.
Gender social identity focuses on how individuals identify with others of the same gender in a group (Brewer & Gardner, 1996). Whereas gender personal identity focuses on the individual and does not depend on the presence of others, gender social identity focuses on other people who are perceived to be part of the individual’s self-concept (Brewer & Gardner, 1996; Tajfel & Turner, 1986). Individuals with a strong gender social identity would be expected to be concerned with making those of the same gender in the workgroup look good, and are motivated to protect those group members, as they see similar others’ fates as their own (Randel & Jaussi, 2008).

Social identity theory helps explain how individuals’ gender social identity affects their judgments of others. This theory posits that individuals place themselves into what they perceive to be meaningful social categories (e.g., sex, race, class, nationality), and that this placement forms the way that they interact with those who are in their own social categories (e.g., in-group members) and those who are in other social categories (e.g., out-group members; Tajfel & Turner, 1986). According to social identity research (e.g., Tajfel & Turner, 1986), individuals with strong gender social identities categorize in-group members differently than they do out-group members, exhibiting favoritism toward the in-group. Related to the idea of differential categorization of one’s in-group and out-group, there is evidence that individuals hold a more complex representation of their in-group compared to their out-group, indicating that less complex representations of out-group members results in more homogenous evaluations (Linville, 1982). Therefore, raters are expected to evaluate out-group members more homogenous than in-group members (Linville, 1982).

The hypotheses regarding gender social identity are distinct from the hypotheses regarding gender personal identity in that they focus on more on whether or not raters associate ratees as a part of their gender group over whether or not OCBs are judged as congruent or incongruent with gender norms. Specifically, following the logic of social identity theory, this thesis hypothesized that raters with stronger gender social identity would exhibit in-group favoritism when providing performance ratings of others’ OCBs compared to raters with weaker gender social identity. On the other hand, when gender social identity is low, we would expect similar a similar relationship between OCB and performance ratings for men and women. More formally stated:
**Hypothesis 4a:** Female raters with lower gender social identity will give both women and men higher performance ratings when they perform more OCB. On the other hand, female raters with higher gender social identity will give women higher performance ratings when they perform more OCB, but will give men generally lower performance ratings regardless of whether or not they perform OCB.

**Hypothesis 4b:** Male raters with lower gender social identity will give both women and men higher performance ratings when they perform more OCB. On the other hand, male raters with higher gender social identity will give men higher performance ratings when they perform more OCB, but will give women generally lower performance ratings regardless of whether or not they perform OCB.

**Rater Gender Identity Salience**

In addition to proposing rater gender personal and social identities as moderators, this thesis also investigated the role of rater gender identity salience in the relationship among OCB, ratee gender, and performance ratings. This was done to further investigate the potential role of gender identity, as it is inaccurate to assume that raters identify with their gender (e.g., Mannix & Neale, 2005; Riordan, 2000). In this section, gender identity salience will be defined, relevant previous literature will be described, and formal hypotheses will be presented.

Identity salience is defined as the extent to which a demographic category (in this case, gender) is used to describe an individual’s work group members (Randel, 2002). This construct is conceptualized at an individual level, and is based on the assumption that we categorize others using demographic information because of our cognitive limitations and desire to make sense of our social environment (Fiske & Neuberg, 1990; Weick, 1979). Self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) may help us explain how an individual’s gender identity salience can affect her or his perceptions of others in the workplace. According to this theory, individuals categorize themselves in order to make sense of the world. The way in which we categorize ourselves changes depending on the social environment and which identity is activated in the situation. In a particular situation, we also use demographic-based categories to classify people who we encounter into an in-group and an out-group based on whether or not they align with our position in that category (Tajfel, 1982; Turner, 1987). Along with making these in-group and out-group distinctions, individuals also attempt to increase their self-esteem by holding positive
opinions of those in their in-group, while holding negative opinions of those in their out-group (e.g., Tajfel & Turner, 1986).

Although identity salience has been proposed through theory to affect the behavior of individuals (e.g., Kramer, 1993; Stryker, 1968; Shamir, 1990), empirical research focused on outcomes of identity salience is not as well-established (Randel, 2002). This thesis, therefore, seeks to expand the research on this area by testing the role of gender identity salience on the relationship among OCB, ratee gender, and performance ratings. Gender identity salience is expected to explain the relationship among these variables more precisely than gender because it accounts for the extent to which gender among group members is noticed by an individual rather than merely being present (Randel, 2002). The hypotheses presented in this section took into account the position of self-categorization theory that individuals categorize themselves and others into demographic-based categories which they then use to form an in-group and an out-group, as well as the position that individuals hold a higher opinion of their in-group compared to their out-group. Specifically, it was hypothesized that individuals whose gender identity is salient reward in-group members more than out-group members for performing OCB, whereas individuals whose gender identity is not salient will reward both women and men for performing OCB. More formally stated:

- **Hypothesis 5a:** Female raters whose gender identity salience is lower will give both women and men higher performance ratings when they perform more OCB. On the other hand, female raters whose gender identity salience is higher will give women higher performance ratings when they perform more OCB, but will give men generally lower performance ratings regardless of whether or not they perform OCB.
- **Hypothesis 5b:** Male raters whose gender identity salience is lower will give both women and men higher performance ratings when they perform more OCB. On the other hand, male raters whose gender identity salience is higher will give men higher performance ratings when they perform more OCB, but will give women generally lower performance ratings regardless of whether or not they perform OCB.

**Rater Modern Sexism**

In addition to proposing rater gender personal identity, gender social identity, and gender identity salience as moderators, this thesis sought to investigate the moderating role of rater modern sexism in the relationship among OCB, ratee gender, and performance ratings. Since research suggests that both men and women can exhibit modern sexism (though men do so more often than women; Swim et al., 1995; Watkins et al., 2006), this construct may
provide clarification for the role of the rater. In this section, modern sexism will be defined, previous research that provides logic for the inclusion of this construct as a moderator will be described, and formal hypotheses will be given.

According to Swim et al. (1995), modern sexism is characterized by a denial that gender discrimination continues to be an issue, an antagonism toward women who make political and economic demands, as well as feelings of resentment toward policies that help women. Interestingly, individuals who are high on modern sexism do not regard themselves (Swim, Mallett, & Stangor, 2004) or others (e.g., Swim, Mallett, Russo-Devosa, & Stangor, 2005) as sexist for endorsing these characteristics. Rather, they attribute inequality in certain areas as being caused by individual shortfalls rather than broader discrimination against the group. In theory, modern sexism’s characteristics signify opposition to efforts that address sexism and a desire that the status quo be maintained. Thus, the construct was designed by Swim et al. (2005) to covertly assess a preference for inequality for women, effectively assessing sexist beliefs in a more subtle manner.

Modern sexism more covertly predicts sexism compared to old-fashioned sexism (which are more traditional sexist beliefs about women, characterized by support for traditional gender roles, disparate treatment of women and men, and stereotypes regarding the lesser competence of women; Swim et al., 1995; Swim et al., 2005), which is reflected in the behaviors modern sexists perform. Modern sexists are affected by concerns for social desirability, and so will not show their prejudices outright—instead, revealing them only in “safe” circumstances where their behavior could be attributed to something other than sexism (Swim & Cohen, 1997). Modern and old-fashioned sexism are distinct constructs but still are correlated, so individuals who endorse modern sexist beliefs are also more likely to hold more traditional beliefs about women (Swim et al., 1995).

As interest in the construct of modern sexism has grown, research has begun to focus on the potential impact of being a modern sexist in the workplace. Watkins and colleagues (2006) investigated the relationship between modern sexism and workplace outcomes, and found that individuals high in modern sexism reported receiving 31% more promotions than individuals low in modern sexism. This finding alone may have important implications, not only for those low in modern sexism who receive fewer promotions, but also for the employees with whom individuals high in modern sexism work. Those modern sexists who
are receiving promotions may be entering positions where they manage others. Within the responsibilities of managing others, there is the act of providing performance ratings for subordinates that likely affect their pay and promotion rate. Considering this, along with evidence that suggests that modern sexism is related to differential perceptions and treatment of women and men in the workplace (Swim et al., 1995) and that modern sexists perceive men as more competent than women, especially in terms of workplace competence (e.g., Beaton, Tougas, & Joly, 1996; Tougas, Brown, Beaton, & St-Pierre, 1999), it is plausible that modern sexists may differentially rate the behavior of women and men who they manage.

Considering evidence that individuals high in modern sexism hold more traditional beliefs about women (e.g., Swim et al., 1995), one may posit that they will react unfavorably when a woman exhibits behavior that goes against traditional gender norms for women (e.g., civic virtue OCB). And if the individual high in modern sexism is in a higher position in the organization (which is plausible considering evidence that individuals high in modern sexism receive more promotions than individuals low in modern sexism; Watkins et al., 2006), that unfavorable reaction may then be reflected in overall performance ratings, since an overall performance rating may be a “safer” circumstance in which behavior can be attributed to something other than sexism (Swim & Cohen, 1997). Because modern sexists perceive and treat men differently than women (Swim et al., 1995), they may not react so unfavorably when a man exhibits behavior that goes against traditional gender norms for men. In sum, this thesis offers the following hypotheses:

- **Hypothesis 6**: Women who are rated by a rater with lower modern sexism will receive higher performance ratings when they perform more voice behavior, whereas women who are rated by a rater with higher modern sexism will receive lower performance ratings when they perform more voice behavior. Men will receive higher ratings when they perform more voice behaviors across all levels of rater modern sexism.

As was done with rater gender personal identity, this thesis followed the evidence of previous research in the field (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005) and the theoretical background on the perceived levels of warmth and competence of women and men (Fiske et al., 2002) and on patronizing behavior (Biernat & Vescio, 2002), in generating hypotheses regarding circumstances when individuals perform behaviors that are congruent with their gender’s traditional norms (e.g., helping behavior for women and voice behaviors for men). In general, across levels of modern sexism, it was expected that
raters would give men higher performance ratings than women when they perform helping behavior (see Hypothesis 2a) as well as when they perform voice (see Hypothesis 2b).
CHAPTER 2

METHOD

PARTICIPANTS

A total of 234 San Diego State University students working in a variety of industries were recruited to participate in a two-part study (which will be referred to as Study 1 and Study 2). Students were recruited to volunteer for participation in the confidential study through the Psychology department web-based participant pool in exchange for course credit, as well as through two upper-division Industrial/Organizational Psychology courses for extra credit. The sample was 73.1% female and 26.9% male, and the average age of participants was 21.52 (SD = 4.02). The racial/ethnic composition was fairly consistent with the location of the university; 39.3% of participants were Caucasian, 12.8% were Asian, 35% were Hispanic or Latino, 4.7% were Black or African-American, and the remaining 8.2% were of other ethnicities. In terms of employment, 88% of participants were currently employed, and 95% of participants had some job experience. Participants had an average of one year and nine months of work experience at their current job (SD=2.13 years), and worked an average of 20.78 hours per week (SD=10.24). In terms of type of position, 80.9% of participants were in non-managerial positions, 16.6% were in managerial positions, and 1.3% were in executive positions. It should be noted that because of an administrative error, a smaller portion of the sample participated in the entirety of the second study, so data from only 175 of the original participants (70.9% women and 29.1% men) were used for the second study. All participants were given course credit for participating in the studies.

A power analysis was conducted using G*Power software (Faul, Erdfelder, Buchner, & Lang, 2009) to identify the target sample size for this study. Although the conventional guidelines for effect sizes suggests that 0.02 is small, 0.15 is medium, and 0.35 is large (Cohen, 1988), Aguinis, Beaty, Boik, and Pierce (2005) found evidence that the average effect size in studies examining moderation is only 0.009. They suggest that the guidelines of 0.005, 0.01, and 0.025 for small, medium, and large effect sizes, respectively, might be more practical for tests of moderation. Keeping this in mind, in order for the study to obtain a
power of .80, the power analysis determined that the target sample size needed was between 159 ($f^2 = 0.05$) and 316 ($f^2 = .025$) participants.

**STUDY 1 PROCEDURE**

Participants were recruited to take part in the two-part study through the SDSU Psychology department participant pool. Before beginning the survey, the informed consent script for both of the studies was read by each participant, and their written consent to participate in the studies was obtained. Consistent with a policy capturing design, consenting participants were then be instructed to read 16 scenarios describing hypothetical employees and respond to four questions regarding the performance of each. The order of the scenarios was randomized to control for possible order effects. Participants then responded to survey items measuring their gender personal identity, gender social identity, gender identity salience, modern sexism, and OCB expectations by gender. Additionally, as per the advice of Huang, Curran, Keeney, Poposki, and DeShon (2011) insufficient effort items were placed at five points throughout the survey, each asking the participant to choose a specific response (e.g., “strongly disagree”), so that data from participants who did not pay adequate attention while taking part in the studies (i.e., 3 or more incorrect responses) could be discarded. As a result of this procedure, 18 of the participants were removed from the analyses. Participants had the opportunity to ask questions at any point during this process. The researcher’s contact information was also provided in case participants had questions at any point after taking the survey.

**STUDY 1 MATERIALS**

The measures that were included in the survey are discussed in detail below. A complete list of the survey items can be found in Appendices A through L.

**Policy Capturing Scenarios**

The 16 scenarios describing hypothetical employees are comprised of a fully-crossed, 2 (woman vs. man) X 2 (low vs. high task performance) X 2 (low vs. high helping behavior) X 2 (low vs. high voice) within-subject policy-capturing design (see Appendix A). Policy-capturing involves presenting raters with scenarios in which independent variables are manipulated, obtaining raters’ opinions relating to the dependent variable, and using
multilevel analyses to calculate how raters combine and weight the information given to them in the scenarios to arrive at their opinions (see Rotundo & Sackett, 2002, for an example of how raters weight different behaviors when forming performance ratings).

In order to control for the effect of gender-typed job, the scenarios described post office employees, as this is considered a gender-neutral job (49.6% of postal service clerks are female; U.S. Bureau of Labor, 2010). The gender of the target employee was manipulated using gender-typed names (e.g., Christopher vs. Christina) and gendered pronouns. The gender-typed names were chosen from among the 200 most popular names for babies born in 1990 – 1999 (U.S. Social Security Administration, 2012). Task performance was manipulated by stating whether the target had been rated “above average” or “below average” in their recent performance reviews. Although task performance is not a main variable in this proposal, it is included in the scenarios so that the relative weight given to OCBs (which are extra-role) in the presence of task performance (which is in-role) can be considered. Helping behavior was manipulated by stating whether the target “often” or “rarely” helped coworkers with their duties when they fell behind. Voice was manipulated by stating whether the target “often” or “rarely” speaks up with ideas for new projects or changes in procedures. The scenarios were adapted from a previous study conducted by Soltani, Ali, Ehrhart, Schultz, and Ensey (2013), with exception of the voice manipulation, which was adapted from Van Dyne and LePine (1998). Following the advice of Aiman-Smith, Scullen, and Barr (2002), three of the scenarios (each varying in terms of level of the manipulated conditions) were duplicated for the purpose of examining test-retest reliability. Bivariate correlations were conducted for the Scenarios 3, 10, and 14 for this purpose, and all were significant $r=0.52, p<0.001$, $r=0.35, p<0.001$, and $r=.40, p<0.001$.

**Performance Ratings**

After reading each scenario, participants were prompted to respond to four items relating to the extent to which they would consider each individual scenario an example of excellent overall performance, each on a 5-point Likert scale (ranging from 1 = *to little or no extent* to 5 = *to a great extent*; see Appendix A). The items asked the following: “To what extent is this person an example of an ideal employee?”; “To what extent would you consider this description to be an example of excellent overall performance?”; “To what extent would
you recommend rewarding this employee with a promotion?”; and “To what extent would you recommend rewarding this employee with an increase in salary?” These items were created as part of a previous study conducted by Soltani et al. (2013), in which a reliability of .92 was reported (this reliability was the average of the individual reliabilities calculated for each scenario). Like the study in which this scale was developed, the alpha reliability for the scale in this study was calculated separately for each scenario; the average alpha was .90 and the range was .87-.92.

**Gender Personal Identity**

The importance of gender to an individual’s personal identity was measured using Luhtanen and Crocker’s (1992) four-item measure (see Appendix B). Participants reported the extent to which they agree or disagree with each item using a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items stated “In general, my gender is an important part of my self-image,” “My gender is unimportant to my sense of what kind of a person I am,” “Overall, my gender has little to do with how I feel about myself,” and “My gender is an important reflection of who I am.” The alpha for this measure in the study was 0.75.

**Gender Social Identity**

Randel and Jaussi’s (2003) measure was used to measure gender social identity in this study (see Appendix C). The instructions for this three-item measure asked participants to report the extent to which they agree or disagree with each item using a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items stated “When others of my gender are recognized for their accomplishments, I feel like I have accomplished something too,” “I share in the successes of others of my gender,” and “I think about my gender more when someone of my gender is commended on his or her work.” The alpha for this measure in the study was 0.88.

**Gender Identity Salience**

Randel’s (2002) measure of gender identity salience was adapted for this study (Appendix D). This three-item measure asked participants to rate the extent to which they agree or disagree with each item using a 5-point Likert scale ranging from 1 (*strongly
disagree) to 5 (strongly agree). The items stated “When people ask me about who I work with, I initially think of describing my coworkers in terms of gender composition (e.g., two women and three men),” “It is not intentional, but when I think of my coworkers, what comes to mind initially are the names of individuals of the same gender as me,” and “Even though I don’t mean to, I think of gender as the most prominent characteristic of my coworkers.” The alpha for this measure in the study was 0.66.

**Modern Sexism**

Participants’ levels of modern sexism was measured using Swim et al.’s (1995) Modern Sexism Scale (see Appendix E). Using this eight-item scale, participants were instructed to rate how much they agree with each item according to their attitudes, beliefs, and experiences on a 5-point Likert scale (ranging from 1 = strongly disagree to 5 = strongly agree). Example items include “Discrimination against women is no longer a problem in the United States” and “Society has reached a point where women and men have equal opportunities for achievement.” It should be noted that this scale is coded so that lower scores on the scale suggest higher levels of modern sexism, and higher scores on the scale suggest lower levels of modern sexism. The alpha for this measure in the study was 0.80.

**OCB Expectations by Gender**

This scale was included in the study to inform exploratory analyses regarding whether there are differences between participants’ explicitly stated OCB expectations by gender, and those expectations which are implied by the performance ratings they give. Two separate measures of OCB expectations by gender were included in this study. In the first measure, which is adapted from Soltani et al. (2013), participants were instructed to read descriptions of what helping behavior and voice entails, and report the extent to which they believe each behavior is more expected of women or men in their workplace on a 7-point scale ranging from 1 = much more expected of women to 7 = much more expected of men (see Appendix F). The second measure used the same instructions and response scale, but included multiple items for each type of behavior instead of a longer, single description (see Appendix G). The helping behavior items were adapted from P. M. Podsakoff, Mackenzie, Moorman, and Fetter (1990), and the voice behavior items were adapted from Van Dyne and LePine (1998).
The alpha for the helping behavior measure was 0.91, and the alpha for the voice behavior measure was 0.85. These measures appeared at the end of the survey to avoid priming participants to rate the hypothetical employees in the scenarios in a way that is aligned with their explicitly stated perceptions of OCB expectations by gender.

**STUDY 1 STATISTICAL ANALYSES**

The proposed hypotheses were analyzed using hierarchical linear modeling (HLM) with the Mplus software package (Muthén & Muthén, 2010). HLM allows for the examination of within- and between-group variance and has been shown to be an effective analytical technique when dealing with policy-capturing data (Kristof-Brown, Jansen, & Colbert, 2002; Rotundo & Sackett, 2002). For the HLM analyses, ratee gender, task performance, helping behavior, voice, and the cross products of ratee gender and each of the OCB dimensions were variables at level 1, and rater gender, rater gender personal identity, rater gender social identity, rater gender identity salience, and rater modern sexism were variables at level 2. For Hypotheses 1, 2a and 2b, a random coefficients model in HLM was performed by regressing performance ratings on ratee gender, task performance, helping behavior, voice, and the cross products of ratee gender and each of the OCB dimensions, with rater gender included as a control. In order to analyze Hypotheses 3 and 6, the interaction terms for these analyses at level 1 (genderXhelping, and genderXvoice) were then be regressed as outcome variables on the level 2 predictor variables of rater gender personal identity and rater modern sexism, controlling for rater gender, in two separate models. Similarly, in order to analyze Hypotheses 4a, 4b, 5a, and 5b, the interaction terms for these analyses at level 1 (genderXhelping, and genderXvoice) will be regressed as outcome variables on the level 2 predictor variable of rater gender social identity (for Hypotheses 4a and 4b) and rater gender identity salience (for Hypotheses 5a and 5b), controlling for rater gender, but with the dataset split by rater gender before running analyses.

**STUDY 2 PROCEDURE**

After completing the first study, the same participants who participated in Study 1 were prompted to begin a second, separate study. Participants began this study by rating the overall performance of one male and one female coworker. Then, they responded to items
about the extent to which their two peers engage in different types of OCBs in their workplace. Finally, they were given flyers to give to those same two peers inviting them to participate in the second portion of the study in which they were asked to rate their own OCBs. Peers were chosen to participate in the second portion of the survey because past research suggests that peers are more likely than supervisors to see the performance of OCB (Lievens, Conway, & De Corte, 2008), and so are likely to provide more accurate accounts of the behaviors. In order to provide incentive for participants to complete the first portion of the study, and to encourage peers to complete the second portion, participants and their peers who completed the survey were individually added to a drawing for a $100 gift card. Participants were assured that they would receive course credit for participating in the study regardless of whether or not their peers responded to the second portion of the survey.

General demographic information was collected at the end of the survey for use in both Study 1 and 2. Participants were asked to report their gender, race/ethnicity, age, class status, major, overall work tenure, tenure at their current or most recent job, and how many hours a week they work at their current or most recent job.

After the participants completed the survey, the researchers requested that they give flyers to their two peers inviting them to participate in the second portion of the study. The flyer inviting the peers to participate in the study briefly described the purpose of the study, provided information regarding how to contact the researchers if the peer was interested in participating in the study (this was done via email), list a numerical code that the peer needed to provide when emailing the researchers so they can be linked to the original participant, and briefly described the incentive for participating (submission of their name in a drawing for a $100 gift card). When a peer emailed the researchers with their numerical code to participate in the study, the researchers responded with a link to a short, confidential online survey. Additionally, peers were reminded in the email that their name would be put in a drawing for a $100 gift card once they completed the survey. Upon clicking on the link, they were directed to an informed consent page that more thoroughly described the study. They were also given the contact information for the researchers in case they had any questions during or after taking the survey. Once they provided their consent (by electronic signature) they began the study. The first page required participants to input the numerical code listed on their flyer. Next, they were asked to provide self-ratings of their performance of helping and
voice behaviors. Then, they were prompted to complete demographic information about their age, race/ethnicity, their tenure at the organization at which they work with the original participant, and how many hours they work a week. Upon completing the survey, each peer’s name was put in the drawing for a $100 gift card. Unfortunately, this process resulted in a sample of only eleven peers, and therefore no analyses were conducted with the peer data.

**STUDY 2 MATERIALS**

In addition to the performance rating, gender personal identity, gender social identity, gender identity salience, modern sexism, and OCB expectations by gender scales used in Study 1, this study included the use of scales that measure overall performance, self- and other- ratings of OCB, and job interdependence. Details regarding these scales are presented herein.

**Overall Performance**

Two separate measures of the overall performance of the participants’ two coworkers were included in this study. The first was adapted from the four-item Soltani et al. (2013) performance rating measure used in the policy-capturing scenarios in Study 1 (see Appendix H). Like the original scale, each item was on a 5-point Likert scale (ranging from 1 = to little or no extent to 5 = to a great extent). However, in place of “this person” or “this employee” gendered pronouns for each respective coworker were used. The alpha for this measure was 0.96 for female coworkers, and 0.94 for male coworkers.

**Task Performance**

A measure of task performance was included in the study for the purpose of running exploratory analyses to explore the effect of OCBs on overall performance when task performance is taken into account. The measure was adapted from Williams and Anderson’s (1991) seven-item measure of in-role behavior (see Appendix I). This measure included items such as “(S)he adequately completes her/his assigned duties,” “(S)he meets formal performance requirements of the job,” and “(S)he neglects aspects of the job (s)he is obligated to perform.” This scale used a 5-point Likert response scale (1 = never and 5 = always). The alpha for this measure was 0.86 for female coworkers, and 0.89 for male coworkers.
Self-Ratings of OCB

As part of this study, the male and female peers of the original participant were instructed to provide ratings of the extent to which they engage in different OCBs in the workplace (see Appendix J). The five-item measure for helping behavior was adapted from P. M. Podsakoff et al. (1990), and included statements such as “I help co-workers with help-related problems” and “I help orient new workers even though it is not required.” The six-item measure for voice behavior was adapted from Van Dyne and LePine (1998), and included statements such as “I speak up and encourage others in my group to get involved in issues that affect the group” and “I speak up in my group with ideas for new projects or changes in procedures.” All self-rating of OCB items used the five-point Likert response scale (1 = to a very small extent and 5 = to a great extent) from P. M. Podsakoff et al.’s (1990) original scale. The alpha of these scales were not calculated, as the sample size of peers was too small to be included in any of the analyses (n=11).

Other-Ratings of OCB

Taking the suggestion of Allen, Barnard, Rush, and Russell (1996), ratings of OCB were collected from multiple sources per subject (i.e., both the original participant and their peer). The items for this scale were adapted from the self-ratings of OCB scale (voice behavior items used the original, non-adapted items from Van Dyne and LePine, 1998), and prompted participants to rate the extent to which their peers individually perform helping and voice behaviors, using a five-point Likert response scale (1 = to a very small extent and 5 = to a great extent; see Appendix K). The alpha scores for helping and voice behaviors were 0.94 and 0.93 for ratings of female peers, and 0.93 and 0.94 for ratings of male peers.

Job Interdependence

Previous research has suggested that the relationship between OCB (specifically helping behavior) and performance depends on the level of job or task interdependence between or among group members (Bachrach, Powell, Collins, & Richey, 2006). Specifically, these findings suggested that OCB more strongly affects performance ratings as task interdependence increases. Taking this literature into account, job interdependence was controlled for in this study using Pearce, Sommer, Morris, and Fridiger’s (1992) six-item
scale. These items, which were on a 5-point Likert scale (1 = strongly disagree and 5 = strongly agree), included statements such as “The tasks related to my job can be performed fairly independently of this person” and “The tasks related to my job require frequent coordination with the efforts of this person.” See Appendix L for the full scale.

**STUDY 2 STATISTICAL ANALYSES**

Similar to what was done for Study 1, the proposed hypotheses were analyzed using HLM with the Mplus software package (Muthén & Muthén, 2010) in Study 2. The level 1 variables were ratee gender, helping behavior, voice, and the cross products of gender with helping and voice (with overall performance ratings as the outcome variable). The following variables were at level 2: rater gender personal identity, rater gender social identity, rater gender identity salience, and rater modern sexism, along with rater gender and job interdependence as control variables. For Hypotheses 1, 2a and 2b, a random coefficients model in HLM was performed by regressing performance ratings on ratee gender, helping behavior, voice, and the cross products of ratee gender and each of the OCB dimensions, with rater gender and job interdependence included as controls. In order to analyze Hypotheses 3 and 6, the interaction terms for these analyses at level 1 (genderXhelping, and genderXvoice) were then regressed as outcome variables on the level 2 predictor variables of rater gender personal identity and rater modern sexism, controlling for rater gender and job interdependence, in two separate models. Similarly, in order to analyze Hypotheses 4a, 4b, 5a, and 5b, the interaction terms for these analyses at level 1 (genderXhelping, and genderXvoice) were regressed as outcome variables on the level 2 predictor variable of rater gender social identity (for Hypotheses 4a and 4b) and rater gender identity salience (for Hypotheses 5a and 5b), controlling for rater gender and job interdependence, but with the dataset split by rater gender before running analyses.
CHAPTER 3

RESULTS

Once the data were collected using the two methodologies discussed in the previous section, statistical analyses were run to test the proposed hypotheses. In the following section, I will discuss the initial analyses, the findings for Study 1, and then the findings for Study 2.

INITIAL ANALYSES

Prior to testing the hypotheses proposed by this thesis, correlational analyses were run for all of the between-subjects variables (see Table 1). Gender personal identity was significantly related to both gender social identity \( (r=0.19, p<0.01) \) and gender identity salience \( (r=0.30, p<0.01) \). This suggests that raters with higher gender personal identity were also likely to have higher gender social identity and higher gender identity salience. Gender social identity was also significantly related to gender identity salience, \( (r=0.22, p<0.01) \), suggesting that raters with higher gender social identity were also likely to have higher gender identity salience. Finally, participant sex was significantly related to gender social identity \( (r=-0.30, p<0.01; \ M=3.04, \ SD=0.93 \text{ for females}, \ M=2.37, \ SD=0.85 \text{ for males}) \) and modern sexism \( (r=0.23, p<0.01; \ M=2.22, \ SD=0.50 \text{ for females}, \ M=2.51, \ SD=0.64 \text{ for males}) \). This suggests that female raters were more likely than male raters to have a higher gender social identity as well as higher levels of modern sexism (since lower scores on this scale indicated higher levels of modern sexism).

In addition to conducting correlational analyses, means and standard deviations of the two OCB expectations by gender scales were calculated. Though formal hypotheses were not described, participants were expected to state that helping and voice behaviors are equally expected of both women and men, regardless of their personal beliefs, because of the explicit nature of the items (they directly ask whether each behavior is more expected women or men, as opposed to doing so in a more covert manner). In line with this, the means of both the single-item and multiple-item measures for helping behavior \( (M=3.80, SD=0.93 \text{ and } M=3.76, SD=0.92) \).
Table 1. Means, Standard Deviations, and Intercorrelations among between-Subjects Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Gender personal identity</td>
<td>3.65</td>
<td>.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Gender social identity</td>
<td>2.86</td>
<td>.96</td>
<td>.194**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Gender identity salience</td>
<td>2.50</td>
<td>.90</td>
<td>.304**</td>
<td>.218**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Modern sexism</td>
<td>2.30</td>
<td>.55</td>
<td>-.065</td>
<td>-.069</td>
<td>-.051</td>
<td></td>
</tr>
<tr>
<td>(5) Sex a</td>
<td>.27</td>
<td>.44</td>
<td>-.024</td>
<td>-.304**</td>
<td>-.100</td>
<td>.233**</td>
</tr>
</tbody>
</table>

Note. N=234.  
* Coded 0=female, 1=male.  
** Correlation is significant at the 0.01 level (2-tailed).

SD=0.78, respectively) and voice behavior (M=3.98, SD=0.93 and M=3.87, SD=0.68, respectively) were near the midpoint. The response scale for these items was on a 7-point scale ranging from 1 = much more expected of women to 7 = much more expected of men (with 4 = equally expected of men and women), so the average response was close to equal expectations for women and men, though they erred a bit on the side of being more expected of women. A one-sample t-test was conducted to see if scores on the single- and multiple-item scales differed significantly from the midpoint, and it was found that they did for both of the helping behavior measures (t(226)=-3.27, p<0.01 and t(226)=-4.70, p<0.001, respectively), but only for the multiple-item item measure for voice behavior (t(226)=0.36, ns and t(226)=-2.90, p<0.01, respectively). Additionally, a paired-samples t-test was conducted to see if the means of the single- and multiple-item measures of helping and voice behavior significantly differed from each other, and it was found that they did (t(226)=-2.71, p<0.01 and t(226)=2.16, p<0.05, respectively). These results suggest that participants considered helping behavior to be a bit more expected of women than men compared to voice behavior.

STUDY 1

Before running HLM analyses for Study 1, an intercept-only model was run that revealed an ICC coefficient of 0.007. This indicates that 0.7% of the variance in overall performance ratings of the scenarios was between raters and 99.3% of the variance in overall performance evaluations of the scenarios was within individual raters (i.e., individual responses to the different conditions within scenarios). The percent of variance that existed between individual did not meet the 5% cutoff suggested by Reise, Ventura, Nuechterlein,
and Kim (2005), suggesting that these data do not have sufficient between level variance to warrant HLM analyses. Subsequent analyses, which are discussed next, only partially confirmed that the only significant effects for Study 1 are at level 1.

To test Hypotheses 1a and 1b, a random coefficients model in HLM was performed by regressing performance ratings of the hypothetical scenarios on the gender, helping behavior, voice behavior, and task performance of the hypothetical employee (otherwise known as the target). The results of these analyses can be found in Table 2. Hypothesis 1a predicted that helping behavior would positively predict overall performance ratings. In support of this hypothesis, the unstandardized regression coefficient relating helping behavior to overall performance ratings was statistically significant ($b=0.93, p<0.001$).

Hypothesis 1b predicted that voice behavior would positively predict overall performance ratings. In support of this hypothesis, the unstandardized regression coefficient relating voice behavior to overall performance ratings was statistically significant ($b=0.78, p<0.001$). Though formal hypotheses were not laid out relating task performance to overall performance ratings, it should be noted that this relationship was statistically significant as well ($b=1.73, p<0.001$). Additionally, it should be noted that the relationship between target gender and overall performance ratings was not significant in this model ($b=0.01, ns$), suggesting that female and male targets did not receive significantly different performance evaluations, controlling for the task, helping, and voice behaviors of those targets. To get a better understanding of the relationship between our level 1 variables and evaluation ratings, we computed the variance explained in line with recommendations from Hofmann, Griffin, and Gavin (2000). The calculated $R^2$ value, presented at the bottom of Table 2, suggested that, averaged across all conditions, our main effects accounted for roughly 84% of the variance in evaluation ratings.

To test Hypotheses 2a and 2b, interaction terms between target gender and each of the behaviors were added to the level 1 variables in separate models. Hypothesis 2a predicted that women would not be rewarded in their overall performance ratings for performing helping behavior, whereas men would be rewarded for performing helping behavior. This was not supported by the unstandardized regression coefficient relating the cross product of gender and helping behavior to overall performance ratings ($b=-0.05, ns$). Thus, female
Table 2. Level 1 Analysis: The Effects of Scenario Conditions on Overall Performance Ratings

<table>
<thead>
<tr>
<th>Variables</th>
<th>Evaluation rating</th>
<th>Variance Component$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b^a$</td>
<td>SE$^b$ $t$</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.00**</td>
<td>0.033 30.66 0.14</td>
</tr>
<tr>
<td>Target gender</td>
<td>0.01</td>
<td>0.012 0.94 0.00</td>
</tr>
<tr>
<td>Helping behavior</td>
<td>0.93**</td>
<td>0.026 35.84 0.09</td>
</tr>
<tr>
<td>Voice behavior</td>
<td>0.78**</td>
<td>0.024 31.92 0.07</td>
</tr>
<tr>
<td>Task performance</td>
<td>1.73**</td>
<td>0.043 39.80 0.32</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: N = 234; **= $p < 0.001$.

$^a$ = Unstandardized coefficient.
$^b$ = Average estimated standard error of the Level 1 regression coefficients.
$^c$ = Variance in Level 1 parameter estimates and chi-square test of significance of variance.
$^d$ = Percentage of variance in the dependent variable explained by main variable.

targets did not receive significantly different performance ratings compared to male targets. Hypothesis 2b predicted that women would not be rewarded in their overall performance ratings for performing voice behavior, whereas men would be rewarded for performing voice behavior. The unstandardized regression coefficient relating the cross product of gender and voice behavior to overall performance ratings was significant ($b=0.06, p<0.05$). As illustrated in Figure 1, overall performance ratings differed by target gender, such that men were evaluated more favorably when voice behavior was high (3.52 vs. 3.48), though further probing showed that this difference was not significant (simple slope=0.039, $ns$). Conversely, when voice was low, women were evaluated more favorably than men (2.73 vs. 2.71), which further probing found was a significant difference (simple slope=-0.02, $p<0.001$). As an exploratory analysis, the effect of the cross product of target gender and task performance on performance ratings was tested, and was supported by the unstandardized regression coefficient ($b=0.06, p<0.05$). As illustrated in Figure 2, overall performance ratings differed by target gender, such that men were evaluated more favorably when task performance was high (4.47 vs. 4.43), though further probing showed that this difference was not significant (simple slope=0.039, $ns$). Conversely, when task performance was low, women were evaluated slightly more favorably than men (2.73 vs. 2.71), which further probing found was a significant difference (simple slope=-0.02, $p=0.001$).

Prior to running analyses for the remaining hypotheses, models were run to examine whether the sex of the rater had any influence on the relationship between OCB and overall
Figure 1. Effect of ratee gender and voice behavior on performance ratings (Hypothesis 2b).

Figure 2. Effect of ratee gender and task performance on performance ratings.

performance ratings. To do this, rater sex was included as a level 2 between-subjects variable, and then the level 1 variables from the previous analyses and the slope of the cross product of target gender and each of the behaviors (helping, voice, and task) were regressed on rater sex in three separate models. None of the 3-way interactions tested were significant
(γ_{31}=-0.001, \textit{ns}; \gamma_{31}=0.00, \textit{ns}; \text{and } \gamma_{31}=0.001, \textit{ns}, respectively), nor did rater sex significantly predict overall performance ratings in general (γ_{01}=0.001, \textit{ns}; \gamma_{01}=0.001, \textit{ns}; \text{and } \gamma_{01}=0.001, \textit{ns}, respectively).

To test Hypothesis 3, rater gender personal identity was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of target gender and voice behavior were regressed on rater gender personal identity. Hypothesis 3 predicted that men would receive higher overall performance ratings when they perform voice behavior regardless of the gender personal identity level of the rater, whereas women would be rewarded for performing voice behavior when the rater has low gender personal identity, and would be punished for performing voice behavior when the rater has high gender personal identity. This was not supported by the unstandardized regression coefficient relating the cross product of scenario gender, voice behavior, and rater gender personal identity to overall performance ratings (γ_{31}=-0.001, \textit{ns}). As an exploratory analysis, the same model was run with helping behavior in place of voice behavior, and this was not supported by the unstandardized regression coefficient relating the cross product of target gender, helping behavior, and rater gender personal identity to overall performance ratings (γ_{31}=0.001, \textit{ns}).

To test Hypotheses 4a and 4b, gender social identity was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of target gender and voice behavior and the slope of the cross product of target gender and helping behavior were regressed on gender social identity in two separate models. These analyses were run three separate times: once with the full sample, once with just the female raters, and once with just the male raters. Hypothesis 4a and 4b predicted that female and male raters with low gender social identity would give both women and men higher performance ratings when they perform OCB compared to when they do not, whereas female (Hypothesis 4a) and male (Hypothesis 4b) raters with high gender social identity would give their gender higher performance ratings when they perform OCB compared to when they do not, but will give the other gender generally lower performance ratings regardless of whether or not they perform OCB. Hypothesis 4a was not supported by the unstandardized regression coefficient relating the cross product of target gender, voice behavior, and rater gender social
identity to overall performance ratings in the female portion of the sample ($\gamma_{31}=0.00$, $ns$). The same model with helping behavior was not significant either ($\gamma_{31}=0.001$, $ns$). Hypothesis 4b was also not supported by the unstandardized regression coefficient relating the cross product of target gender, voice behavior, and rater gender social identity to overall performance ratings in the male portion of the sample ($\gamma_{31}=0.00$, $ns$). The same model with helping behavior was not significant either ($\gamma_{31}=0.001$, $ns$). Additionally, the models for these hypotheses run with the full sample were not significant for either voice behavior ($\gamma_{31}=0.00$, $ns$) or helping behavior ($\gamma_{31}=0.001$, $ns$).

To test Hypotheses 5a and 5b, gender identity salience was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of target gender and voice behavior and the slope of the cross product of target gender and helping behavior were regressed on gender identity salience in two separate models. As was done for Hypotheses 4a and 4b, these analyses were run three separate times: once with the full sample, once with just the female raters, and once with just the male raters. Hypothesis 5a and 5b predicted that female and male raters with low gender identity salience would give both women and men higher performance ratings when they perform OCB compared to when they do not, whereas female (Hypothesis 5a) and male (Hypothesis 5b) raters with high gender identity salience would give their gender higher performance ratings when they perform OCB compared to when they do not, but will give the other gender generally lower performance ratings regardless of whether or not they perform OCB.

Hypothesis 5a was not supported by the unstandardized regression coefficient relating the cross product of target gender, voice behavior, and rater gender identity salience to overall performance ratings in the female portion of the sample ($\gamma_{31}=-0.001$, $ns$). The same model with helping behavior was not significant either ($\gamma_{31}=0.00$, $ns$). Hypothesis 5b was also not supported by the unstandardized regression coefficient relating the cross product of target gender, voice behavior, and rater gender social identity to overall performance ratings in the male portion of the sample ($\gamma_{31}=0.00$, $ns$). As was the case for Hypothesis 5a, the same model with helping behavior was not significant either ($\gamma_{31}=0.002$, $ns$). Additionally, the model for these hypotheses run with the full sample was not significant for either voice behavior ($\gamma_{31}=-0.001$, $ns$) or helping behavior ($\gamma_{31}=0.001$, $ns$).
To test Hypothesis 6, rater modern sexism was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of target gender and voice behavior were regressed on rater modern sexism. Hypothesis 6 predicted that men would receive higher overall performance ratings when they perform voice behavior regardless of the modern sexism level of the rater, whereas women would be rewarded for performing voice behavior when the rater is low in modern sexism, and would be punished for performing voice behavior when the rater is high in modern sexism. This was not supported by the unstandardized regression coefficient relating the cross product of scenario gender, voice behavior, and rater modern sexism to overall performance ratings ($\gamma_{31}=0.001$, ns). As an exploratory analysis, the same model was run with helping behavior in place of voice behavior, and this was not supported by the unstandardized regression coefficient relating the cross product of scenario gender, helping behavior, and rater gender personal identity to overall performance ratings ($\gamma_{31}=0.00$, ns).

**STUDY 2**

As was done in Study 1, an intercept-only model was run for Study 1 prior to running HLM analyses that revealed an ICC coefficient of 0.124. This indicates that 12.4% of the variance in overall performance ratings of the scenarios was between raters and 87.6% of the variance in overall performance evaluations of the scenarios was within individual raters (i.e., individual responses to the different conditions within scenarios). The percent of variance that existed between individual did meet the 5% cutoff suggested by Reise and colleagues (2005), suggesting that these data have sufficient between level variance to warrant HLM analyses.

Like in Study 1, to test Hypotheses 1a and 1b, a random coefficients model in HLM was performed by regressing the overall performance ratings of the raters’ two coworkers (one female, one male) on the gender, helping behavior, voice behavior, and task performance of the coworkers, as well as job interdependence between the rater and each of their coworkers. The results of these analyses can be found in Table 3. Hypothesis 1a predicted that helping behavior would positively predict overall performance ratings. In support of this hypothesis, the unstandardized regression coefficient relating helping behavior to overall performance ratings was statistically significant ($b=1.25$, $p<0.001$).
### Table 3. Level 1 Analysis: The Effect of Coworker Gender and OCBs on Overall Performance Ratings

<table>
<thead>
<tr>
<th>Variables</th>
<th>Evaluation rating</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b(^a)</td>
<td>SE(^b)</td>
<td>t</td>
<td>Variance Component(^c)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.72(*)</td>
<td>0.286</td>
<td>2.51</td>
<td>0.05</td>
</tr>
<tr>
<td>Coworker gender</td>
<td>0.05</td>
<td>0.214</td>
<td>0.23</td>
<td>0.001</td>
</tr>
<tr>
<td>Helping behavior</td>
<td>1.25(**)</td>
<td>0.191</td>
<td>6.53</td>
<td>0.00</td>
</tr>
<tr>
<td>Voice behavior</td>
<td>0.51(**)</td>
<td>0.075</td>
<td>6.83</td>
<td>0.00</td>
</tr>
<tr>
<td>Job interdependence</td>
<td>0.03</td>
<td>0.113</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>R(^d)</td>
<td></td>
<td></td>
<td></td>
<td>0.63</td>
</tr>
</tbody>
</table>

Note: N = 171; \(*\) = p < 0.05; \(**\) = p < 0.001.

\(^a\) = Unstandardized coefficient.
\(^b\) = Average estimated standard error of the Level 1 regression coefficients.
\(^c\) = Variance in Level 1 parameter estimates and chi-square test of significance of variance.
\(^d\) = Percentage of variance in the dependent variable explained by main variables.

Hypothesis 1b predicted that voice behavior would positively predict overall performance ratings. In support of this hypothesis, the unstandardized regression coefficient relating voice behavior to overall performance ratings was statistically significant (b=0.51, p<0.001).

Additionally, it should be noted that the relationship between coworker gender and overall performance ratings, like in Study 1, was not significant in this model (b=0.05, ns), suggesting that male and female coworkers did not receive significantly different performance ratings, controlling for their level of helping and voice behaviors as well as job interdependence. Though formal hypotheses were not laid out relating task performance to overall performance ratings, it should be noted that this relationship was statistically significant in a model that included this behavior in addition to the OCBs (b=0.69, p<0.001).

To get a better understanding of the relationship between our level 1 variables and evaluation ratings, we computed the variance explained (in line with recommendations from Hofmann et al., 2000). The calculated R\(^d\) value, presented at the bottom of Table 3, suggested that, averaged across all conditions, our main effects accounted for roughly 63% of the variance in evaluation ratings.

To test Hypotheses 2a and 2b, interaction terms between coworker gender and each of the behaviors were added to the level 1 variables in separate models. Hypothesis 2a predicted that female coworkers would not be rewarded in their overall performance ratings for performing helping behavior, whereas male coworkers would be rewarded for performing helping behavior. Like in Study 1, this was not supported by the unstandardized regression
coefficient relating the cross product of gender and helping behavior to overall performance ratings ($b=0.43, ns$). Thus, female coworkers did not receive significantly different performance ratings compared to male coworkers. Hypothesis 2b predicted that female coworkers would not be rewarded in their overall performance ratings for performing voice behavior, whereas male coworkers would be rewarded for performing voice behavior. Unlike in Study 1, this was not supported by the unstandardized regression coefficient relating the cross product of gender and voice behavior to overall performance ratings ($b=0.31, ns$). Thus, female coworkers did not receive significantly different performance ratings compared to male coworkers. Additionally, the effect of the cross product of target gender and task performance on performance ratings was tested as an exploratory analysis, and unlike in Study 1, was not significant ($b=0.02, ns$).

As was done in Study 1, prior to running analyses for the remaining hypotheses, models were run to examine whether the sex of the rater had any influence on the relationship between OCB and overall performance ratings. To do this, rater sex was included as a level 2 between-subjects variable, and then the level 1 variables from the previous analyses and the slope of the cross product of target gender and each of the behaviors (helping, voice, and task) were regressed on rater gender in three separate models. None of the 3-way interactions tested were significant ($\gamma_{31}=-0.17, ns; \gamma_{31}=-0.14, ns; \text{ and } \gamma_{31}=-0.03, ns$, respectively), nor did rater gender significantly predict overall performance ratings in general ($\gamma_{01}=-0.30, ns; \gamma_{01}=-0.14, ns; \text{ and } \gamma_{01}=0.33, ns$, respectively).

To test Hypothesis 3, rater gender personal identity was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of coworker gender and voice behavior were regressed on rater gender personal identity. Hypothesis 3 predicted that male coworkers would receive higher overall performance ratings when they perform voice behavior regardless of the gender personal identity level of the rater, whereas female coworkers would be rewarded for performing voice behavior when the rater has low gender personal identity, and would be punished for performing voice behavior when the rater has high gender personal identity. This was not supported by the unstandardized regression coefficient relating the cross product of coworker gender, voice behavior, and rater gender personal identity to overall performance ratings ($\gamma_{31}=0.06, ns$). As
an exploratory analysis, the same model was run with helping behavior in place of voice behavior, and this was not supported by the unstandardized regression coefficient relating the cross product of coworker gender, helping behavior, and rater gender personal identity to overall performance ratings ($\gamma_{31}=0.11$, ns).

To test Hypotheses 4a and 4b, gender social identity was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of coworker gender and voice behavior as well as the slope of the cross product of coworker gender and helping behavior were regressed on gender social identity in two separate models. These analyses were run three separate times: once with the full sample, once with just the female raters, and once with just the male raters. Hypothesis 4a and 4b predicted that female and male raters with low gender social identity would give both female and male coworkers higher performance ratings when they perform OCB compared to when they do not, whereas female (Hypothesis 4a) and male (Hypothesis 4b) raters with high gender social identity would give coworkers of their gender higher performance ratings when they perform OCB compared to when they do not, but will give coworkers of the other gender generally lower performance ratings regardless of whether or not they perform OCB. Hypothesis 4a was not supported by the unstandardized regression coefficient relating the cross product of coworker gender, voice behavior, and rater gender social identity to overall performance ratings in the female portion of the sample ($\gamma_{31}=0.008$, ns). Additionally, the same model hypothesized by Hypothesis 4a, but with helping behavior, was not significant for female raters ($\gamma_{31}=0.05$, ns). Hypothesis 4b was also not supported by the unstandardized regression coefficient relating the cross product of scenario gender, voice behavior, and rater gender social identity to overall performance ratings in the male portion of the sample ($\gamma_{31}=-0.005$, ns). As was the case for Hypothesis 4a, the same model with helping behavior was not significant either ($\gamma_{31}=0.20$, ns). Additionally, the model for these hypotheses run with the full sample was not significant for the interaction of coworker gender with neither voice behavior ($\gamma_{31}=-0.002$, ns) or helping behavior ($\gamma_{31}=0.08$, ns).

To test Hypotheses 5a and 5b, gender identity salience was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of coworker gender and voice behavior as well as the slope of the cross product of
coworker gender and helping behavior were regressed on gender identity salience in two separate models. As was done for Hypotheses 4a and 4b, these analyses were run three separate times: once with the full sample, once with just the female raters, and once with just the male raters. Hypothesis 5a and 5b predicted that female and male raters with low gender identity salience would give both female and male coworkers higher performance ratings when they perform OCB compared to when they do not, whereas female (Hypothesis 5a) and male (Hypothesis 5b) raters with high gender identity salience would give coworkers of their gender higher performance ratings when they perform OCB compared to when they do not, but will give coworkers of the other gender generally lower performance ratings regardless of whether or not they perform OCB. Hypothesis 5a was not supported by the unstandardized regression coefficient relating the cross product of coworker gender, voice behavior, and rater gender identity salience to overall performance ratings in the female portion of the sample ($\gamma_{31}=-0.05, ns$). The same model with helping behavior was not significant either ($\gamma_{31}=-0.01, ns$). Hypothesis 5b was also not supported by the unstandardized regression coefficient relating the cross product of scenario gender, voice behavior, and rater gender social identity to overall performance ratings in the male portion of the sample ($\gamma_{31}=0.32, ns$). As was the case for Hypothesis 5a, the same model with helping behavior was not significant either ($\gamma_{31}=0.13, ns$). Additionally, the model for these hypotheses run with the full sample was not significant for either voice behavior ($\gamma_{31}=0.05, ns$) or helping behavior ($\gamma_{31}=0.03, ns$).

To test Hypothesis 6, rater modern sexism was included as a level 2 variable. Then, the level 1 variables from previous analyses and the slope of the cross product of coworker gender and voice behavior were regressed on rater modern sexism. Hypothesis 6 predicted that male coworkers would receive higher overall performance ratings when they perform voice behavior regardless of the modern sexism level of the rater, whereas female coworkers would be rewarded for performing voice behavior when the rater is low in modern sexism, and would be punished for performing voice behavior when the rater is high in modern sexism. This was not supported by the unstandardized regression coefficient relating the cross product of coworker gender, voice behavior, and rater modern sexism to overall performance ratings ($\gamma_{31}=0.08, ns$). As an exploratory analysis, the same model was run with helping behavior in place of voice behavior, and this was not supported by the
unstandardized regression coefficient relating the cross product of coworker gender, helping behavior, and rater gender personal identity to overall performance ratings ($\gamma_{31}=-0.14$, ns).
CHAPTER 4

DISCUSSION

The purpose of this thesis was to explore the mechanisms behind gender-based differential treatment in organizational settings. It was hypothesized that, while helping behavior and voice behavior would be positively related to overall performance ratings, women would be rewarded less than men in their overall performance ratings for performing helping behavior and voice behavior. Taking into account previous research findings, as well as established theories related to gender identity and sexism (e.g., status characteristics theory, social identity theory, self categorization theory), it was also proposed that the interaction between gender and voice behavior would be differentially related to overall performance ratings based on participants’ levels of gender personal identity, gender social identity, gender identity salience, and modern sexism. In the following section, the results of this thesis will be examined, the limitations of the two studies will be discussed, and future directions for research on gender and OCB will be detailed.

INITIAL ANALYSES

The initial correlational analyses run for the between-subjects variables of interest revealed significant correlations among gender personal identity, gender social identity, and gender identity salience, as well as between participant sex and both gender social identity and modern sexism. The significant relationships among gender personal identity, gender social identity, and gender identity salience were expected, as each of these constructs simply focus on a different aspect of an individual’s identification with her or his gender (the first relating to defining oneself as an individual, the second relating to defining oneself as part of a group, and the third relating to how prominent the gender of others is to an individual; Brewer & Gardner, 1996; Randel, 2002; Randel & Jaussi, 2008; Rosenberg, 1979). It should be noted that, although the correlations were significant, they were not extremely high, suggesting that the scales did effectively capture different aspects of participants’ identities.
It should also be noted that previous research has found a significant correlation between gender personal identity and gender social identity ($r=0.36, p<0.01$; Randel & Jaussi, 2008).

The analyses conducted with the OCB expectations by gender scales supported the informal expectation that, when explicitly asked to state whether helping and voice behaviors are expected more of women or men, participants will err on the side of stating that behaviors are equally expected of women and men. This likely occurred because participants felt that it would be socially undesirable to explicitly state that women and men are expected to perform different kinds of behaviors (as this might be considered sexist), even though they may feel otherwise in reality (Krumpal, 2013). The analyses also suggested that, although the means for both helping and voice behavior were near the center of the response scale, ratings of the gender expectations of helping behavior showed that it is slightly more expected of women than men compared to voice behavior. Though this difference is slight, it suggests that participants tended to consider helping behavior to be a bit more expected of women even when explicitly asked.

**Studies 1 and 2**

In this section, the findings of the two studies will be discussed as they relate to this thesis’ hypotheses, as well as to a number of exploratory analyses. Throughout this section, various explanations for why specific hypotheses were or were not supported will be explored, focusing on both theoretical and methodological issues. Since the two studies had largely comparable findings, they will be discussed together. However, some distinctions will be made when discussing potential methodological reasons for why particular hypotheses were not supported.

Hypotheses 1a and 1b, which predicted that helping and voice behavior would each positively predict overall performance ratings, were supported in both Studies 1 and 2. This aligns with previous research in this area that showed that OCBs are given substantial weight when overall performance ratings are formed (e.g., Borman et al., 1995; Johnson, 2001; N. P. Podsakoff et al., 2009; Rotundo & Sackett, 2002; Van Dyne & LePine, 1998; Whiting et al., 2008). Additionally, both studies found that task performance predicts overall performance ratings. This also aligns with the previous research mentioned above, and provides evidence that OCBs are given significant weight when forming performance ratings even when task
performance is explicitly discussed (in the case of Study 1) or measured (in the case of Study 2). It is important for research in this area to account for the effects of task performance when measuring the effect of OCBs on overall performance ratings, since all these behaviors are taken into account when performance ratings are formed in real organizational settings (as opposed to just one or the other; Rotundo & Sackett, 2002). Notably, the effect of target or coworker gender was not significant in either study. It may be the case that participants were affected by social desirability bias since it is not generally socially accepted to explicitly favor one gender over the other (as this is considered sexist, and sexism is not typically socially desirable behavior; Krumpal, 2013).

Hypotheses 2a and 2b predicted that female coworkers would not be rewarded in their overall performance ratings for performing helping (Hypothesis 2a) or voice (Hypothesis 2b) behavior, whereas male coworkers would be rewarded for performing each behavior. Contrary to previous research (e.g., Allen, 2006; Allen & Rush, 2001; Heilman & Chen, 2005) and theory (e.g., Fiske et al., 2002; Kidder & McLean Parks, 2001; Ridgeway, 1991), Hypothesis 2a was not supported by either study, and Hypothesis 2b was only supported by Study 1. It should be noted that the interaction between gender and task performance, which was run as an exploratory analysis, was significant in Study 1 but not in Study 2. After further probing the two significant interactions in Study 1 for voice and task performance, it was found that male targets were punished more than female targets for not performing voice behavior as well as task behavior. The non-significant findings for helping behavior, as well as the non-significant simple slopes for high voice and high task performance for female and male targets, run directly contrary from research by Heilman & Chen (2005) and Allen (2006), which reported that men are rewarded more than women for performing OCBs regardless of whether those OCBs are congruent or incongruent with masculine stereotypes. The significant simple slopes for low voice behavior and low task performance, which unexpectedly show that men are punished more than women for not performing these behaviors, may parallel Heilman and Chen’s (2005) findings that women are punished more for performing gender congruent behavior. The findings for voice behavior may suggest that men are punished for not performing this behavior because they are expected to do so, and so when they do not, they are held more accountable than women would be. Since this proposition isn’t supported by previous evidence (namely Allen, 2006), more research should
be done in this area to see if the findings presented herein can be replicated. A potential explanation for the non-significant findings in Studies 1 and 2 (as well as the unique significant findings) is that the participants were young adults, most of whom did not work in managerial positions, and so are considered novices in terms of experience providing performance ratings (Aiman-Smith et al., 2002). Previous research has shown that novices and experts have different approaches to making decisions (Hardiman, Dufresne, & Mestre, 1989; Mackay, Barr, & Kletke, 1992; Polson & Jeffries, 1982), and researchers have been cautioned against using novices for such tasks in research and then extrapolating findings to managers and other, more experienced professionals (e.g., Barr & Hitt, 1986; Shanteau & Stewart, 1992). Because the raters in these studies were novices, they likely have not yet formed a nuanced policy for giving performance ratings, and so may not serve as a representative sample for how performance ratings are affected by different traits and behaviors of the ratees.

Before discussing each of the remaining hypotheses (Hypotheses 3-6) in sequence, a number of notes should be made that may explain why these proposed models in particular were not supported by the findings of the two studies. First, the sample sizes for the two studies (especially the second study) were not very large. This is especially the case when you consider the complexity of the models proposed, as each of the hypotheses predicted either 3-way (Hypotheses 3 and 6) or 4-way (Hypotheses 4a-5b) mixed-level interactions. Thus, there may not have been enough statistical power to find significant results for these high-level interactions if the relationships do actually exist. Additionally, the ICC for Study 1 revealed that only 0.7% of the variance in overall performance ratings was accounted for by between-subjects variables. This suggests that it was very unlikely that any between-level variables would affect overall performance ratings, much less the interaction of between-level variables with within-level variables. However, the ICC of 12.4% for Study 2 was acceptable (Reise et al., 2005), and yet the proposed between-subjects variables did not have an effect on overall performance ratings in that study. This may potentially suggest that some other between-subjects construct that was not tested by these hypotheses may have affected overall performance ratings.

Hypotheses 3, 4a, 4b, 5a, and 5b made predictions about the role different aspects of the participants’ gender identity (gender personal identity, gender social identity, and gender
identity salience, respectively) play in the way they rate the OCBs and overall performance of female and male targets or coworkers. Each of these hypotheses posited that individuals who are high in a particular type of identity would differentially reward females and males for performing OCBs, while individuals who are low in a particular type of identity would reward females and males in a similar manner for performing OCBs. None of these hypotheses were supported by the two studies presented herein. While there is theoretical support for the idea that individuals’ gender identities affect the way they perceive the behavior of others (e.g., Linville, 1982; Ridgeway, 1991; Tajfel & Turner, 1986), it is possible that participants’ gender identities did not greatly affect their perceptions of the OCBs and overall performance of the targets and their coworkers because they are inexperienced as raters of performance. It has been recognized that experts and novices use different approaches to making decisions (Hardiman et al., 1989; Mackay et al., 1992; Polson & Jeffries, 1982), so we might expect that novice raters who do not have already-formed “policies” for forming ratings simply do not have enough experience to bring additional information regarding their gender identity into the rating “policy” they may have only formed as they completed the studies. Using this logic, we might find that different aspects of gender identity do affect older individuals who have managerial experience, but not younger individual who have little to no managerial experience.

Additionally, recent research on optimal distinctiveness theory (Brewer, 1991) suggests that young college students with low gender social identities may have a greater need to satisfy their needs for assimilation and differentiation, which then leads them to feel greater identification with their work group if it meets those needs (Randel, Zatzick, & Pearce, 2013). If the work groups of the participants in the current studies satisfy their needs for assimilation and differentiation, then they may identify with the work group as a whole more strongly than with different subgroups (e.g., the in-group for their gender). And if this is the case, then it may not be plausible to expect that the strength of their gender social identities would affect the way they rate the performance of others. Another potential explanation for why the hypotheses related to gender identity were not supported in these studies is that gender identity may be more important for college students in certain majors compared to others. For instance, there may be a difference in the strength of gender identities between psychology majors and engineering majors (especially female engineering
majors, who are typically a minority in their departments). The participants in the current studies were primarily psychology majors, so analyses that examine whether or not the strength of participants’ gender identities related significantly to their majors could not be conducted. With these issues in mind, future studies should consider replicating this study to examine the effects in an older study and in contexts where gender is a very salient individual difference.

Hypothesis 6 predicted that men would be rewarded for performing voice behavior regardless of the modern sexism level of the rater, whereas women would be rewarded for performing voice behavior only when the rater was low in modern sexism, and would be punished for performing voice behavior when the rater was high in modern sexism. The results of the two studies did not support this hypothesis, nor were the results of the same model significant when helping behavior replaced voice behavior. While the mean score for modern sexism in the sample indicates the presence of some modern sexist beliefs (the mean was 2.3; a lower score on the 5-point scale suggests stronger modern sexist beliefs), this may be an artifact of naïveté regarding gender issues of the relatively young participants as opposed to their sexist beliefs. Previous research on modern sexism has posited that college-aged women may be more sensitive to gender issues compared to non-college-aged women (Swim et al., 1995), but this was not the case in this sample, as female participants’ ratings suggested greater levels of modern sexism compared to the ratings of male participants. If the scale did not accurately measure their modern sexist beliefs, we should not expect those scores to relate to the performance ratings participants gave others. An alternative explanation is that participants who did really hold modern sexist beliefs did not consider the studies a “safe” situation in which they could reveal their modern sexists beliefs (Swim & Cohen, 1997), and so succumbed to social desirability bias and rated the targets and their coworkers in a more fair manner than they may have in another, less experimental, context.

LIMITATIONS

A number of limitations relating to this thesis should be noted. These limitations relate to different issues surrounding the sample for the studies, as well as the design of the two studies. In this section, these limitations will be discussed in detail.
There are several limitations related to the sample used for this thesis. As was discussed earlier in this section, one limitation of the two studies was that they had low sample size and statistical power considering the complexity of the models that were proposed. Bearing in mind that this thesis utilized HLM analyses, and that 3- and 4-way interactions were present in the models, it was unlikely that the findings would be significant with the sample size utilized in Studies 1 and 2. This is especially the case when you consider the fact that the dataset was split by the gender of the participants before running the analyses for the proposed four-way interactions, so the samples for these analyses were especially low relative to the sample size suggested by the power analyses that were conducted prior to data collection. Additionally, it was a limitation that the sample was not balanced in terms of participant sex. Considering the sample size was generally small and there were few males in the sample, conducting analyses for Hypotheses 4b and 5b (which focus on male participants) were particularly problematic. Finally, the use of relatively young participants who are mostly not working in managerial positions presented a challenge for the generalizability of this thesis’ findings (e.g., Barr & Hitt, 1986; Shanteau & Stewart, 1992). This is because these participants are considered novices in making decisions regarding the performance evaluations of employees (Aiman-Smith et al., 2002), and previous research has shown that novices and experts use different processes when making these sorts of decisions (Hardiman et al., 1989; Mackay et al., 1992; Polson & Jeffries, 1982).

Though a strength of this thesis was that it used multiple methodologies for testing its hypotheses, there are a few limitations related to each of the methods used that should be noted. Several of these ideas relate specifically to Study 1. One such limitation is that this study was done in an experimental setting (using “paper people”) with novice raters, which threatens its external validity (Rotundo & Sackett, 2002). Relatedly, because of the experimental within-subjects design of the study, participants may have been affected by social desirability bias (Krumpal, 2013), thus responding in ways they believe are expected, as opposed to responding in ways they would in a real life situation. Additionally, the scenarios focused on a single job, which is a general criticism of policy capturing design (Rotundo & Sackett, 2002). Furthermore, in an effort to minimize boredom among participants, the number of scenarios were limited to 16, which does not meet the recommendation by Aiman-Smith and colleagues (2002) to have 5 to 10 scenarios per
independent variables (this study only had four scenarios per independent variable). However, with regard to the scenarios, it was a strength of the study that the scenarios were randomized and that three scenarios were duplicated so that reliability of the ratings could be examined (Aiman-Smith et al., 2002).

For Study 2, one limitation relates to the general issue of same-source bias, since data on coworker behavior and performance were collected from single individuals (cf. P. M. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Williams & Brown, 1994). This thesis worked to combat this issue by requesting that participants recruit the two coworkers they rated to participate in a short survey (in line with the suggestion of Allen, 2006, to have ratings of OCB collected from multiple sources), but these data were not usable because of a rather low response rate (i.e., n=11).

Finally, it should be noted that the order of the studies may have affected the way participants provided ratings. If the participants, who were novice raters, formed their “policy” for rating the performance of others while completing Study 1 (as opposed to coming into the study with an already formed “policy;” Aiman-Smith et al., 2002), this newly formed “policy” may have affected the way they rated their coworkers’ job performance in Study 2. While recognizing this issue, the studies in this thesis were put in this order because there was a practical concern for the likelihood that participants would become bored or fatigued if the policy capturing study (the items of which can seem a bit repetitive) were later in the survey. The next section will discuss some ways in which future studies might combat this limitation.

**FUTURE DIRECTIONS**

There are a number of potential directions for future research on the relationship between gender and OCBs that should be noted. Some of these address specific limitations that were identified in this thesis, while others are general areas that would be interesting to research. In this section, a number of these directions are discussed.

Several steps should be taken in future research to combat some of the limitations of the two studies presented herein. With regard to the participants in future studies, there should be an effort to have a larger sample size with equal amounts of women and men. Additionally, participants should vary more in terms of age. Ideally, participants should have
more experience in managerial positions, so they will have real life experience evaluating the job performance of others, and so will not be considered novice raters who do have well-developed performance evaluation “policies” (Aiman-Smith et al., 2002). A future direction for Study 1 is to vary the job type of the person in the scenario according to feminine and masculine stereotypes (e.g., nurse vs. engineer). Doing so would be beneficial because it would expand the research that considers the gender stereotypes of the job as well as the OCBs of targets. A future direction for Study 2 is to have participants rate a larger number of their coworkers (as opposed to letting them choose one female and one male coworker, as was done in the current study). The ideal would be to have participants rate all of their coworkers, but this may not be feasible their work groups are large. Finally, future studies should consider randomizing the order of the two studies to see if there is a difference in the way participants rate their coworkers when they are not first exposed to a policy capturing study. An alternative option is to conduct the two studies separately with different samples so that we can be confident that the presence of one study does not affect responses in the other study.

There are also several areas for research on gender and OCB that should be explored in future research. There is still a need to investigate the individual difference variables that lead organizational members to vary the way they rate different employees. This study focused on different aspects of gender identity as well as modern sexism, but there may be other avenues worth examining as well. Another potential route for future studies is to use priming (Baars & McGovern, 1994) to get participants to think about gender stereotypes before they rate the performance of others and see this affects their ratings. One other possibility is to explore the possibility of a non-linear relationship between OCB and performance ratings, as well as how OCBs interact with each other to affect performance ratings. MacKenzie, Podsakoff, and Podsakoff (2011) began to explore these avenues, and found that voice behavior is only rewarded to a certain point, and that helping behavior moderates this relationship, but it would be useful to replicate these findings.

**CONCLUSION**

Overall, it was found that both helping behavior and voice behavior significantly contribute to overall performance ratings, based on ratings from experimentally manipulated
scenarios and from actual work settings. Additionally, there was support in the experimental study that men are given significantly lower overall performance ratings compared to women when they are low in voice behavior as well as in task performance. There was no support for the moderating between-persons variables of interest. Considering the limitations of these studies, future research should replicate these studies in line with the future directions outlined above, and see whether or not the hypotheses from this thesis may be supported in a larger sample of women and men who have managerial experience. Although there were few significant findings in this thesis, there is still a great deal of research that suggests that OCBs are important because they have an effect on decisions that influence career advancement and success (P. M. Podsakoff et al., 2000). Considering the influence of this behavior, it is vital to continue investigating how OCBs contribute to the overall performance ratings of individuals in the workplace. This means not only focusing on the individual employees being rated, but also contemplating the role of the rater. Doing so may shed light on how performance ratings are formed, and in a broader sense, how those ratings may eventually lead to a society with considerable gender gaps in career advancement.
REFERENCES


APPENDIX A

POLICY CAPTURING SCENARIOS AND
PERFORMANCE RATINGS
Instructions: Please read the following scenarios, answering the set of questions following each scenario. Although the scenarios may seem redundant, each scenario contains unique details. Please pay extra attention to these subtle differences, and answer the following questions appropriately.

There are no right or wrong answers to the questions - we are interested in your personal opinion.

Please use the following scale to answer questions about each scenario.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>To little or no extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. To what extent is this person an example of an ideal employee?
2. To what extent would you consider this description to be an example of excellent overall performance?
3. To what extent would you recommend rewarding this employee with a promotion?
4. To what extent would you recommend rewarding this employee with an increase in salary?

Scenario #1:
Christopher has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “below average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Christopher often speaks up with ideas for new projects or changes in procedure. In addition, Christopher often helps his fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #2:
Christina has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “below average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Christina often speaks up with ideas for new projects or changes in procedure. In addition, Christina often helps her fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #3:
Michael has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “above average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording
information such as charges and the destination for each item. His reviews also note that Michael often speaks up with ideas for new projects or changes in procedure. In addition, Michael rarely helps his fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #4:
Michelle has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “above average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Michelle often speaks up with ideas for new projects or changes in procedure. In addition, Michelle rarely helps her fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #5:
Joshua has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “below average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Joshua often speaks up with ideas for new projects or changes in procedure. In addition, Joshua rarely helps his fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #6:
Jessica has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “below average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Jessica often speaks up with ideas for new projects or changes in procedure. In addition, Jessica rarely helps her fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #7:
Nicholas has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “below average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Nicholas rarely speaks up with ideas for new projects or changes in procedure. In addition, Nicholas rarely helps his fellow coworkers complete their mail sorting duties if they fall behind.
Scenario #8:
Nicole has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “below average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Nicole rarely speaks up with ideas for new projects or changes in procedure. In addition, Nicole rarely helps her fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #9:
Andrew has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “below average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Andrew rarely speaks up with ideas for new projects or changes in procedure. In addition, Andrew often helps his fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #10:
Amanda has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “below average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Amanda rarely speaks up with ideas for new projects or changes in procedure. In addition, Amanda often helps her fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #11:
Brandon has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “above average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Brandon rarely speaks up with ideas for new projects or changes in procedure. In addition, Brandon often helps his fellow coworkers complete their mail sorting duties if they fall behind.

Scenario #12:
Brittany has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “above average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that
Brittany rarely speaks up with ideas for new projects or changes in procedure. In addition, Brittany often helps her fellow coworkers complete their mail sorting duties if they fall behind.

**Scenario #13:**
David has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “above average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that David rarely speaks up with ideas for new projects or changes in procedure. In addition, David rarely helps his fellow coworkers complete their mail sorting duties if they fall behind.

**Scenario #14:**
Diana has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “above average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Diana rarely speaks up with ideas for new projects or changes in procedure. In addition, Diana rarely helps her fellow coworkers complete their mail sorting duties if they fall behind.

**Scenario #15:**
Sean has been a mail clerk at his hometown Post Office for over five years. According to his recent performance reviews, his supervisors rated him as “above average” on his daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. His reviews also note that Sean often speaks up with ideas for new projects or changes in procedure. In addition, Sean often helps his fellow coworkers complete their mail sorting duties if they fall behind.

**Scenario #16:**
Sarah has been a mail clerk at her hometown Post Office for over five years. According to her recent performance reviews, her supervisors rated her as “above average” on her daily tasks, which include sorting through letters and packages to direct them to the correct bin, affixing postage, weighing packages, inspecting the wrapping of packages, and recording information such as charges and the destination for each item. Her reviews also note that Sarah often speaks up with ideas for new projects or changes in procedure. In addition, Sarah often helps her fellow coworkers complete their mail sorting duties if they fall behind.
APPENDIX B

GENDER PERSONAL IDENTITY
Instructions: Read each of the following statements and decide how much you agree with each according to your attitudes, beliefs, and experiences. It is important for you to realize that there are no “right” or “wrong” answers to these questions. People are different, and we are interested in how you feel. Please indicate the appropriate response on the following 5-point scale:

<table>
<thead>
<tr>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Neither Agree nor Disagree</th>
<th>4 Agree</th>
<th>5 Strongly Agree</th>
</tr>
</thead>
</table>

1. In general, my gender is an important part of my self-image.
2. My gender is unimportant to my sense of what kind of a person I am.
3. Overall, my gender has little to do with how I feel about myself.
4. My gender is an important reflection of who I am.
APPENDIX C

GENDER SOCIAL IDENTITY
Instructions: Read each of the following statements and decide how much you agree with each according to your attitudes, beliefs, and experiences. It is important for you to realize that there are no “right” or “wrong” answers to these questions. People are different, and we are interested in how you feel. Please indicate the appropriate response on the following 5-point scale:

<table>
<thead>
<tr>
<th>1</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Disagree</td>
</tr>
<tr>
<td>3</td>
<td>Neither Agree nor Disagree</td>
</tr>
<tr>
<td>4</td>
<td>Agree</td>
</tr>
<tr>
<td>5</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. When others of my gender are recognized for their accomplishments, I feel like I have accomplished something too.
2. I share in the successes of others of my gender.
3. I think about my gender more when someone of my gender is commended on his or her work.
APPENDIX D

GENDER IDENTITY SALIENCE
Instructions: Read each of the following statements and decide how much you agree with each. It is important for you to realize that there are no “right” or “wrong” answers to these questions. People are different, and we are interested in how you feel. Please indicate the appropriate response on the following 5-point scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. When people ask me about who I work with, I initially think of describing my coworkers in terms of gender composition (e.g., two women and three men).
2. It is not intentional, but when I think of my coworkers, what comes to mind initially are the names of individuals of the same gender as me.
3. Even though I don’t mean to, I think of gender as the most prominent characteristic of my coworkers.
APPENDIX E

MODERN SEXISM SCALE
Instructions: Read each of the following statements and decide how much you agree with each according to your attitudes, beliefs, and experiences. It is important for you to realize that there are no “right” or “wrong” answers to these questions. People are different, and we are interested in how you feel. Please indicate the appropriate response on the following 5-point scale:

<table>
<thead>
<tr>
<th>1 Strongly Disagree</th>
<th>2 Disagree</th>
<th>3 Neither Agree nor Disagree</th>
<th>4 Agree</th>
<th>5 Strongly Agree</th>
</tr>
</thead>
</table>

1. Discrimination against women is no longer a problem in the United States.
2. Women often miss out on good jobs due to sexual discrimination.
3. It is rare to see women treated in a sexist manner on television.
4. On average, people in our society treat husbands and wives equally.
5. Society has reached a point where women and men have equal opportunities for achievement.
6. It is easy to understand the anger of women’s groups in America.
7. It is easy to understand why women’s groups are still concerned about societal limitations of women’s opportunities.
8. Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women’s actual experiences.
APPENDIX F

OCB EXPECTATIONS BY GENDER – SINGLE ITEM MEASURE
**Instructions:** Think about the expectations of behavior for men and women *in your workplace*. Then respond to each of the following behavior categories according to whether you think that particular behavior is more expected of men or of women. Please note that there are no right or wrong answers to these questions – we are interested in your personal opinion. In addition, please note that you should be thinking about *what actually happens* in your workplace -- not what should happen.

*Helping behavior*
Voluntary acts of helping others with work-related problems. Examples include orienting new people, assisting someone with a heavy workload, helping those who have been absent, and encouraging fellow employees.

In my workplace, these behaviors are…

<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>Much More Expected of Women</td>
<td>More Expected of Women</td>
<td>Somewhat More Expected of Women</td>
<td>Equally Expected of Men and Women</td>
<td>Somewhat More Expected of Men</td>
<td>More Expected of Men</td>
<td>Much More Expected of Men</td>
</tr>
</tbody>
</table>

*Voice behavior*
Responsible, constructive involvement in the political process of the organization. Examples include engaging in relevant policy debates, speaking up with ideas for new projects, and expressing one’s opinion about organizational strategy.

In my workplace, these behaviors are…

<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>Much More Expected of Women</td>
<td>More Expected of Women</td>
<td>Somewhat More Expected of Women</td>
<td>Equally Expected of Men and Women</td>
<td>Somewhat More Expected of Men</td>
<td>More Expected of Men</td>
<td>Much More Expected of Men</td>
</tr>
</tbody>
</table>
APPENDIX G

OCB EXPECTATIONS BY GENDER – MULTIPLE ITEM MEASURE
Instructions: Think about the expectations of behavior for men and women in your workplace. Then respond to each of the following behavior categories according to whether you think that particular behavior is more expected of men or of women. Please note that there are no right or wrong answers to these questions – we are interested in your personal opinion. In addition, please note that you should be thinking about what actually happens in your workplace -- not what should happen.

Please indicate the appropriate response on the following 7-point scale:

<table>
<thead>
<tr>
<th></th>
<th>1 Much More Expected of Women</th>
<th>2 More Expected of Women</th>
<th>3 Somewhat More Expected of Women</th>
<th>4 Equally Expected of Men and Women</th>
<th>5 Somewhat More Expected of Men</th>
<th>6 More Expected of Men</th>
<th>7 Much More Expected of Men</th>
</tr>
</thead>
</table>

1. Develops and makes recommendations concerning issues that affect the work group.
2. Helps others who have been absent.
3. Speaks up and encourages others in the work group to get involved in issues that affect the group.
4. Helps others who have heavy workloads.
5. Communicates his/her opinions about work issues to others in the group even if his/her opinion is different and others in the group disagree with him/her.
6. Helps orient new people even though it is not required.
7. Keeps well informed about issues where his/her opinion might be useful to the work group.
8. Willingly helps others who have work-related problems.
9. Gets involved in issues that affect the quality of work life here in this group.
10. Is always ready to lend a helping hand to those around him/her.
11. Speaks up in this group with ideas for new projects or changes in procedures.
APPENDIX H

OVERALL PERFORMANCE: SOLTANI
(adapted from Soltani et al., 2013)

Instructions: For the following questions, please think of a specific, individual (FE)MALE peer at your current job. Please rate the extent to which each item reflects the behavior of that (fe)male peer at your current job.

Note: Please remember the (fe)male peer who you are rating in this section. We will be asking you more about her/him later on.

<table>
<thead>
<tr>
<th></th>
<th>To little or no extent</th>
<th>To some extent</th>
<th>To a great extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. To what extent is (s)he an example of an ideal employee?
2. To what extent would you describe her/his behavior as an example of excellent overall performance?
3. To what extent would you recommend rewarding her/him with a promotion?
4. To what extent would you recommend rewarding her/him with an increase in salary?
APPENDIX I

OVERALL PERFORMANCE: WILLIAMS & ANDERSON
(adapted from Williams & Anderson, 1991)

Instructions: Read each of the following statements regarding the specific, individual (FE)MALE coworker you just rated and rate how frequently that (fe)male coworker engages in each behavior.

<table>
<thead>
<tr>
<th>1 Never</th>
<th>2 Rarely</th>
<th>3 Sometimes</th>
<th>4 Often</th>
<th>5 Always</th>
</tr>
</thead>
</table>

1. (S)he adequately completes her assigned duties
2. (S)he fulfills responsibilities specified in her job description.
3. (S)he performs tasks that are expected of her/him.
4. (S)he meets formal performance requirements of the job.
5. (S)he engages in activities that will directly affect her/his performance evaluation.
6. (S)he neglects aspects of the job s(he) is obligated to perform.
(S)he fails to perform her/his essential duties.
APPENDIX J

SELF-RATINGS OF OCB
**Instructions:** Please use the following scale to rate the extent to which each item reflects your behavior on your current job:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>To a very small extent</td>
<td>To a limited extent</td>
<td>To some extent</td>
<td>To a considerable extent</td>
<td>To a great extent</td>
</tr>
</tbody>
</table>

1. I develop and make recommendations concerning issues that affect the work group.
2. I help others who have been absent.
3. I speak up and encourage others in the work group to get involved in issues that affect the group.
4. I help others who have heavy workloads.
5. I communicate my opinions about work issues to others in the group even if my opinion is different and others in the group disagree with me.
6. I help orient new people even though it is not required.
7. I keep well informed about issues where my opinion might be useful to the work group.
8. I willingly help others who have work-related problems.
9. I get involved in issues that affect the quality of work life here in this group.
10. I am always ready to lend a helping hand to those around me.
11. I speak up in this group with ideas for new projects or changes in procedures.
APPENDIX K

OTHER-RATINGS OF OCB
Instructions: Think of an individual male (or female) peer at your current job. Please use the following scale to rate the extent to which each item reflects the behavior of that male (or female) peer at your current job:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>To a very small extent</td>
<td>To a limited extent</td>
<td>To some extent</td>
<td>To a considerable extent</td>
<td>To a great extent</td>
</tr>
</tbody>
</table>

1. He (she) develops and makes recommendations concerning issues that affect the work group.
2. He (she) helps others who have been absent.
3. He (she) speaks up and encourages others in the work group to get involved in issues that affect the group.
4. He (she) helps others who have heavy workloads.
5. He (she) communicates his (her) opinions about work issues to others in the group even if his (her) opinion is different and others in the group disagree with him (her).
6. He (she) helps orient new people even though it is not required.
7. He (she) keeps well informed about issues where his (her) opinion might be useful to the work group.
8. He (she) willingly helps others who have work-related problems.
9. He (she) gets involved in issues that affect the quality of work life here in this group.
10. He (she) is always ready to lend a helping hand to those around him (her).
11. He (she) speaks up in this group with ideas for new projects or changes in procedures.
APPENDIX L

JOB INTERDEPENDENCE
**Instructions:** Read each of the following statements about your coworker and decide how much you agree with each. Please indicate the appropriate response on the following 5-point scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. The tasks related to my job can be performed fairly independently of this person.
2. The tasks related to my job can be planned with little need to coordinate with this person.
3. It is rarely required to obtain information from this person to complete the tasks related to my job.
4. The tasks related to my job are relatively unaffected by the performance of this person.
5. The tasks related to my job require frequent coordination with the efforts of this person.
6. The tasks related to my job are dependent on receiving accurate information from this person.