

Does Gender Matter in the Evaluation of Successful Physicians? Examining How Evaluators Use
Stereotype-Based Attributions in Determining Outcomes at Work

Dyan Angela Ludeña Ferraris-Baron

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

©2017
Dyan Angela Ludeña Ferraris-Baron
All rights reserved

ABSTRACT

Does Gender Matter in the Evaluation of Successful Physicians? Examining How Evaluators Use Stereotype-Based Attributions in Determining Outcomes at Work

Dyan Angela Ludeña Ferraris-Baron

The purpose of this study was to understand whether physician leaders are subject to gender bias in the form of differential work outcomes. Specifically, the primary goal was to examine whether the gender of a successful physician leader, the medical specialty in which he or she works (surgery or pediatrics), and participant level of social dominance orientation (SDO; level of egalitarianism) influenced the allocation of workplace outcomes (i.e., evaluations of performance, promotion recommendations, and characterizations of ability and effort). This study further explored if ability and effort characterizations mediated the relationship between gender, specialty, and participant SDO on evaluations of performance and promotion recommendations. Attempting to elucidate the atypical but increasing phenomenon in which successful female leaders in male-typed jobs receive higher performance evaluations, but lower rates of promotion as compared to equivalent males; this study drew on attribution theory to explain that characterizations of successful women as “hard workers” (effort) may be seen as deserving of high evaluations of performance but not promotions, while being “brilliant” (ability) may be seen as deserving of promotions and reserved for successful men. Results revealed an unexpected overall boost for female surgeons, awarded especially by participants low in SDO (those most egalitarian) such that female surgeons received significantly better outcomes as compared to female pediatricians and equivalent outcomes as compared to male physicians. Male surgeons and pediatricians were largely awarded equivalent outcomes across all levels of participant SDO. Further, mediation was supported only for female surgeons, such that higher characterizations of effort explained higher evaluations of performance, particularly by

those low in SDO. Further research is required to understand why successful women receive higher evaluations of performance, but not promotions.

TABLE OF CONTENTS

LIST OF TABLES	iii
LIST OF FIGURES	iv
ACKNOWLEDGEMENTS	v
CHAPTER ONE: Introduction	1
CHAPTER TWO: Literature Review	13
Leader Gender and Outcomes at Work	13
Explanations for Success: Attribution-based Gender Bias	32
Participant Social Dominance Orientation as a Moderator	44
Effort and Ability Attributions as Mediators	46
CHAPTER THREE: Method	48
Participants	48
Research Strategy & Design	48
Procedure	49
Experimental Manipulations	53
Measures	54
CHAPTER FOUR: Results	60
Manipulation Checks	60
Preliminary Analysis	61
Main Analysis	62
Mediation Analysis	76
Summary of Findings	80

CHAPTER FIVE: Discussion	83
Overview	83
Discussion and Interpretation of Results	84
Limitations and Future Directions	97
Implications for Theory	104
Implications for Practice	106
Conclusion	110
REFERENCES	112
APPENDICES	
Appendix A. <i>Cover Story, Informed Consent, & Participants Rights</i>	157
Appendix B. <i>Study Instructions</i>	162
Appendix C. <i>Stimulus Materials: Background Information</i>	163
Appendix D. <i>Stimulus Materials: Dean’s Letter</i>	165
Appendix E. <i>Stimulus Materials: Summary Performance Evaluation</i>	167
Appendix F. <i>Questionnaire Instructions</i>	171
Appendix G. <i>Study Debrief</i>	172
Appendix H. <i>Study Measures – Promotion & Performance</i>	175
Appendix I. <i>Study Measures – Ability & Effort</i>	176
Appendix J. <i>Social Dominance Orientation Measure & Distribution of SDO Scores by Tertiary Split Category</i>	177
Appendix K. <i>Manipulation Check Questions</i>	179
Appendix L. <i>Demographic Questionnaire</i>	180
Appendix M. <i>Additional Results</i>	183

LIST OF TABLES

Table 1. <i>Number and Percentage of Active Physicians by Sex and Specialty: 2014</i>	128
Table 2. <i>Means, Standard Deviations, and Intercorrelations for All Study Variables</i>	130
Table 3. <i>Means and Standard Deviations for Outcome Measures by Condition</i>	131
Table 4. <i>Means and Standard Deviations for Outcome Measures by Condition and Participant SDO</i>	132
Table 5. <i>Summary of Research Questions, Hypotheses, and Findings</i>	133
Table 6. <i>Univariate ANCOVA Results for Study Predictors on Promotion</i>	137
Table 7. <i>Univariate ANCOVA Results for Study Predictors on Performance</i>	138
Table 8. <i>Univariate ANOVA Results for Study Predictors on Ability Characterizations</i>	139
Table 9. <i>Univariate ANCOVA Results for Study Predictors on Effort Characterizations</i>	140
Table 10. <i>Results of Hierarchical Regression of Effort and Ability Characterizations on Promotion</i>	141
Table 11. <i>Results of Hierarchical Regression of Effort and Ability Characterizations on Performance</i>	142
Table 12. <i>Mediational Analyses: Effort Attributions as a Mediator</i>	143
Table 13. <i>Mediational Analyses: Ability Attributions as a Mediator</i>	144

LIST OF FIGURES

Figure 1. <i>Theoretical Process Map of the Central Purpose of this Study</i>	145
Figure 2. <i>Model of Moderation of Specialty to Gender on Promotion and Performance</i>	146
Figure 3. <i>Attributional Explanations Categorized by Attribution Dimension</i>	147
Figure 4. <i>Moderation of Specialty to Gender on Ability and Effort</i>	148
Figure 5. <i>The Impact of Success Attributions on Employee Outcomes</i>	149
Figure 6. <i>Full Mediated Moderation of Specialty Congruency and Success Attribution on Work Outcomes</i>	150
Figure 7. <i>Graphs of Significant Three-way Interaction on Performance by SDO Level</i>	151
Figure 8. <i>Graphs of Significant Three-way Interaction on Performance by Target Gender</i>	152
Figure 9. <i>Graphs of Significant Three-way Interaction on Attribution Characteristics by SDO Level</i>	153
Figure 10. <i>Graphs of Significant Three-way Interaction on Attribution Characteristics by Target Gender</i>	154
Figure 11. <i>Graphs of Significant Three-way Interaction on Effort Characteristics by SDO Level</i>	155
Figure 12. <i>Graphs of Significant Three-way Interaction on Effort Characteristics by Target Gender</i>	156

ACKNOWLEDGEMENTS

I am immensely grateful to all of my family, friends, and mentors who supported me along the way! Thank you for your encouragement and for your patience – and for believing that I am capable of achieving these goals. First, I'd like express my profound gratitude to my mentors Caryn Block, Debra Noumair, and Elissa Perry – the ultimate trifecta! Caryn, I came to the program to learn specifically from you, as your research and scholarship lit (and continues to light) a fire in my intellectual and personal heart and helped me realize where I'd like to contribute in this world. Thank you for being my dissertation sponsor, my scholarly inspiration, my compassionate mentor, and for your dedication to helping me become a better scholar and researcher. Debra, thank you for teaching, guiding, and showing me how to be the best scholar-practitioner I can be. I admire you, your work, and your way of working deeply. You encouraged me to bring my whole self to the work, which was not easy, but my work is that much more authentically me. You also had the unique 6th sense to say the thing I needed to hear most at exactly the time I needed to hear it. So, my blood pressure thanks you too. Elissa, your high standards and rigorous methods molded me into a scientist who has immense theoretical knowledge and data analysis skills. Thank you for all of the time and dedication this type of training takes. I looked to you for that standard, so once I got it, I knew I was good to go!

I am also very thankful for the friendship, guidance and training I received from members of Social-Organizational Psychology Program community. Particularly, my professors Loriann Roberson, Sarah Brazaitis, and Matthew Tye – your teaching and insight during my earlier years impacted me greatly and set me up well to be where I am now. And thanks to my dear friends and colleagues I know through the program: Mekayla Castro, Naomi Stutzman, Chris St. John, Joaquin Roca, Ambar Ureña, Paul Hanvongse, Regina Kim, Joe Dillard, and Marissa Magno.

A huge and heartfelt thank you to my very best friends. First to those who helped me along my doctoral program as they went through their own: Mateo Cruz, Jennifer Wallach, and Faith Kares. Mateo, my partner in crime in the cohort – thank you for keeping me sane and also providing some insane times when we needed some derailment. I am in awe of your talent and am deeply thankful to have you as a fellow scholar and close friend. Jenny, my roomie for life – thank you for all the laughs, support, brainstorming sessions, and for sharing that special slice of pizza with me. I looked to you to see how this doctoral thing and life thing is done, and see you as one of my most beloved role models and friends. Faith, my Lil One, almost 20 years ago, the world brought us together and since then you have accomplished so much as a scholar and now a scholar-practitioner! Thank you for being a best friend as well as an inspiration – you and your work move me every single day. Lastly, a huge thank you to my oldest and bestest friends Gwen Salazar and Stephanie Kim – you two are my rocks and have been for decades! You both are unbelievably strong and amazing women who know how to work hard and play hard. Thank you for sticking by me as we’ve tried to figure out life together, I have so much love for you.

Lastly, salamat kaayo to my loving family. First to my husband, Gerard Baron. Thank you for supporting and loving me the best you know how. You fill my heart and life with so much – let’s continue holding hands throughout it all. To my brothers, Mark and Blase Ferraris – thank you for putting up with me and showing me the loving and constructive aspects to sibling rivalries. I know you both will always have my back – I will always have yours. To my parents, Rony and Joy Ferraris, my dissertation is wholeheartedly dedicated to you. I cannot thank you enough for your selfless and never-ending dedication to me and my education. You both have rich stories full of love and sacrifice – I am lucky to share in some of these stories and look forward to how our stories continue to unfold together. All of my success is because of you.

CHAPTER ONE: INTRODUCTION

Consider a well-known riddle, “A patient is brought into the emergency room, and the surgeon says, “I cannot operate on this patient: he is my son.” The surgeon is not the patient’s father. Why can’t the surgeon operate?” (Williams, 2009, p. 15). The explanation is that the surgeon is the patient’s mother. Gender-stereotyped based assumptions such as this (i.e., the notion that surgeons are male) persist even though women have made tremendous progress in workforce participation. It is estimated that 75% of women between the ages of 25 and 55 are currently employed, this tremendous increase occurring predominantly in the last three decades (Fassinger, 2008; Fitzgerald & Harmon, 2001). Despite these gains, women remain alarmingly underrepresented in traditionally male occupations, receive inequities in compensation, and are rarely found at the leadership levels of organizations, especially at the very top (Catalyst, 2010). Unfortunately for those exceptional women who do make it to the higher levels, there is an overwhelming amount of evidence that suggests that partaking in leadership roles comes at a cost (i.e., lower rates of promotions as compared to men), especially for women in male-dominated fields.

These disparities are mirrored in the field of medicine. As such, patterns of gender participation within medicine, specifically physician leadership is of primary focus in this study. Overall and across medical specialties, the pattern of gender participation is encouraging and seemingly balanced. Some scholars even say there is a “feminizing” of medical education, as women currently comprise almost half of enrolled students in medical schools (47%) and of residencies (46%) (Akers, van Leerdamn, Dielissen, & Lagro-Janssen, 2014; Association of American Medical Colleges, 2014). Women have also been entering academic medicine as faculty members in numbers equal to their male counterparts for several decades (Ash, Carr, &

Goldstein, 2004). Further, the percentage of female faculty at medical schools in the United States has doubled in just three decades, increasing from 13% in 1967 to 27% in 1992 (Bickel, Wara, & Atinson, 2002). Currently, 29% of practicing physicians in the United States are women, and they make up 38% of full-time faculty members of medical schools (Association of American Medical Colleges, 2014). Although these numbers are not quite balanced, the shift is moving in the direction of equal representation.

While more women are entering the field of medicine as compared to only a few decades ago, they are starkly underrepresented in leadership, high-paying, and prestigious roles. Most noticeably, there is a *vertical segregation* within medicine, where disproportionately few women occupy senior positions (Kilminster, Downes, Gough, Murdoch-Eaton, & Roberts, 2007). In academic medicine, female physicians make up only 15% of department chairs, 16% of deans, and 21% of full professors (Association of American Medical Colleges, 2014; Dacre, 2008; Lautenberger, Dandar, & Raezer, 2014). With regards to leadership outside of academic medicine, a report released by the American College of Physician Executives (2009) supports that physician executive leadership positions in hospitals are male-dominated, estimating that men compose 90% of chief medical officers, 90% vice president of medical affairs, and 92% chief executive officer/president (Dister, 2009).

In addition to vertical segregation, there is also evidence of *horizontal segregation* across specialties (Alers et al., 2014). For example, according to the American Association of Medical Colleges (2014) women are overwhelmingly under-represented in surgical specialties (e.g., orthopaedic surgery, 4%; neurological surgery, 7%) and relatively equally represented in primary care specialties (e.g., pediatrics, 60%; obstetrics and gynecology, 52%) (Williams et al., 2013). See Table 1 for number and percentage of active physicians by sex and specialty. Even as more

women enter the occupation, gender segregation among specialties remains constant. The proportion of specialty dissimilarity based on gender has been about the same since 1985 (Association of American Medical Colleges, 2014).

Further, there is a stark pay gap between what male and female physicians earn. Women physicians suffer the largest gender pay gap of any of the professions, earning only \$0.62 for every \$1.00 a male physician earns (US Census Bureau, 2010). Female physicians and surgeons make 79% of what their male colleagues earn; and even though women are the majority of pediatricians, they earn 66% of what their male counterparts earn (Boulis & Jacobs, 2010; Bureau of Labor Statistics, 2011). Reflecting the pattern of horizontal segregation, women are disproportionately overrepresented in lower paying specialties. For example, pediatrics, one of the only specialties dominated by female physicians, is also one of the lowest paying specialties (Williams et al., 2013). The income disparity between male and female physicians remains even when controlling for age, education, specialty, and hours worked. This remaining income gap and paucity of women in leadership positions is reportedly not fully understood and even “perplexing” to medical scholars and practitioners (Darvies, 2012; Langton, 2008).

These alarming numbers have spurred much inquiry and investigation in the medical field and other male-dominated fields. Although well-intentioned investigations seek to explain why these differences exist along gender lines, many result in perpetuating the issue by adopting a “blame the woman” or “fix the women” approach (Ely, Ibarra, & Kolb, 2011; Ely & Meyerson, 2000). These types of explanations locate the issue in women themselves, rationalizing that “women have not been socialized to compete successfully in the world of men, and so they must be taught the skills their male counterparts have acquired as a matter of course” (Ely et al., 2011, p. 475). While this approach may begin to address the lack of women in certain specialties and

leadership roles, it severely lacks an adequate and complete narrative of the organizational realities women confront.

For example, many “blame the woman” reasons center on the women themselves contributing to a *pipeline problem* that results in a lack of qualified female candidates for these male-dominated positions (Alers et al., 2014; Eagly & Karau, 2002; Zhuge, Kaufman, Simeone, Checn, & Velazquez, 2011). An often-cited reason why women choose not to go into surgery is due to family and lifestyle priorities. Studies report that female students hold the opinion that because of the long and unpredictable work hours, pursuing surgery is incompatible with a satisfying family life, happy marriage, or raising children; and are consequently more likely to choose a caring profession, namely pediatrics (Cancian & Olike, 2000; Drinkwater, Tully, & Dornan, 2008; Novielli, Hojat, Park, Gonnella, & Veloski, 2001).

Another often cited “blame the woman” explanation is that women are somehow deficient in characteristics needed for success in those roles. For example, in a recent article examining why there are so few women in surgical specialties, a physician was quoted saying, “Generally speaking, female physicians don’t do as good a job at self-promotion as their male counterparts do; and many think that if they simply work harder they’ll get ahead by virtue of their qualifications and track record,” (Darves, 2012, p. 6). Additional explanations for lack of advancement of women in medicine include, lower productivity (measured by publications or grants and fewer hours working) and lower ambition. For example Nonnemaker (2000) found that women are “less likely than men to join departments in which the overall probability of promotion is low,” (p. 402) because these departments are seen as more competitive, thus avoided by women. These explanations cite women’s inherent tendencies to display fewer of the traits and characteristics required to obtain and succeed in high-level positions (Alers et al.,

2014; Bickel, 2001; Eagly et al., 2002).

In response to the tremendous evidence of vertical and horizontal gender segregation in medicine, there is currently considerable nationwide energy being devoted to recruiting, retaining, and developing women into and within specialties and positions in which they are a minority (Bickel et al., 2002; Institute of Medicine (US) - Committee on Maximizing, 2007). However, many of these programs may not address the entirety of the problem, because many of them originated as remedies to pipeline and deficiency explanations.

As such, the “blame the woman” reasons may be insufficient given evidence supporting that women do have the characteristics necessary to succeed in male-typed jobs. For example, there is an abundance of leadership and personality scholarship that supports that women and men do not differ in either (1) ambition or desire for advancement and leadership (Konrad, Ritchie, Lieb, & Corrigan, 2000); or (2) the personality traits associated with effective leadership (Halpern, 2001; Judge, Bono, Ilies, & Gerhardt, 2002; Judge, Colbert, & Ilies, 2004; Schmitt, Realo, Voracek, Allik, J., 2008). Further, there is a sizeable amount of human capital evidence (i.e., women hold about half of the jobs and possess half of bachelor’s and advanced degrees in the U.S.) that suggest that women are willing and capable of occupying leadership roles (Eagly et al, 2002; Snyder & Dillow, 2012). What then are other viable reasons to explain the vertical and horizontal barriers for women leaders in medicine? This study examines stereotype-based gender bias as a central explanation to the disparity in workplace outcomes (e.g., promotions and performance ratings) to otherwise equally successful male and female physician leaders.

Social science, and more recently medical, scholars maintain there are unexplored opportunities to gain a deeper understanding and better address inequities in male-dominated fields. These scholars explain that women’s absence and lack of advancement into prestigious,

high-level, and leadership positions (that are also the highest paying) is due to gender bias against women and preference for men. The *glass ceiling* metaphor has been used to explain gender bias in the form of invisible barriers and covert obstacles that inhibit women from ascending the corporate or academic organizational ladder despite increased participation of women in higher education and fields traditionally held by men (Williams & Dempsey, 2014). Further, it has been suggested that the *old boy's club* mentality continues to persist and promote organizational cultures that favor men by excluding women from access to resources such as formal and informal networking and mentoring (Longo & Straechley, 2008). The glass-ceiling and old boy's club phenomena are largely discussed as problematic in traditional male careers, such as finance, law, and business management, however it is also observed in medical professions (Zhuge et al., 2011).

Research supporting the glass ceiling for women in the workplace explains that the economic and social penalties working women experience are largely due to gender bias, and that these biases are rooted in gender *stereotypes*. Stereotypes are culturally based cognitive shortcuts that inform generalizations about groups of people (Macrae, Milne, & Bodenhausen, 1994). The use of stereotypes become problematic when group-level oversimplifications are applied to individual members of the group simply because they belong to that group (Link & Phelan, 2001). Gender-based stereotypes are used to fill in missing pieces or to interpret the behavior of men and women in general (Dunning & Sherman, 1997).

The opening riddle exemplifies how gender stereotypes can provide missing pieces for men and women in the workplace. Specifically, because men typically and historically have occupied the role of surgeon, we expect that surgeons are male. Women who occupy these masculine roles face bias in the form of lowered expectation of performance as compared to

men, because gender stereotypes contribute to a perceived mismatch between a “typical woman” and the requirements of roles typically occupied by men (Heilman, 2001). Empirical evidence supports that this lowered expectation often results in a preference for men and a bias for women in the form of workplace outcomes such as selection for and evaluation, especially within *male-typed* or more stereotypically masculine roles (e.g., surgeon) and less within *female-typed* or more stereotypically feminine roles (e.g., nurse) (Heilman, 1983, 2001). This bias may be amplified for female physician leaders because not only is the field of medicine a stereotypically masculine work domain, but the role of “leader” has also been supported to be stereotypically masculine in nature (Eagly et al., 2002; Heilman, Block, Martell, 1989).

Despite gender stereotype based bias identified by previous scholarship, there are a few studies that begin to surface a different pattern for female leaders in male-typed jobs. For example, in a field study conducted in 2006, researchers found that female leaders in a male-typed job received higher performance evaluation ratings, but lower rates of promotion than males in equivalent positions (Lyness & Heilman, 2006). Further, they found a strong significant relationship between performance ratings and promotion rates for women, but no relationship was found for men. In other words, the study supported that performance ratings mattered for female leaders, but not for male leaders; men received promotions no matter what their performance ratings were (Lyness et al, 2006). Lyness and Heilman (2006) deduced that these female leaders had to work harder than their male counterparts at getting the same outcome (i.e., a promotion).

Eagly and Karau (1992) and more recently, Joshi, Son, and Roh (2015) found a similar pattern in their meta-analyses of studies examining gender and leadership; a pattern which Eagly et al. (1992) described as “somewhat puzzling” (p. 17). The majority of the studies supported a

clear bias for male leaders compared to female leaders with regards to promotion outcomes (e.g., Joshi et al., 2015; Lyness & Judiesch, 1999; Maume, 1999; Williams, 1992, 1995). However, a handful of studies indicated a pattern favoring female leaders over male leaders on performance outcomes, even in male-typed jobs (e.g., Abramson, Goldberg, Greenberg, & Abramson, 1977; Heilman et al., 1988; Joshi et al., 2015; Taynor & Deaux, 1973). Eagly and colleagues (1992) speculate that perceivers ascribed higher levels of performance for these women compared to men because the female leaders were seen as “competent enough to withstand the countervailing pressures from the traditional gender hierarchy,” (p.17). Joshi and colleagues (2015) further speculate that in prestigious contexts like medicine, law, and academia, employment practices serve to maintain social hierarchies that favor men, awarding women with good performance evaluations and men with promotions. They explain that many of these high prestige contexts practice an “up or out” norm for promotions, providing men with promotions and women with a way out (p. 1533).

These studies suggest a reoccurring phenomenon specifically for *successful* females, women who occupy leadership positions. The explanations, however, are not entirely consistent themselves. Lyness and Heilman (2006) explain that pattern is due to women *actually* having to work harder, Eagly and colleagues (1992, 2002) speculate that there is a *perception* that these women are better performers because not many women made it as far as they did, and Joshi and colleagues (2015) attribute this phenomenon to employment practices in highly prestigious *contexts*. These researchers provide pieces of the puzzle; however, the underlying stereotyped-based mechanisms that drive these patterns remain largely unexplored and unexplained.

Additionally, the majority of previous research examining leader gender, evaluation, and the impact on workplace outcomes (i.e., performance and promotion) is limited to leaders within

particular industries (e.g., military, education, and business) and functional areas within organizations (e.g., human resources, finance, production, marketing) (Eagly et al., 2002). Studies that include the healthcare industry, typically examine *female-typed* jobs, jobs that are stereotypically more feminine in nature, such as nursing (e.g., Eagly et al., 2002; Maume, 1999; Ott, 1989; Williams, 1992, 1995). Thus, despite the alarming statistics cited earlier, little is known about how gender stereotypes impact physicians and physician leaders.

This research will contribute to understanding of gender bias in the following four ways. First, I will examine whether *attributions* of success (i.e., explanations of hard work or natural ability) are the underlying stereotyped-based mechanisms that explain the uneven pattern of workplace outcomes for male and female leaders. Attribution theory is a cognitive social psychological theory that focuses on the human inclination to explain why people behave the way they do (Heider, 1959). Specifically, the theory suggests that people engage in a sense-making process to form a judgment about why an event happened or why person acts in a particular way (Heider, 1959; Kelley, 1967; Wiener, 1985). There is considerable evidence that differential attributions for success are made for women and men (Deaux & Emswiller, 1974). For instance, Heilman and Guzzo (1978) found that explanations for the successful performance of men were likely to be attributed to his innate ability, while the most likely attribution for women was due to her hard work, luck, or ease of the task. Like Heilman et al's (1978) work, most of the studies in this body of literature stop at point of attributions, meaning that they do not assess impact of gender and attributions on workplace outcomes (Swim & Sanna, 1996). The current study investigates not only the impact gender has on attributions, but also how certain types of attributions may differentially impact evaluations of performance and allocation of promotion.

Second, the study of the attributional process has implications for understanding the objectivity of performance evaluation of leaders. The process of performance evaluation fundamentally involves the use of socio-cognitive sense making mechanisms that may lead to inaccurate interpretations about the reasons, sources, and causes of behavior. In a performance evaluation process, a supervisor or evaluator will attempt to discern the cause of the target's work performance in order to determine the appropriate action (i.e., recommendation for promotion or demotion) following the evaluation. However, the human tendency to discern and explain behavior does not necessarily mean that the correct explanation has been discovered, especially if the sense-making process is influenced by gender stereotypes (Gedeon & Rubin, 1999; Heilman, 1983). Further, acting on stereotype-based explanations may have substantial negative impact on the targets of these judgments. The relevance of attribution theory in performance evaluation is critical. The current study will contribute to these practical implications on the accuracy of performance evaluation processes at work by linking gender stereotypes, causal attribution theory, and an evaluation situation.

Third, this study will focus specifically on physician leaders in a promotion and evaluation scenario. The author is aware of only one article, written more than 40 years ago, that examines gender stereotypes and attributions of physicians. Feldman-Summers & Kiesler (1974) found that success on a masculine task (performing as a physician) was attributed to higher levels of motivation or effort for the female physician and higher levels of ability for the male physician. Although the researchers did not assess subsequent impact on workplace outcomes (i.e., performance ratings or promotion recommendations), they did explore two levels of job-type: pediatricians and surgeons. Feldman-Summers and colleagues (1974), however, found no interaction effects of physician gender and specialty; results were the same across both

specialties. Despite the non-significant interaction effects (physician gender and specialty), the current study will explore male and female physician leaders in both pediatrics and surgical specialties.

The fourth contribution of this study is that it takes into account the traits of the evaluators themselves by examining participant *social dominance orientation* as a possible moderator. Social dominance orientation (SDO) is considered a personality trait, defined by social-psychologists as a measure of a person's preference to preserve social hierarchy such that higher-status groups maintain domination over lower-status groups (Sidanius & Pratto, 2001). With regards to gender as a social-identity group, those high in SDO are anti-egalitarian and see men as more dominant than women, and thus have a strong preference for men having greater access to power and resources than women (Sidanius et al., 2001). Previous research found that for equivalent men and women in a male-typed job, participants low in SDO demonstrated a "pro-female" bias while participants high in SDO demonstrated an "anti-female" bias (Hoyt, 2012). Researchers explain that especially in highly masculine contests such as law or academia, organizational structures maintain the social dominance hierarchy by favoring men over women in the allocation of organizational rewards such as for promotion practices (Joshi et al., 2015; Sinclair, Sidanius, Levine, 1998). Some scholars assert that an evaluator's SDO can strongly influence the typical "gender-matching heuristic that matches the gender-type of the applicant with that of the job," (Pratto et al., 1997, p. 47). Because the current study focuses on a promotion scenario in a highly masculine context (medicine), we include participant SDO as a possible predictor of bias and workplace outcomes.

The central purpose of this study is to examine the specific case of *exceptional women*, that is successful women in a male typed job. In this case, a physician with a leadership role. See

Figure 1 for a visual depiction and theoretical process map of the following case. As previously stated, it is supported that for a male-typed job (e.g., physician leader) male employees will achieve better outcomes (e.g., promotions) because men are assumed and expected to have attributes necessary to be successful on the job (Heilman, 1983). However, there is an emerging and “puzzling” phenomenon; some studies found that successful women receive higher performance ratings, but lower rates of promotion than men; while successful men receive higher rates of promotion as compared to women no matter what their performance ratings are (Eagly et al., 2002; Eagly et al., 1994; Joshi et al., 2015; Lyness et al., 2006).

The current study proposes a novel explanation to why this phenomenon occurs by drawing on attribution theory. Attributions help us explain why an event occurred (e.g., she performs well because she tries really hard) (Heider, 1959). This study proposes that gender-based attributions of success are the underlying socio-cognitive mechanisms that influence the disparate allocation of workplace outcomes. We propose that attributions of ability will be associated with successful men, and thus rewarded with better promotion recommendations than women. Further, we propose that attributions of effort will be associated with successful women, and thus rewarded with better performance ratings than men. Taken together, theories concerning gender-stereotyping (Heilman, 1983; Eagly et al., 1997) and assigning gender-based attributions in organizations (Heider, 1959; Swim et al, 1996) lay the foundation to uncovering and explaining what happens to exceptional women in male-typed jobs such as physician leadership.

CHAPTER TWO: LITERATURE REVIEW

Leader Gender and Outcomes at Work

Descriptive Gender Stereotypes. The social cognitive process of stereotyping has been used by scholars across disciplines to explain how perceivers form impressions of a person (Fiske & Taylor, 1991). Stereotyping is a cognitive tool a perceiver uses for two main functions: (1) to organize information about different groups of people; and (2) to generate general expectations about newly encountered members of those groups (Kulik & Bainbridge, 2006). Stereotypes serve as cognitive shortcuts, allowing perceivers to quickly form impressions in order to make sense of the world and more easily respond to their surroundings (Macrae, Milne, & Bodenhausen, 1994). Furthermore, stereotypes can unconsciously influence the perceiver because they can operate without awareness. Studies support that the activation of stereotypes is automatic when encountering a stereotyped group member, though they are not always acted upon (Devine, 1989; Heilman, 2012). The use of stereotypes become problematic when group-level oversimplifications are applied to individual members of the group simply because they belong to that group (Link & Phelan, 2001).

Gender stereotypes that dictate what women and men *can do* are called descriptive stereotypes. These types of stereotypes refer to beliefs about how women and men typically are (Heilman, 1983, 2001). These beliefs ascribe communal traits to women; outlining that “women in general” are affectionate, kind, warm, helpful, nurturing, and interpersonally sensitive. Conversely, these beliefs assign agentic traits to men; outlining that “men in general” are assertive, confident, ambitious, dominant, ambitious, and prone to act like a leader (Burgess & Borgida, 1999; Eagly et al, 2002; Fiske et al., 1991; Heilman, Block, Martell, & Simon, 1989). Gender stereotypes are formed largely from the compounded observation of individuals in sex-

typical roles in society, particularly “women’s in general” occupancy of homemaker and lower status roles and “men’s in general” occupancy of high status breadwinner roles (Cuddy, Fiske, & Glick, 2002; Heilman et al., 1989). This normative conception of how men and women typically are is important because gender roles are pervasive and surreptitiously impact society as a whole. In fact, social cognitive research supports that an individual’s gender provides the strongest source of stereotypic categorization, over and above age and race (Eagly et al., 2002). These strongly held beliefs about how men and women are naturally extend into the microcosm of the workplace.

Sex-typing of Jobs. Gender stereotypes apply to men and women as well as the jobs they occupy. Two main complementary theoretical approaches explain why and how some jobs become typed as female or male: (1) a compositional demography approach (Cejka & Eagly, 1999; Krefting, Berger, & Wallace, 1978); and (2) a cognitive-based approach (Heilman, 1983, 1995). The compositional demography framework suggests that a job becomes sex-typed based on which gender dominates that job. This compositional approach suggests that the actual distribution of persons in the job determines the sex-type of that job (Cleveland & Hollmann, 1990; Nieva & Gutek, 198). Accordingly, we come to associate a job with the gender of the typical job-holder (Lyness et al., 2006). For example, the job of “physician” is male-typed because most physicians are men.

The cognitive-based approach complements this compositional explanation by proposing that gender stereotypes associated to a job determine the sex-based typing of that job. Specifically, particular elements of a job (e.g., job tasks) are cognitively linked to specific gender stereotypes (Cleveland et al., 1990; Lyness et al., 2006). For example, male-typed jobs, including executive roles and management positions, are believed to require characteristics congruent with

stereotypically male attributes and behaviors (i.e., agentic and competitive). Further, specific elements of a job (e.g., job responsibilities) also inform perceptions of job worth, associating high status and prestige with particular jobs. Ragins et al. (1989), explain that managerial jobs, especially at executive levels, are traditionally considered to be male-typed because of the high levels of authority and responsibility, while female-typed jobs are considered less prestigious. For example, stereotypical male-typed jobs include administrator, doctor, lawyer, and politician; and stereotypical female-typed jobs include elementary school teacher, nurse, secretary, and social worker (Ragins et al., 1989).

There is also evidence that the notion of sex-typing can apply to arenas more encompassing than just the role itself. In support of applying sex-typing from level of roles to the higher level of industry, Heilman (2012) stated, “Maleness is determined not only by the job itself, but by occupation (e.g., the military vs. education), subfields or professional specialties (e.g., surgery vs. pediatrics), academic fields (e.g., sciences vs. humanities), and function and level within an organization,” (p.118). In each of her examples, Heilman provides a highly male-typed arena (i.e., military, surgery, sciences) to a highly female-typed arena (i.e., education, pediatrics, humanities). Scholars further explain that not only does male-typing occur when work responsibilities are outlined with traits typically associated with men (Gaucher et al., 2011), but also that male-typing is highly likely when men constitute the overwhelming majority proportionally (Cejka & Eagly, 1999; Eagly et al., 1995); the latter explanation overlapping with the aforementioned compositional demography approach. The sex-typing of jobs, industries, occupations, and subfields has been shown to negatively impact (i.e., lower performance ratings, lower rates of promotion) the people in those jobs based on his or her gender.

Lack of Fit & Role Congruity Theories. The lack of fit model (Heilman, 1983; Heilman

& Parks-Stamm, 2007) and role congruity theory (Eagly et al., 2002) explain that bias occurs when there is a perceived incongruence between the gender of an applicant or employee and the sex-type of a job. Specifically, the expectations about how successful an individual will be at a job are “determined by the fit between the perception of an individual’s attributes and the perception of the job’s requirements in terms of skills and abilities” (Heilman, 1983, p. 278). According to these theories, women are believed to be unfit for male-typed jobs because of the incongruity between the requirements for success in the job (i.e., masculine attributes) and the attributes women stereotypically are deemed to possess (i.e., feminine attributes). Leadership and management jobs are strongly male-typed because the norm-based necessities of the role require stereotypically masculine traits; this is supported widely as the “think manager think male” paradigm (Heilman et al., 1989). Based on this reasoning, men are considered a good fit, perceived as capable for leadership, and expected to be successful in a leadership job, due to a high degree of congruence between the male gender role and a leader role. Thus, if a job requires masculine skills and abilities, then a woman would be evaluated as ill-suited for the job due to the incongruity, and expectations of failure would follow.

These expectations of failure are problematic for women seeking upward mobility into leadership roles. Heilman (1983) emphasized that expectations of failure may prevent women from being perceived as qualified for or successful at a job, and consequently are often passed over in employment decisions, especially for leadership roles. Heilman (1983) refers to biases during the hiring process as “pre-entry discrimination,” (p. 280) because women are perceived as incapable for a male-type job and are not hired even before entry into the organization. This notion can be used to explain the horizontal barriers women face when applying for masculine-typed residency positions such as those in orthopedic surgery. Because this type of bias is rooted

in gender stereotypes, the selection process becomes discriminatory to women when decision makers, many times unintentionally, rely more on stereotypes of women to provide them with information, versus the objective qualities of the female applicants themselves. For male applicants, if decision makers rely on stereotypes of men to provide information, according to lack of fit and role congruity theories, male applicants would be considered a good fit, in possession of the necessary skills, and expected to be successful in the job, and accordingly offered a position as an orthopaedic surgeon over a female applicant.

There is an abundance of empirical support for the preference of male candidates over equally skilled female candidates in selection situations. For example, in a meta-analysis exploring sex discrimination in selection procedures, Davison & Burke (2000) found support that males were selected over equivalent females across 49 lab studies. Within all of these studies, the exact same resumes or job applications were presented to participants, the only difference being the gender indicated on the application (i.e., applicant's name); typically, one male and one female were simultaneously evaluated for hire. Davison et al. (2000) found that female applicants were chosen less frequently than males when gender of the applicant was salient and even when the gender of the applicant was not salient. Davison et al. (2000) operationalized gender salience of the applicant within the research design of the studies. Specifically, gender was identified as salient in a *within subjects* design, where research participants simultaneously reviewed a male and female candidate; gender was identified as not salient in a *between subjects* design, where participants reviewed one candidate, either a male or a female. This is important because even in those studies where there was not a built in comparison of a male and female candidate (between subjects design), males were still hired over equivalent females. Differential pre-entry selection of otherwise equivalent male and female candidates is outside the boundaries

of the current study, as the focal physician-leaders are currently employed. However, these findings have important implications on what happens post organizational entry, particularly on the ratings of performance and rewarding of promotions to leaders, as both workplace outcomes involve an evaluative judgment similar to selection.

Social cognitive scholars assert that decision makers use information from stereotypes most when there is little other information to use (Meyers, 1999; Smith, 2007). For example, in a hiring situation, decision makers do not have many or any experiences, observations, or interactions with the applicant to draw upon, so they are more likely to use stereotypic information to guide decisions. Consequently, when lack of fit or incongruity is perceived, females will not even be considered for the male-typed job. Burgess and Borgida (1999) proposed that descriptive sex stereotypes lead to a unique form of discrimination against women called disparate impact. They described *disparate impact* as “hiring and promotion decisions that are biased against a class of people” (p. 666). Similar to lack of fit and role congruity theory, Burgess et al. (1999) also argued that disparate impact, for women stems from a mismatch between descriptive gender stereotypes of the female applicants and the masculine gender stereotypes used as evaluation criteria. These authors provided support of their hypothesis by discussing hiring and promotion practices of organizations such as Sears, who were litigated against for favoring applicants with deep voices and who enjoyed hunting and fishing.

Heilman (1983) suggests that stereotype-based gender bias is not entirely removed for women who make it past the pre-entry stage and are hired into a male-typed job. Heilman (1983) refers to this as “post-entry discrimination,” (p. 283). This perception of a lack of fit has detrimental consequences for women already occupying a male-typed job, as perceivers attach a host of negative expectations about ability and performance (Heilman, 2001). Even in the post-

entry stage in a male-typed job, the congruence (or lack thereof) between gender and the gender-type of the job creates positive performance expectations for men and negative expectations for women (Heilman, 1983; Eagly et al., 2002). Because of these negative performance expectations due to lack of fit, women receive poorer workplace outcomes as compared to men with equivalent qualifications.

Evidence in support of post-entry lack of fit demonstrates that women, especially in higher levels of leadership, are less likely to be promoted than men (Lyness & Judiesch, 1999). For example, in a field study, Lyness et al., (1999) tracked the advancement of 30,000 managers; controlling for age, organizational tenure, and education. Their results indicated that promotion was progressively difficult for women as compared to men as they moved up the organizational ladder. The researchers explain that women face additional obstacles as they reached positions of greater success because the positions became increasingly male-typed (Lyness et al., 1999). Further, in a meta-analysis of 61 lab experiments focused on evaluations of leaders, Eagley, Makhijani, and Klonsky, (1992) found that women received poorer performance evaluations than men in jobs that are highly male-typed (i.e., coaches of athletic basketball).

Previous research clearly supports a bias that works against women for male-typed jobs, especially for leadership roles. Lack of fit theory and role congruity theory explain that this is due to the mismatch between a woman's gender and the masculine nature of the role (Heilman, 1983; Eagly et al., 2002). Past studies reveal that the impact of this mismatch are negative workplace outcomes such as lower rates of promotion (Lyness et al., 1999) and lower performance ratings (Eagly et al., 1992) for female leaders in male-typed jobs. However, there is a body of scholarship that has identified contextual buffers that may defend women from these negative consequences (Heilman, 1983; Eagly et al., 2002).

Potential Buffer to Women's Lack of Fit: Female-Congenial Contexts. In the discussion of role incongruity and women's lack of fit with leadership roles, it is important to consider how leadership is defined and enacted across different contexts. For example, definitions of leadership roles vary widely across leadership domains and industries (e.g., military, healthcare, education, and business) and within functional areas in organizations (e.g., human resources, finance, production, marketing) (Eagly et al., 2002). This is an important consideration because research supports that gender bias is determined by the magnitude of incongruity between a women's inferred attributes and perceived requirements of a masculine sex-typed job. Thus, variations of either of these factors should impact the resulting magnitude of gender bias (Heilman, 1983). Situational variables such as leadership domains and functional areas can influence the perceived sex-typed requirements of leadership job, and resultantly the degree of bias. Although leadership is typically ascribed masculine qualities, less incongruity should result for women when the position is defined in more feminine terms. As such, women may be perceived as more competent leaders in feminine jobs because those roles require communal qualities (i.e., interpersonal sensitivity) more stereotypically associated with women (Eagly et al., 2002; Mueller, 1986; Sapiro, 1983).

Perry and Bourhis (1998) found support for typing of a job on a continuum in their study of a different category of social identity: age. Though their study focused on the age-typing versus sex-typing of a job, their results supported that older applicants were evaluated more positively when in a less strongly young-typed job (i.e. fast-food worker) than when in a more strongly young-typed job (i.e., pizza deliverer). Perry et al. (1998) conclude that the degree to which a job is "typed" can act as a buffer to age bias. In other words, when applying to stereotypically young-typed jobs, older workers were less discriminated against in the less

strongly young-typed job. A similar fit process can be applied to the degree of fit between gender roles and degree of masculinity or femininity of a job.

The fit between gender roles and particular leadership roles has also been described as the *gender congeniality* of the leadership role (Eagly, Karau, & Makhijani, 1995). Eagly and colleagues (1995) defined gender congeniality as the extent to which a leadership role is perceived as a better fit for men or women. They outline that roles defined in masculine terms (i.e., requiring a high degree of agenticism and task orientation) will be perceived as a better fit for men, making the role male-congenial. Similarly, if a role is defined in more feminine terms (i.e., requiring interpersonal and communal skills), women will be perceived as a better fit for the role than men, making the role female-congenial.

The evaluation of gender congeniality of a role enabled the researchers to make predictions concerning whether the evaluation of men and women within a leadership role was related to how the role was defined. They argued that because leadership roles in general are perceived as a better fit for men (think manager-think male paradigm), leadership roles defined in relatively androgynous terms (i.e., required masculine *and* feminine gender stereotypic qualities) or in feminine terms would decrease the tension between the leader role and female role as compared to leadership roles defined more in masculine terms (Eagly & Johnson, 1990; Eagly et al., 1995). In their studies, Eagly and colleagues (1995) measured the congeniality of leadership roles by asking male and female participants (1) how effective they perceived they would be in the role; and (2) the extent to which each role was judged to require male or female gender-stereotypic qualities. Further, they took into consideration the demographic representation of men and women currently in those leadership roles to examine whether demographic representation within a role was related to the extent to which the role was

perceived as male-congenial or female-congenial.

Two meta-analyses provide support for predictions claiming that although leadership roles in general are considered male-congenial, there are female congenial roles that provide a buffer for female leaders. In a meta-analysis of 61 predominantly lab-based experiments investigating gender and the evaluation of leaders, Eagly, Makhijani, & Klonsky (1992) found a tendency for men to be evaluated more favorably than women in roles perceived as male-congenial (i.e., athletic, business, and manufacturing contexts). Further, they found a greater preference for women in roles equally occupied by men and women (less male-congenial: i.e., non-business contexts). A second meta-analysis that included 96 predominantly field-based experiments similarly investigated gender and the performance of leaders. In their analysis, Eagly, Karau, & Makhijani (1995) found support that female leaders were perceived as “modestly” (p.29) more effective than men in the domains of education, government, and social service. They found a similar effect in middle-level leadership positions, as opposed to line or supervisory positions. The researchers explain that these findings are consistent with the definition of certain industries (i.e., social service) and levels (i.e., middle management), as they are seen as requiring the type of interpersonal skills that are categorized as communal, thus feminine.

Overall, these meta-analyses support that gender congeniality of roles matter in the evaluation of men and women in those leadership roles, such that men were advantaged in male-congenial roles and females were advantaged in female-congenial roles. Their results maintain that gender congeniality is important not only for a particular leadership role, but also for level of leadership within an organization (i.e., middle and upper management), as well as on the more encompassing industry level (i.e., social service and business). Further, they found support for

their prediction that the gender congeniality of a role is directly related to the extent to which the role is dominated by men or women. Additionally, the extent to which a role, level of leadership, or industry was dominated by men or women was consistently related to the degree of gender congeniality (Eagly et al., 1990; Eagly et al., 1992; Eagly et al., 1995). The advantage of using *gender congeniality* as an indicator of potential bias is because concept and supporting research take into consideration the gender stereotypicality of a role (similar to lack of fit and role congruity theories) as well as the gender demographic composition of those currently in the role as indicators of the extent to which a male or female will be perceived as a better fit for a given leadership role.

In addition to the above research, there are studies that use the lack of fit theory to examine a potential contextual buffer from gender bias for women in leadership positions. For example, in a field study assessing the performance evaluation of 489 upper-middle-level and senior-level managers, Lynness and Heilman (2006) found that although female managers were evaluated significantly less favorably than male managers in male-typed line jobs (i.e., business management, operations management, and sales), females received a boost in ratings when in staff jobs (i.e., human resources, administration, and external affairs). Specifically, female managers in staff jobs received better performance evaluations than female managers in line jobs. Further analysis revealed that performance ratings for women in staff jobs did not differ significantly from pooled ratings for men in line jobs and staff jobs. Lynness et al. (2006) explain that because line management positions are likely to be perceived as more strongly male-typed than staff management positions, less biased evaluations of women managers occurred due to the decreased perceived lack of fit between job requirements and attributes of women.

Applying the reasoning of lack of fit theory to the male workers in Lynness and Heilman's (2006) study, one possibility is that that men in staff jobs could face bias because staff jobs are more feminine. According to the lack of fit appraisal process, men should receive lower performance ratings because of the decreased congruence with the gender-type of staff jobs, as staff jobs require more interpersonal skills, stereotypically associated with women (Paollilo, 1981). However, their results show that performance ratings did not differ significantly for men in line jobs and men in staff jobs. Lynness et al's (2006) findings suggests that men are exempt from lack of fit evaluations, such that they are not subject to the same penalties for incongruity as women are. Applying this finding more broadly, this exemption of men from lack of fit penalties could arguably occur in a male-congenial industry as it does for a male-congenial or male-typed job. In other words, although staff jobs are female-typed, the job itself is a leadership role (upper-middle-level and senior-level managers), embedded in a male-typed industry (i.e., financial services). Though the researchers' interpretations mainly focused on women, a presumable interpretation for their results for men could be that the male congeniality of the leadership roles and of the industry was enough to negate any potential disadvantage for men in a gender incongruent situation.

Pazy and Oron (2001) had similar findings in their study of the performance ratings of a large sample of 3,014 high-ranking officers in the military. They found that in the units with the lowest proportion of women, performance ratings of women were lower than men. However, in the unit with the highest proportion of women, performance ratings of women were higher than men. Interestingly, ratings of men stayed the same no matter what proportion of women were present in the unit. In other words, male officers' performance ratings were not related to the gender composition of their work units. Pazy and colleagues (2001) explain that their finding

lends support for increased stereotyping of women in conditions where women are the minority. In other words, evaluators relied more on gender stereotypes to inform their appraisals of women in a male-dominated versus a less male-dominated context. Further, although men arguably received a disadvantage in the female-dominated units, consistent with Lynness et al.'s (2006) finding for men, performance ratings for men did not differ across geniality of contexts.

Boost for Males in Female-Congenial Contexts. The performance evaluation of leaders is often times used as information to make decisions about promotions, thus central to the investigation of vertical barriers for women in male-congenial contexts. Following the predictions of lack of fit theories, although women may receive fewer promotions as compared to men in male-congenial roles, women may experience a boost or at least reach equal levels of promotions awarded in more female-congenial roles. Further, if this reasoning explains the underlying mechanisms involved when women face the vertical barriers then the same justification should hold for males in feminine gender-typed professions. That is, males who work in stereotypically female professions such as nursing or elementary education should be subjected to similar biases. Males in roles typically thought of as female-congenial should likewise be perceived to be unfit due to the stereotypically feminine qualities associated with these jobs. This lack of fit for men in female-congenial roles should then lead to similar types of bias that face women in male-congenial roles. However, studies that have examined males in female-congenial roles support that males are typically exempt from such biases, and instead are promoted above and beyond their female counterparts, often at much faster rates (e.g., Maume, 1999; Williams, 1992). Given this, the vertical barriers experienced in the workplace are almost exclusively applicable to women in the workplace.

Empirical evidence supports this male-advantage over females in promotion scenarios.

For example, in a large-scale longitudinal sociological study examining rates of promotions, Maume (1999) analyzed data from 8,534 men and 7,778 women who completed a national survey on “the fortunes of American families” (p. 499). Maume (1999) found that overall, women’s progress in reaching supervisory positions was significantly slower than men’s, such that women were rewarded with promotions at a decreased rate as compared to men. Further, he found evidence that male respondents experienced an increase in upward mobility as the percentage of females in the occupation increased. Similarly, in a qualitative investigation of males and females in female-congenial careers (i.e., librarians, nurses, elementary school teachers, and social workers), Williams (1992) found comparable results. Williams’ interviews revealed that males in these female congenial occupations were disproportionately advantaged in promotion decisions. These studies suggest that men are not only exempt from vertical barriers, but experience an advantage over equally capable women in receiving promotions in female-congenial arenas.

In summary, the theoretical and empirical analysis provides support that performance ratings and promotions are allocated in differing patterns according to gender. With respect to performance ratings, evidence suggests that female leaders experience a buffer against gender bias and may also experience a boost over male leaders in female congenial roles (e.g., Eagly et al., 1992). To the extent that leader roles are defined as less masculine in nature, leadership roles should be perceived as more congruent with a feminine gender role, and therefore the tendency to view women as less qualified than men should weaken or even disappear.

With respect to promotion, evidence suggests that males are rewarded with promotions over women across male-congenial and female-congenial leadership roles (e.g., Maume, 1999). Although these results for performance and promotion appear opposite in nature, the evidence

suggests that work outcomes (performance and promotion) for men do not differ across masculine and feminine domains. Specifically, performance ratings and promotion ratings for males in male-congenial and males in female-congenial fields largely do not differ. One explanation for this finding for men, could be due to the leadership, thus masculine, nature of the roles and responsibilities included in the studies.

For women, the overall pattern and explanation is less clear. However, in some cases within this set of studies, the researchers begin to uncover a phenomenon suggesting that female leaders in male-congenial contexts may be seen as exceptions to the “lack of fit” rule. This paper proposes that because these female leaders as seen as exceptions, they could receive higher performance ratings, but lower promotion recommendations relative to equivalent male leaders.

Exceptional Women: Higher Performance Ratings, Lower Rates of Promotion than Men.

As mentioned previously, Lyness et al. (2002) found that women in line jobs (masculine jobs) had the lowest ratings of performance as compared to men in line jobs, and men and women in staff jobs (feminine jobs). However, the researchers discovered an interesting result for promotions. Lyness and colleagues (2002) performed a supplementary analysis in which they only included managers who had received promotions. The researchers found that of those managers that had received promotions, women’s performance ratings were significantly higher than men’s in the same role. In the discussion, they explain that these female leaders had to work harder (operationalized by higher performance ratings) than their male counterparts at getting the same outcome (i.e., a promotion). Further, overall correlational analyses revealed that performance evaluation ratings were related to promotions only for women; suggesting that men received promotions no matter what their performance ratings were. Aligned with this reasoning, Eagly and Karau (2002) identified a similar phenomenon in their review of Eagly et al.’s (1995)

meta-analysis. Eagly and colleagues (2002) explain that although women in the meta-analytic study were rated as lower performing than men when leadership roles were defined in predominantly masculine terms, “those [few] women who actually obtained these [masculine leadership] positions might on average be more competent than their male counterparts and thus could have a performance advantage due to the higher standard that they have met,” (p. 587).

There is also a compositional demography based explanation for the phenomenon that female leaders could receive higher performance ratings than men. For example, in their meta-analysis looking at gender and the performance evaluation of leaders, Eagly et al. (1992) discovered a “puzzling” finding (p. 17). Their results support that subjects preferred female leaders to male leaders with more male subordinates, but favored male leaders over female leaders with more female subordinates. They explain that, “[one] possibility is that subjects engaged in relatively subtle attribution reasoning that ascribed a special competence to women who are in charge of men because such women have to be competent enough to withstand the countervailing pressures from the traditional gender hierarchy,” (p.17). This result has implications especially for female leaders in male-congenial specialties, as they will presumably have mostly male subordinates (i.e., physicians without a leadership role).

Most recently, Joshi, Son, and Roh (2015) uncovered a similar pattern in their immense meta-analysis examining gender and the allocation of outcomes at work, which included “a total of 190 effect sizes and 474,732 individuals from 142 studies (73 for performance evaluations and 69 for organizational rewards [such as promotions])” (p.1524). Their study had multiple significant findings. First, across a range of industries, occupations, and jobs they found that sex differences were nearly 14 times larger in organizational rewards (promotions) than in performance evaluations, with men having the advantage in terms of rewards. In other words,

men received significantly more promotions, and equivalent evaluations of performance as compared to women across a variation of workplace contexts. Secondly, consistent with Lyness et al.'s (2006) finding, men received promotions over women no matter what their performance evaluation ratings were. Further, Joshi and colleagues (2015) found that in *highly prestigious occupations*, or occupations with a high socioeconomic value (psychologist, physician), the sex difference in promotion but not performance was enhanced, again significantly favoring men in terms of promotion. Specifically, in highly prestigious occupations, women received equally good performance evaluations, but significantly fewer promotions than men. Lastly, consistent with prior studies suggesting a buffer against gender bias for female leaders in female congenial contexts (e.g., Eagly et al., 1992), Joshi et al. (2015) found that, “only in industries with a higher proportion of female executives did women reverse the gap,” particularly in obtaining promotions as compared to men (p. 1532).

In the discussion, the scholars highlight the particularly “striking” pattern in their findings that, especially in highly prestigious occupations, men were rewarded with promotions significantly more than women, even though women did not perform worse than men (Joshi et al., 2015). They emphasize the importance and the need for further study of disentangling promotion allocation outcomes from performance evaluation outcomes particularly in highly prestigious settings such as law and academia. Joshi and colleagues (2015) attempt to explain these findings by proposing that in these settings, employment practices likely function as, “hierarchy-enhancing agents that maintain social hierarchies facilitating male dominance,” (p. 1533). For example, professions such as academia, on one hand, tend to have “up or out” promotion norms with processes that are subjective and opaque in nature; and on the other hand, tend to have performance evaluation criteria that are more objective in nature (e.g., research

productivity). These norms and practices maintain the dominance of men in highly prestigious contexts, awarding them with the more valuable outcome (promotion) that provides occupational mobility and organizational power. However, they conclude by revisiting their finding that in occupations with a higher proportion of female executives, this bias against women in terms of promotions is buffered. Joshi and colleagues (2015) call for an examination of contexts with a high proportion of female leaders to further understand how and why employment practices become more egalitarian in nature.

Lastly, it is relevant to consider another set of recent studies conducted by Leslie, Manchester, and Dahm (2016). Leslie and colleagues (2016) contribute to the current discussion of gender and the allocation of organizational outcomes, because they proposed a priori hypotheses outlining that in certain male dominated organizations, *high potential* women who have the skills needed to succeed to upper levels in organizations exceedingly dominated by men, will receive better outcomes (i.e., pay) than equivalent men. They propose and find evidence of a *female premium*, “unique to women with high potential, driven by perceptions that these women are valuable for achieving organizational diversity goals, and larger in organizations where diversity goals are stronger,” (p. 36). They explain that because women are the minority in these settings, high potential women are perceived as more valuable than men due to the larger organizational diversity goals. Although Leslie et al. (2016) focus on a different outcome (pay) in an organizational context with a different characteristic (value diversity goals), the scholars provide evidence that the typical gender/role congruity matching process (Eagly et al., 1991; Heilman, 1983) may be different for exceptional, successful, and high potential women in highly male-typed contexts.

Taken together, these studies provide evidence that exceptional women may receive higher performance ratings but lower recommendations for promotion than men because they have to or are perceived to work harder to get to the same level. Applied to the highly masculine context of physicians and physician leadership, female physician leaders could be awarded higher performance ratings, but lower promotion recommendations than male physician leaders. Additionally, it is arguable that this effect will be stronger more in male-congenial specialties (i.e., surgery) than female-congenial specialties (i.e., pediatrics), because females may be perceived as having worked very hard to get promoted to a leadership position in specialties where women are the minority. For male physician leaders, there is evidence that their performance and promotion outcomes could stay relatively equivalent across male and female-congenial specialties. Results for male physician leaders as compared to female physician leaders is clearer for the outcome of promotions, as the studies cited consistently report that men receive more promotions than women. The results regarding performance evaluations is less clear, as women can be seen as an exception, especially in male-congenial specialties; but consistent with prior findings (e.g., Joshi et al., 2015) perhaps less so in female congenial specialties because they are more equal in proportion and the specialty itself may be perceived as more female-typed. For example, pediatrics is arguably more stereotypically female in nature because (1) the specialty has more women as compared to all other specialties; and (2) the specialty involves the treatment and care of infants, which is stereotypically feminine in nature.

The following hypotheses describe medical specialties in gender-congenial terms because all specialties are situated in the higher, industry-level male-congenial context of medicine. Thus, a female dominated specialty (i.e., pediatrics) may not necessarily be female-congenial per se, due to being embedded in a higher level masculine industry. However, the lower, specialty-level

context could be described as less male-congenial, or more female-congenial. The following set of hypotheses propose a potential buffer for or bias against female physician leaders. As such, pursuing Joshi et al.'s (2015) call to examine highly prestigious or highly masculine contexts, the “most” male-congenial specialty (i.e., surgery) and in contrast, the “most” female-congenial specialty (i.e., pediatrics) were chosen as comparison conditions to ensure the clearest pattern of results. Lastly, see Figure 2 for a model of the moderation of specialty to gender on promotion and performance.

Gender x Specialty → Promotion

- Hypothesis 1a** Successful male physicians will be perceived to be more promotable than successful female physicians, no matter which specialty.
- Hypothesis 1b** Successful female physicians in pediatrics will be perceived to be more promotable than successful female physicians in surgery.
- Hypothesis 1c** Successful male physicians will be perceived to be equally promotable in pediatrics and surgery.
- Hypothesis 1d** Differences in promotion between successful male and female physicians will be more amplified in surgery than in pediatrics.

Gender x Specialty → Performance

- Hypothesis 2a** Successful female physicians will be perceived to be better performing than successful male physicians, no matter which specialty.
- Hypothesis 2b** Successful female physicians in surgery will be perceived to be better performing than successful female physicians in pediatrics.
- Hypothesis 2c** Successful male physicians will be perceived to be equally performing in pediatrics and surgery.
- Hypothesis 2d** Differences in performance ratings between successful male and female physicians will be more amplified in surgery than in pediatrics.

Explanations for Success: Attribution-based Gender Bias

The discussion of descriptive stereotypes, lack of fit, and role congruity outlined how female leaders may experience a boost in recommendations for promotion in female-congenial

contexts as compared to female leaders in male-congenial contexts, while male leaders most likely will benefit from a higher rate of promotions over female leaders because the leadership role is associated with masculine traits (Heilman et al., 1989; Schein, 1973). With regards to performance ratings, although female leaders overall may receive higher ratings than men, researchers have explained that it is presumably because those women are seen as having to work much harder than men to obtain the same results (i.e., promotion) (e.g., Lynness et al., 2006). Overall, these predictions support what Heilman (1983) called post-entry discrimination for women.

In addition to the performance evaluation and promotion biases reviewed above, Heilman (1983) noted another form of post-entry discrimination. Heilman (1983) includes biased causal inferences regarding successful performance, as a third challenge women face in the workplace. Specifically, when a woman successfully performs a masculine job task, it is often viewed as the result of something other than her skill or competence. According to attribution theory, when an individual performs outside of what is expected, the results are often credited to factors beyond the individual's ability (Deaux & Emswiller, 1974). Thus, because women are expected to fail or underperform in a male-typed job, when a woman is instead successful, her achievements are viewed as a result of working especially hard, an anomaly, or due to chance (Deaux et al., 1974; Heilman, 1983; Swim & Sanna, 1996). Such attributions support the perpetuation of descriptive stereotypes of women and inhibit the use of objective evidence of her performance.

To date, empirical studies on attribution and performance have largely involved attributions as evaluations of one's own performance (e.g., Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971); an outcome variable (e.g., Swim et al., 1996); or a result of a singular event or performance on a task (e.g., Deaux et al., 1974; Garcia et al., 2006). However, there is little

known about whether attributions can act as a mediator, explaining the link between gender and workplace outcomes (i.e., performance and promotion). The current study simultaneously explores attributions as outcomes and predictors, building the case for that attributions may contribute to gender bias by mediating the impact of leader gender on performance evaluations and recommendations for promotion. Additionally, because individuals use attributions to make sense of achievement related behaviors, namely success or failure on a task (Weiner et al., 1971), the current study presumes that because the physician leaders already occupy a leadership role, they should be perceived as successful in his or her career. In support of this assumption, in a recent review of gender bias in organizations, Elsessner (2015) explained that, “individuals who do not possess the traits desirable for leadership in a particular organization will simply not be promoted to the leadership level, and individuals chosen to manage will typically reflect the norms and expectations of their organization,” (p. 165).

The following sections will explore what is known about attribution theory in order to examine attributions as a possible mediator of gender bias on performance and promotions for physician leaders.

Attribution Theory. Attribution theory is rooted in foundational scholarship produced by Kurt Lewin, Julian Rotter, John Atkinson, Fritz Heider, and Harold Kelly (Weiner, 1990). Heider (1959) explains that forming attributions about others is a fundamental cognitive process, motivated by the basic human need to make sense of the world. According to the theory, this is part of an automatic social cognitive sensemaking process that humans engage in to explain why people behave the way they do. By using contextual information to form assumptions and formulate rationalizations of another person’s intentions, motivations, and behaviors, humans are able to explain past behavior and predict future behavior, thereby decreasing anxiety and fear by

making the unknown, known (Hewstone, 1989). However, the human need to make sense of surroundings does not necessarily lead to accurate or objective explanations of others' behaviors. Applied to the work context, inaccurate explanations of behavior have significant implications to employees, especially those in a performance evaluation process.

Causal Attribution Theory in the Work Context. Weiner and colleagues (1971) extended the scholarship on attribution theory and focused on the individual's tendency to engage in a causal search process following an action or performance event enacted by another individual. They describe that individuals generate causal explanations to explain and identify why an event occurred. Specifically, *attributions* are the causal explanations generated to explain an individual's performance (Heider, 1958). Focusing on the achievement domain (i.e., success or failure), Weiner and colleagues (1971) provided ways of identifying characteristics of causal attributions and detailed how certain causal explanations can differentially impact subsequent evaluations and decisions. Because the current study aims to examine the role of attributions in influencing the performance evaluation of successful leaders, Weiner's (1985) attributional framework was adopted to explore the predictions in the current study.

According to Weiner (1985), individuals tend to explain the cause of achievement related behaviors by attributing other's success in one of four ways: ability, effort, task difficulty, and luck. Specifically, *ability* refers to a person's inherent general intellectual or physical aptitude; *effort* refers to how hard an individual appears to work; *task difficulty* is the level of ease of a task of job and is often appraised by observing and comparing how individuals perform on the same task; and *luck* often is evaluated by the degree of unpredictability of an observed outcome.

These four dimensions of attributions are explained through a framework of three causal dimensions: locus of control, stability, and controllability. See Figure 3 for a matrix of

Attributional Explanations Categorized by Attribution Dimension. The property of *locus of causality* concerns the location of the cause of performance, specifically whether the cause of the behavior is internal (i.e., caused by aspects of the individual) or external to the individual (i.e., caused by aspects of the environment). *Ability* and *effort* are characterized as internal because they describe qualities within the individual. Conversely, *task difficulty* and *luck* are characterized as external because they involve factors that are not part of the individual, but rather part of the environment (Weiner, 1985).

The second property of *stability* concerns the constancy of a behavior, as it describes whether the cause of behavior is perceived to be invariant (i.e., stable) or changeable over time (i.e., unstable) (Weiner et al., 1971). *Ability* and *task difficulty* are categorized as causally stable because an individual's innate aptitude and the difficulty of a particular situation are generally not variable. On the other hand, *effort* and *luck* are categorized as causally unstable because level of effort and the unpredictability of luck are variable in nature (Weiner, 1985).

The third property of *locus of controllability* was added later by Weiner (1985) to allow for a more nuanced interpretation of a person's behavior. It concerns the controllability of a situation, specifically whether there is control over the behavior or situation (i.e., *effort* and *task difficulty*) or whether there is little control over the behavior situation (i.e., *ability* and *luck*). Of the three overarching categorical properties of causal attributions, the internal-external locus dimension is the most fundamental causal distinction with respect to attributions and has been the focus of most empirical attribution research. However, it is the controllability dimension that provides the most information to perceivers because it helps in determining whether or not the cause of behavior was within or outside of the individual's control (Heider, 1958; Weiner, 1985; Weiner et al., 1971). Supporting this assertion, Weiner & Kukla (1970) found that the attribution

to *effort* generally weighed most heavily in an evaluation situation. They explain that effort is variable (i.e., unstable) and usually under the control of the individual (i.e., internal and controllable). Therefore, from the perspective of an evaluator in a performance evaluation situation, Weiner et al. (1970) explain that degree controllability is a helpful measure in determining the deserved level of reward or punishment.

However, there is also evidence that rewards might be unequally allocated along gender lines due to differential attributions ascribed to each gender. The following section will review differential causal attributions based on gender.

Attribution-based Gender Bias. According to social psychological scholars, the determination of ability, effort, task difficulty or luck is assessed by the degree to which an individual performs consistently or inconsistently with ascribed expectations. These scholars explain that behaviors perceived as consistent with stereotype-based expectations are attributed to stable causes (i.e., ability and task difficulty), whereas behaviors inconsistent with stereotype-based expectations are attributed to external causes (i.e., task difficulty, luck) or internal, unstable causes (i.e., effort) (Deaux, 1984; Hansen & O’Leary, 1985; Swim et al., 1996). Because success on a masculine task by women are less expected than success by men, attributions for women's successful performance are less likely to be attributed to ability than men’s successful performance (Deaux, 1984). In other words, a woman’s success is more likely to be attributed to their effort, the task, or luck and not to their ability, particularly for masculine tasks.

Hansen and O’Leary (1985) have explained that one possible reason for this pattern is, “that attributions are a direct result of an expectation indicating that women try harder (which they suggest may reflect reality), work on easier tasks, and have better luck or that women have

low ability and low performance levels as compared with men,” (Swim et al., 1996, p. 508). Another possible reason is that, “the attributions are consciously driven by a desire to maintain the status quo,” (Swim et al., 1996, p. 508). In accordance with descriptive stereotype based theories of fit, because women are incongruous with the masculine requirements of a male-typed job, they are expected to perform less well than men. However, if a woman succeeds at the masculine job, attributing her success to reasons other than her ability maintains the cognitive “status quo.” Therefore, perceivers avoid having to reconcile dissonant pieces of information, thus reserving cognitive resources. The same line of reasoning can be applied to the attribution of *ability* to explain a man’s successful performance in a male-typed job. Specifically, because men are congruous with the masculine requirements of a male-typed job, they are expected to perform well.

There is some evidence linking gender stereotypes and attributions, generally supporting the notion that performance consistent or inconsistent to expectations are attributed to different sets of causal attributions. For example, in their seminal finding, Deaux & Emswiller (1974) found support that a woman’s success on a masculine task was found to be attributed to luck rather than ability; and on the other hand, a man’s success on the same task was attributed to ability rather than luck. Feldman-Summers & Kiesler (1974) found comparable results for attribution of success on a masculine task; effort was attributed more to a woman’s success as compared to an equivalent man. Similarly, Cash et al. (1977) found that success in male-typed tasks is attributed to ability more for males than females. In general, prior research supports that men’s and women’s successes are causally attributed according to the consistency of performance outcome with stereotype-based expectancies, versus the performance itself. In other

words, there is evidence of attributional gender bias such that a man's and a woman's equivalent success on the same task results in attributions of ability for men and not for women.

Interestingly, similar to the mitigating effects of congruency with the context previously discussed, these studies found that when a woman was successful at task that is female sex-typed, differential attributions based on gender was less likely to occur (Deaux et al., 1974). A meta-analytic study conducted by Swim and Sanna (1996) investigated observers' attributions for women's and men's successes and failures provided further support. For masculine tasks, participants attributed men's successes more to ability than women's successes; participants instead attributed women's successes to effort. However, their analysis also found that attributional gender bias (i.e., attributing equivalent success on a task to ability for men, but not for women) was observed more with masculine tasks than with feminine tasks. This finding offered support that gender-typed work context acted as a moderator to the relationship between gender and causal attributions (Swim et al., 1996). Specifically, for success at a feminine task, differences in ratings of effort and ability between male and female targets were not significantly different. The researchers explain that either men are generally seen as more skilled than women, or that feminine tasks are generally less difficult than masculine tasks (Deaux et al., 1974; Swim et al., 1996). In summary, prior research supports that for a male-typed task, men's success is attributed to ability, whereas women's successes are typically attributed to increased effort, good luck, or task ease. Further this main effect can be moderated by the sex-type of the task, such that for feminine tasks, the differential assignment of attribution based on gender should decrease (Cash et al., 1977; Deaux & Emswiller, 1974; Feldman-Summers & Kiesler, 1974; Swim et al., 1996; Taynor & Deaux, 1975).

Although these findings provide clear support for attributional gender bias, these studies largely conclude with the prediction of causal attribution (i.e., attribution type was the outcome variable studied). There are only a few studies that went a step further to investigate how differential attributions impact workplace outcomes awarded to targets under evaluation. One of these studies was conducted by Heilman and Guzzo in 1978. Their study investigated the perceived cause of sex discrimination in organizations and hypothesized that biased causal attributions based on gender explained the allocation of outcomes. In their lab experiment, Heilman et al. (1978) manipulated cause of success, explaining in their stimulus materials that the target's success was caused either by ability, effort, task ease, or luck. For example, ability, effort, task difficulty, and luck conditions were operationalized, respectively, as: "I would say that Paul (Paula) is one of the most capable people I've seen in years; Mark (Marcia) is one of the most hardworking people I've ever met; If you ask me, this job is one of the least demanding ones around; and Michael (Michelle) was simply in the right place at the right time," (Heilman et al., 1975, p. 350). It is important to note how the researchers applied attributions to a more encompassing evaluation of performance to a summarial judgment of performance versus solely on a singular task. Heilman et al. (1975) then asked the participants to rate the degree to which the target (a man or a woman) deserved a pay raise and a promotion. Results indicated that the causal explanations typically used to explain a woman's success (i.e., effort, task difficulty, luck) as compared with what is typically used to explain a man's success (i.e., ability) resulted in fewer and less desirable organizational rewards in the form of promotion (there were no significant effects associated with pay raise). Upon closer inspection, because they manipulated instead of measured attribution, they researchers could only conclude that it was the attribution assigned, and not gender of the employee that determined a promotion.

The second study that investigated outcomes as well as gender bias in attributions was more recently conducted by Garcia-Retamero and Lopez-Zafra (2006). This study is perhaps most relevant to the current study as the researchers investigated this cross-section of phenomena for men and women in a leadership role. As discussed earlier, the role of “leader” is sex-typed as male, and typically associated with masculine-typed tasks and responsibilities. Results supported their predictions: first, female leaders received poorer workplace outcomes (i.e., promotion, performance, and salary recommendations) than male leaders in the masculine industry (i.e., auto manufacturing). In the feminine industry (i.e., clothing manufacturing), female leaders received better outcomes than male leaders. Further, the researchers examined male and female leaders separately for analyses which investigated attribution as a precursor to workplace outcomes. Results supported that for a female leader, success was attributed to internal causes (operationalized as “capacity and preparation”) in the feminine industry and external causes (operationalized as “manager decision”) in the masculine industry. For male leaders, success was attributed to internal causes across both industries. Similar to Heilman et al. (1978), this finding was only supported for the recommendations for promotion outcome.

Though these results are generally consistent with prior studies, Garcia-Retamero et al.’s (2006) operationalization of causal attribution was not directly consistent with prior studies involving gender and attribution, primarily because they combined “capacity and preparation” into one measure of internal control. This is problematic because capacity is similar to ability, and preparation is similar to effort; therefore making it difficult to conclude whether ability was assigned to men over women as many previous studies supported. The only somewhat clear conclusion based on their operationalization is that internal attributions were associated with better recommendations for promotion as compared to external attributions. Although this

conclusion is informative, the distinction between effort versus ability is important because prior scholarship supports that 1) ability is typically reserved for men; and 2) effort was identified as the most significant determinant of rewards because it is internal and under the person's control (Weiner, 1985). Thus, it is difficult to build on past findings with Garcia-Retamero et al.'s (2006) findings.

The following hypotheses on attribution-based gender bias were formulated by combining theoretical predictions and empirical support from previous studies. The first set of studies presented above examined casual attributions as an outcome variable (i.e., those reviewed in Swim & Sanna's meta-analysis in 1996). These studies consistently supported that men's success is attributed to ability, and women's equivalent success is more likely to be attributed to effort, the task, or luck, particularly for masculine tasks. For feminine tasks, men and women's success is more attributed to ability (Swim & Sanna, 1996).

Further, because the current study is situated an overall male-congenial context of medicine, evaluators will most likely expend more cognitive effort in scrutinizing the cause of woman's success (i.e., assign an attribution) because she is doubly incongruous with the context (medicine) and the role (leader). Due to this as well as Weiner et al.'s (1970) finding that attribution to *effort* generally weighed most heavily in an evaluation situation, I suggest that perceivers will likely attribute female physician leader's success more to *effort* than other attributions (i.e., task difficulty and luck), because it has been established that *ability* is reserved for men, and because perceivers will need help forming causal attributions for the dissonant target (i.e., draw information more from stereotypes than the individual). This is consistent with Lyness et al.'s (2006) assertion that exceptional women may receive higher performance ratings but lower recommendations for promotion than men because they have to work harder (i.e., give

more effort) to get to the same level. This lends justification to include only *ability* and *effort* as attributions in this study. Further supporting this justification, Heider (1958) explained that internal attributions used for explaining someone's actions offer the most information about that person and therefore are given more attention than external attributions, because behaviors that are perceived to be internally caused informs a person's level of responsibility of the behavior, and therefore provides useful information about how to react and interact with this particular person or in this particular environment. See Figure 4 for a model of the moderation of specialty to gender on ability and effort. And see Figure 5 for a model of the impact of attributions on employee outcomes.

Gender x Specialty → Ability

- Hypothesis 3a** Participants will attribute ability more to successful male physicians than successful female physicians.
- Hypothesis 3b** Participants will attribute ability more to successful female pediatricians than successful female surgeons.
- Hypothesis 3c** Participants will attribute ability similarly to successful males in pediatrics and surgery.
- Hypothesis 3d** Differences in ability attributions between successful male and female physicians will be more amplified in surgery than pediatrics.

Gender x Specialty → Effort

- Hypothesis 4a** Participants will attribute effort more to successful female physicians than successful male physicians.
- Hypothesis 4b** Participants will attribute effort more to successful female surgeons than successful female pediatricians.
- Hypothesis 4c** Participants will attribute effort similarly to successful males in pediatrics and surgery.
- Hypothesis 4d** Differences in effort attributions between successful male and female physicians will be more amplified in surgery than pediatrics.

Ability & Effort → Promotion & Performance

Hypothesis 5a Effort attributions will be a stronger predictor of performance than ability attributions, and targets with higher effort attributions will be perceived as better performers than targets with lower effort attributions.

Hypothesis 5b Ability attributions will be a stronger predictor of promotion than effort attributions, and targets with higher ability attributions will be perceived as more promotable than targets with lower ability attributions.

Participant Social Dominance Orientation as a Moderator

Apart from target characteristics (e.g., gender) and contextual characteristics (i.e., degree of masculinity of a job), social psychologists have also considered evaluator characteristics as a factor that can potentially buffer or exacerbate gender bias at work. Research on individual differences that moderate gender bias is “in its infancy” as compared to target or contextual moderators, consequently contributing to mixed findings on most often studied factors such as participant gender and attitudes towards women (Rudman & Phelan, 2008, p. 71). For example, some studies support that male and female participants equally penalize incongruous targets because of societal knowledge of gender role stereotypes (Heilman et al., 2004; Rudman & Glick, 1999; Rudman & Glick, 2001), while other studies found support for participant gender differences such that men, as compared to women, penalize incongruous targets more because, “men often have a more masculine construal of leadership than to women,” (Eagly et al., 2002; p. 577). Further, an often included individual level difference, attitudes toward women, has proved an unsuccessful moderator to gender bias (e.g., Rudman & Glick, 2001).

Recently however, there has been consistent evidence of a potentially strong individual difference that may have attenuate or exacerbate gender bias, especially in highly masculine contexts such as physician leadership. Recall the “striking” pattern of results found consistently in the large scale meta-analysis conducted by Joshi and colleagues (2015): specifically in highly prestigious occupations were women acutely rewarded less than men, even if their performance

levels were not lower than men. The researchers concluded that the penalty women received in these highly prestigious and highly masculine settings was due to organizational practices that “maintain social hierarchies facilitating male dominance” (p. 1533). Joshi and colleagues (2015) stress the importance of future research that studies the *social dominance orientation* (SDO) in key organizational decision makers, particularly because these key decision makers tend to be high in SDO in highly masculine contexts.

Therefore, another factor that may affect the hypothesized relationship of physician leader gender and specialty on workplace outcomes is the social dominance orientation of the evaluator. *Social dominance orientation* (SDO) is an individual difference variable that considers the extent to which a person believes that social group hierarchies should exist (Pratto, Sidanius, Stallworth & Malle, 1994). People high in SDO have “the tendency to hold nonegalitarian values and to support hierarchically structured relationships among social groups” (Umphress, Smith-Crowe, Brief, Dietz, & Watkins, 2007, p. 396). Specifically, those high in SDO prefer to maintain identity-based group hierarchies such that some groups have access to greater resources than others. Conversely, those low in SDO prefer identity-based groups to be equal in terms of having access resources (Pratto et al., 1994).

SDO is often used to legitimize or justify why certain groups belong at the top of the social hierarchy (Pratto et al, 1994; Whitley, 1999). Previous research supports that SDO impacts perceptions of and attitudes towards different gender groups, with those high in SDO maintaining that men are more socially dominant and thus should be higher in the social hierarchy than women. A study conducted by Cokley, Tran, Hall-Clark, Chapman, Bessa, Finley and Martinez (2010) supported that participant SDO was the strongest predictor of attitudes toward gender diversity (i.e. beliefs one holds about women and gender equality). Specifically,

the higher the participant's score on SDO, the more negative cognitive attitudes the participant held toward gender diversity. Further, Hoyt (2012) found that for equivalent men and women in a male-typed job, participants low in SDO demonstrated a "pro-female" bias while participants high in SDO demonstrated an "anti-female" bias. Overall, scholars who study social dominance theory strongly assert that participant SDO can strongly influence the typical "gender-matching heuristic that matches the gender-type of the applicant with that of the job," (Pratto et al., 1997, p. 47).

SDO is considered an important individual difference because it may help to explain the social-psychological trait, on the part of the evaluator, that creates and perpetuates social identity based group inequality in organizations. This group-centered social dominance theory is particularly a concern because institutional gender bias is one of the major contributors to the maintenance of gender inequity. Based on this reasoning, I propose the following research question:

Gender x Specialty x Participant SDO → Outcome Variables

Research Question 1 Will participant SDO moderate the effects of the study predictors (gender X specialty) on the outcome variables (promotion, performance, ability, effort)?

Effort and Ability Attributions as Mediators

Although an abundance of evidence supports that effort and ability attributions are operationalized as outcomes as well as predictors, it is largely unknown whether attributions mediate the relationship between gender and workplace outcomes such as performance and promotion. There is one study that claims to investigate attributions as a mediator of gender bias in the form of organizational rewards (i.e., Heilman & Guzzo, 1978). However, in closer review of the manuscript, the researchers did not directly test if attributions mediate the relationship

between gender and outcomes. In their study, the researchers tested (1) if attributions impacted workplace outcomes (i.e., promotion decisions) with ANOVA analyses; and (2) if target gender was related to attributions with a chi square analysis (Heilman et al., 1973). To address this gap, this study explores whether effort and ability attributions mediate the relationship between target gender and specialty on workplace outcomes. Accordingly, the following research question is proposed. See Figure 6 for the full mediated moderation model of specialty congruency and success attribution on work outcomes.

Gender x Specialty → Ability & Effort → Promotion & Performance

Research Question 2 Will effort and ability attributions act as mediators to the relationship of gender X specialty X SDO on performance and promotion?

CHAPTER THREE: METHOD

Participants

Two hundred and seven individuals residing in the U.S. completed the study. Participants in the sample were recruited from Amazon Mechanical Turk, an online crowdsourcing marketplace, where individuals self-selected into this study after reading a brief description of the research. One hundred and two (49%) of the participants were female and 105 (50.7%) were male. The average age of participants was 37.10 years ($SD = 12.55$). Participants self-identified as 156 (75%) White, 19 (9.2%) African American or Black, 14 (6.8%) Asian or Asian-American, 14 (6.8%) Hispanic or Latino, and 4 (2%) American Indian or Alaska Native. Participants had an average of 5 years ($SD = 2.6$) of full time work experience. Thirty-five (16.9%) participants indicated that their highest level of education was a masters or doctoral degree, 83 (40.1%) a four-year degree, 23 (11.1%) a 2-year degree, 41 (19.8%) some college, 24 (11.6%) a high school degree, and 1 (.5%) had less than high school. Participants consisted of individuals working in a diverse range of fields including education, entertainment, sales, government, administration, and information technology. Three (1.4%) participants indicated having executive level experience, 11 (5.3%) senior-level management, 47 (22.7%) management, 70 (33.8%) professional level, and 56 (27.1%) entry level experience.

Research Strategy & Design

This scenario-based study was presented by the researcher and completed by participants online. The study was a 2 (target gender) x 2 (target medical specialty) x 3 (participant social dominance orientation) between-subjects factorial design. Participants were randomly assigned to one of four study conditions by a function available within Qualtrics.com, where the survey was created and maintained. Participants rated one target (male or female) who worked as a

physician specializing in either pediatrics or surgery. The study was part of a larger study looking at the consequences of stereotype violation on interpersonal outcomes at work (liking, boss desirability) (Heilman & Okimoto 2007). Only the measures and data of the current study are reported here.

Procedure

Participants were recruited from Amazon Mechanical Turk (MTurk), a crowdsourcing online marketplace in which tasks are distributed to an anonymous population of workers for completion. Participation was available to MTurk workers residing in the U.S. with an approval rate of at least 95%, meaning that task requesters approved at least 95% of the workers' previous completed tasks. Participants completed the study at their own convenience at a device of their choosing. Participants were compensated \$1.50 for their participation and were told that the study would take about 30 minutes to complete.

To participate, respondents self-selected into the study by reading a very general description of the study, time commitment, and what participation involves. Specifically, potential respondents read, "Short survey study: 30 mins. After reviewing an employee profile, you will be asked a series of questions about the person, then about yourself."

Participants who chose to participate were directed to a link that led to the survey, which was created and maintained on Qualtrics.com.

Participants were first given introductory information to the study including a cover story, a guarantee that participation was voluntary, confidential, and anonymous. They were also given researcher contact information should they have questions regarding the study.

The cover story described the study as a way to understand and improve the performance evaluation system within academic medical institutions. Specifically, the cover story included the following:

“The overall goal of this study is to help improve performance evaluation tools and processes in academic medical institutions. We hope to better understand how people form impressions of faculty in academic medical institutions and how they make employment related decisions about them (e.g., reappointment, promotion, and compensation).

Specifically, we are interested in addressing the balance in providing enough but not too much information to reviewers of performance evaluations.”

See Appendix A1 for the cover story. After clicking the “next” button, the participants reviewed the participant informed consent information. Then followed the participant’s rights page. See Appendix A2 for informed consent and Appendix A3 for participant’s rights. Lastly, participants were informed that consent to participate was provided by clicking the “next” button to advance to the beginning of the study.

The study was a scenario-based study in which participants saw one of four alternative scenarios. Participants were presented with a hypothetical promotion scenario at a fictitious academic medical institution. Participants were asked to play an evaluator role and to provide their first and general impressions about the person under review. The study materials included background information about the target’s specialty (pediatrics or surgery), a letter from the dean of the institution asking for participants’ input regarding the target’s promotion, a completed set of performance evaluation materials for the target, followed by the study measures.

Once in the online survey, participants viewed a screen titled *Instructions*, which explained that they would be asked to play the role of a co-faculty member in the target’s department. It also informed the participants they would be provided with promotion materials and would be asked a series of questions including promotion related decisions and perceptions

of the target based on those materials. The process of having colleagues weigh in on promotion decisions is customary for academic promotions in medical institutions (Simpson, Hafler, Brown, & Wilkerson, 2004). See Appendix B for study instructions.

The next screen participants viewed contained *Background Information* on the specialty of pediatrics or surgery in the U.S. This provided participants with a general definition of the specialty and actual general statistics of the specialty condition (pediatrics or surgery) compared to all other medical specialties in the U.S. (i.e., number of active physicians, number of active physicians by gender, and numbers of active physicians by age) (AAMC, 2014). The purpose for including this information helped ensure that participants had the same basic knowledge of the specialty condition. To continue, the participants were prompted to click “next page” to begin their role as a pediatrics or surgery faculty member. See Appendix C1 for pediatrics background information and Appendix C2 for surgery background information.

The next two screens contained the promotion materials, both containing a logo of the fictitious academic medical institution. The first screen contained a brief memo from the dean of the institution addressed to the faculty members of the target’s department (pediatrics or surgery). The memo reminded faculty that their colleague was up for promotion from Associate Professor to Full Professor and that a completed set of his/her summary performance evaluation materials was available for their review. See Appendix D1 for the pediatrics dean’s letter (female condition) and Appendix D2 for the surgery dean’s letter (male condition).

The summary performance evaluation materials were included to demonstrate that the target was a successful and very qualified for the Full Professor position. Each performance item in the summary evaluation was presented with two numerical averages. One average was the target’s aggregated average across all raters. The other was the aggregated average for all faculty

in the department. The target's average was higher than the aggregated faculty average for each of the 18 items. Specifically, on a scale of 1 to 5 (where 1=below expectations, and 5 = markedly exceeds expectations), the overall average for the target individual was a 4.49, while the overall department faculty norm was a 4.1. The purpose of this was to present the target as slightly more successful than most of the faculty in the department. All information on the summary performance evaluation was identical for the male and female targets, except for specialty type. All targets were presented as physician leaders in that they all occupied a leadership role as, the Residency Program Director in their department. Lastly, the performance indicators in the summary evaluation were adapted from an enrichment and mentor guide to increase the realism of the performance content (Icahn School of Medicine, 2016). See Appendix E1 for the pediatrics summary performance evaluation (female condition) and Appendix E2 for the surgery summary performance evaluation (male condition).

A pilot study of the promotion materials indicated that the target was rated an average of 6.09 on a seven point perceived success item, where 1 = not at all successful and 7 = very successful. Participants of the pilot study included subject matter experts (i.e., medical school professors) as well as non-medical individuals. The summary performance evaluation was adapted from an actual performance evaluation form from an accredited U.S. academic medical institution.

After reviewing these materials, the participants proceeded to the next screen, which informed them of the upcoming questions to which they were to respond. Participants were again reminded to provide their first impressions and immediate reactions to the target. Before proceeding to the next screen, participants were asked to type in the name of the faculty member under review. The purpose of this was to re-focus participants' attention to the target and his or

her gender. It was also an attention check item. See Appendix F for the questionnaire introduction.

The next three pages contained study measures including promotion and performance and attribution items. The following page contained manipulation check questions. The last page contained questions about the participant themselves including the social dominance orientation measure as well as demographic questions.

At the end of the questionnaire, participants read a debrief in which the research objectives were explained and discussed in greater detail. Within the debrief, participants were again provided with contact information of the researcher if they wished to follow up with questions or concerns. Participants then received instructions to enter a unique code in MTurk to verify that they completed the study. See Appendix G for the study debrief.

Experimental Manipulations

Target Gender. The target's gender was manipulated by the name of the target physician: Michelle Anderson or Michael Anderson. Prior research has manipulated gender by using easily recognizable female and male names (e.g., Heilman & Guzzo, 1978; Heilman & Okimoto, 2007). The target physician's name was mentioned four times in the vignette section: twice in the dean's letter (see Appendix D2 for the female condition and Appendix D2 for the male condition), and twice in the summary performance evaluation (see Appendix E1 for the female condition and Appendix E2 for the male condition).

Target Specialty. The sex-typing of the job was manipulated by specialty. Targets were presented as either in the surgery department (i.e., stereotypically male sex-typed) or in the pediatrics department (i.e., stereotypically female sex-typed). The selection of these two specialties was twofold. First, the two specialties are highly contrasting of one another in terms

of gender composition and stereotypical gender roles associated with the job. Specifically, surgery is one of the most highly masculine and highly prestigious jobs due to the overrepresentation of men in the specialty as well as the heightened societal value placed on the job; while pediatrics is the specialty with the largest percentage of women and requires skills associated with being stereotypically feminine in nature (Heilman, 2012; Joshi et al., 2015). Second, the only other known study that examined successful physicians by comparing surgeons and pediatricians was Feldman-Summers and Kiesler (1974). The scholars reported their selection of these two specific specialties was based on their pretest data, which suggested that “males were not expected to succeed more than females in pediatrics, but men were expected to be more successful than women in surgery,” a distinction similar to the stereotypical gender roles and fit associated with the job (p. 849). To ensure equivalent knowledge of the specialty, basic actual U.S. statistics of either surgery or pediatrics were presented in the *Background Information* section. Specifically, number of active physicians by gender was included; pediatricians: 60.4% female (see Appendix C1); and surgeons: 17.6% female (see Appendix C2). Further, in the promotion materials section, target specialty was mentioned nine times: six times in the dean’s letter (see Appendix D1 for pediatrics and Appendix D2 for surgery) and three times in the summary evaluation (see Appendix E1 for pediatrics and Appendix E2 for surgery).

Measures

Dependent Measures

Promotion. The recommendation for promotion measure consisted of a composite rating of three 7-point scale items. The first item asked, “To what extent would you recommend the individual to be placed in a prestigious upper-level position in the organization?” (*not at all – very much*) (Heilman et al., 2004, study 3). The second item asked, “How likely would you be to

recommend this person for a promotion?” (*not at all likely – extremely likely*) (Benard & Corell, 2010). The third item asked, “How likely would you be to recommend that this organization invest in leadership training and education for this individual?” (Heilman et al., 2004, study 3) (*not at all likely – extremely likely*) ($\alpha=.83$). See Appendix H1 for the promotion measure.

Performance. The performance measure consisted of a composite rating of five 7-point scale items adapted from a leader effectiveness scale (Tsui, 1984), leader satisfaction scale (Eagly, Makhijani & Klonsky, 1992) and an overall performance evaluation scale (Heilman et al., 2004). Items were, “Please indicate your overall satisfaction with this individual in his/her role.” (*extremely low – extremely high*), “What is your personal view of this individual’s overall effectiveness?” (*not at all effective – extremely effective*), “Overall how would you rate this individual in his/her role?” (*extremely low – extremely high*), “Rate this individual’s potential to excel in his/her career,” (*extremely low – extremely high*), “How successful do you think this individual is in his/her role?” (*extremely successful – extremely successful*) ($\alpha=.90$). See Appendix H2 for the performance measure.

Attribution Measures

The need for more refined measures of attribution, particularly the attribution of others’ as opposed to one’s own behavior, was acknowledged by Swim and Sanna (2002) and Martinko, Douglas, and Harvey (2006, 2007). To address this issue, two scales were created, one scale to represent the construct of *ability* and the other to represent *effort* as defined by Weiner (1985). The following ability and effort scales incorporated the two most commonly mentioned traits describing men and women in a qualitative study that content analyzed over 300 letters of recommendation (Trix & Psenka, 2003). In addition to having construct similarity, Trix et al.’s (2003) study investigated medical faculty at a large academic medical institution who submitted

their materials as part of their promotion application, which is analogous to the organizational context and promotion scenario of the current study.

Ability Attribution. Characteristics of ability were measured by a composite rating of across eight traits on a 9-point likert scale (1 = *not at all*, to 9 = *very much*). Respondents indicated the extent to which they thought the target was: *unparalleled, exceptional, brilliant, gifted, outstanding, excellent, unique, and superb* ($\alpha=.87$). Trix et al., (2003) found that letters of recommendation prepared for male applicants contained many more mentions of “standout adjectives” than letters prepared for female applicants. They describe these “standout adjectives” as descriptive of men’s *abilities* (Trix et al., 2003, pp. 208, 211), which is consistent with Weiner’s (1985) definition of *ability*, which refers to a person’s inherent general intellectual or physical aptitude. See Appendix I1 for the ability measure.

Effort Attribution. Characteristics of effort were measured by a composite rating of across nine traits on a 9-point likert scale (1 = *not at all*, to 9 = *very much*). Respondents indicated the extent to which they thought the target was: *hardworking, diligent, conscientious, dedicated, perfectionistic, dependable, thorough, careful, and meticulous* ($\alpha=.84$). Trix et al., (2003) found that letters of recommendation prepared for female applicants contained many more “grindstone adjectives” than letters prepared for male applicants. They describe these “grindstone adjectives” as “putting one’s shoulder to the grindstone” which is a common metaphor that denotes working extremely hard (Trix et al., 2003, p. 208). These grindstone adjectives are consistent with Weiner’s (1985) definition of *effort*, which refers to how hard an individual appears to work. In fact, Trix et al. (2003) refers to effort attribution versus ability attribution when discussing *grindstone* versus *standout* adjectives, the scholars explain that “there is an insidious gender schema that associates effort with women, and ability with men in

professional areas. According to this schema, women are hard-working because they must compensate for lack of ability,” (p. 208). See Appendix I2 for the effort measure.

Social Dominance Orientation. Social dominance orientation (SDO) is an individual difference variable that was measured. The SDO scale measures a person’s preference for inequality among particular social identity groups (Sidanius & Pratto, 2001; 2004). SDO is considered a personality trait or a disposition and the measure is designed to distinguish individuals who prefer group dominance and inequality (i.e., high SDO) from those who prefer group inclusion and equality (i.e., low SDO). The SDO scale was a validated short form version and consisted of a composite rating of four items (Pratto, Çidam, Stewart, Zeineddine, Aranda, Aiello, Chryssochoou, Cichocka, Cohrs, Durrheim, & Eicher, 2013). Participants responded on a 7-point scale, from 1= extremely oppose to 7=extremely favor. An example item is, “*Superior groups should dominate inferior groups*” ($\alpha=.88$). See Appendix J1 for the SDO measure.

The internal reliability and predictive validity of the four-item short SDO scale was initially tested among 2,130 adults in 20 countries, using 15 languages. This included 153 people from the U.S., 46% of whom were female (Pratto et al., 2013). The short SDO scale has been shown to have predictive validity, as it was found to predict previously confirmed criterion variables of SDO. As the authors expected, the lower the participants scored on the short SDO scale (i.e., they were lower in SDO), the more they expressed (1) support for more women in leadership positions, (2) greater support for aid to the poor, and (3) stronger advocacy for protecting minorities (Pratto et al., 2013). In addition, the authors found the short SDO scale to be internally consistent, with a Cronbach's alpha of .80 (Pratto et al., 2013). Subsequent studies that utilized the short SDO scale established similar results for reliability and predictive validity

(e.g., Caricati, Mancini, & Marletta, 2016; Pratto, Saguy, Stewart, Morselli, Foels, Aiello, Aranda, Cidam, Chrysochoou, Durrheim, & Eicher, 2014).

For the current sample, an examination of the distribution of each of the four SDO items revealed that each distribution was highly skewed toward low SDO. Similar to prior research that included a skewed personality variable (e.g., Shah & Higgins, 1997; Roman, Moskowitz, Stein, & Eisenberg, 1995; Wheeler, Petty, & Bizer, 2005), we conducted a tertiary split on SDO scores. Preacher, Rucker, and Nicewander (2005) purport that the psychological method of splitting a continuous variable into tertiles aids in the interpretation of results, especially in the case of extreme or skewed scores. Typically, the “high” and “low” groups represent a clearer extreme, and allows for a clearer interpretation of its relationship with a continuous outcome variable. See Appendix J2 for distribution of SDO scores by tertile category.

Manipulation Checks

Success of Target. In order to demonstrate that the target was successful in his or her job, target success was assessed by one item which asked, “How successful do you think this individual is in their role?” Participants responded on a 7-point scale from 1=not at all successful to 7=extremely successful. This was the same item as in the performance measure. See Appendix H2, item 1.

Gender of Target. In order to demonstrate that the gender of the target was correctly manipulated, gender was assessed by one item asking them to, “Please indicate the gender of the person you evaluated.” Participants responded on a categorical measure selecting one of three options (female, male, other). See Appendix K, item 2.

Specialty of Target. In order to demonstrate that the specialty of the target was correctly manipulated, specialty was assessed by one item asking them to, “Please indicate the person’s

specialty.” Participants responded on a categorical measure selecting one of three options (pediatrics, surgery, other). See Appendix K, item 3.

Sex-typing of Job. In order to demonstrate that the extent to which the job was perceived as more masculine or feminine, sex-type of the job was assessed by one item which asked, “In general, do you think more men or more women occupy this individual’s specific role?”

Participants responded on a 7-point scale from 1=more men to 7=more women. See Appendix K, item 4.

Demographic Questionnaire

The demographic questionnaire included gender, age, ethnicity, marital status, years of fulltime work, income, occupation, and occupation level. See Appendix L.

CHAPTER FOUR: RESULTS

Manipulation Checks

Success of target. The stimulus materials were designed so that the target was successful in his or her job across all of the conditions in the study. To evaluate whether this manipulation was effective, participants were asked how successful they thought the individual was in this or her role on a seven-point scale (ranging from 1=not at all successful to 7=extremely successful). As expected, participants perceived the target as successful ($M=6.24$, $SD=0.70$) with all participants rating the target between 4 to 7. Thus, the success manipulation was effective, and the targets were perceived as quite successful across experimental conditions.

Sex-typing of job. Job-type was manipulated to be either stereotypically masculine (i.e., surgeon) or stereotypically feminine (i.e., pediatrician). To evaluate whether the manipulation of job-type was effective, participants were asked to rate if the job was occupied by more men or by more women on a seven-point scale (1=more men to 7=more women). As expected, participants in the surgery condition perceived the job as more masculine ($M=1.75$, $SD=1.15$) and participants in the pediatrics condition perceived the job as more feminine ($M=5.47$, $SD=1.74$). This confirmed that job-type was successfully manipulated.

Specialty of target. To assess if the manipulation of physician specialty was effective, we evaluated participants' answers to a categorical measure asking: Please indicate the person's specialty. Most of the participants responded accurately, 99% of those ($N=96$) in the pediatrics condition responded "pediatrics", 97.3% of those ($N=107$) in the surgery condition responded "surgery", and no participants selected "other". This provided support that the physician specialty manipulation was effective.

Gender of target. Lastly, to determine whether the physician gender manipulation was successful, we assessed participants' answers to a categorical measure asking: Please indicate the gender of the person you evaluated. The majority of the participants responded correctly, 95.4% of the participants (N=103) in the female condition responded "female", 94% of those (N=93) in the male condition responded "male", and no participants selected "other". This provided support that the physician gender manipulation was effective.

Preliminary Analyses

Preliminary analyses were conducted to assess whether it was necessary to statistically control for participant gender, age, ethnicity, marital status, years of fulltime work, income, occupation, and occupation level. Becker (2005) recommends that continuous control variables should be selected if they significantly correlate with one of the outcome variables (i.e., performance, promotion, ability attribution, and effort attribution) or if they have been suggested by prior research.

Continuous demographic variables measured in the study were: participant age, years of fulltime work, and income. Correlations and descriptive statistics for the control variables, predictor variables, and outcome variables are reported in Table 2. None of these continuous demographic variables were significantly correlated with any of the independent or outcome variables (see Table M1 in Appendix J for these correlations). Thus, there were no continuous demographic variables included as controls.

Next, a multivariate analysis of variance (MANOVA) including all of the study outcome variables were separately conducted for each of the potential categorical covariates (i.e., participant gender, ethnicity, marital status, education level, occupation, and occupation level). There was an overall significant effect for participant gender, $F(4, 202) = 5.77, p < .0001$, Wilk's

$\lambda = 0.90$. See Table M2 in Appendix M for results of the MANOVA for participant gender.

There were no significant effects for ethnicity, marital status, education level, occupation, and occupation level.

Follow up univariate analyses followed for the significant MANOVA. For participant gender, univariate results supported a significant relationship with the promotion measure ($p < .01$), the effort attribution measure ($p < .0001$), and a marginally significant relationship with the performance measure ($p = .056$). There was no significant relationship with the ability attribution measure ($p = n.s.$). An examination of the means revealed that female participants rated targets higher than male participants on all of these outcome variables. See Table M3 in Appendix M for ANOVA results for participant gender.

Based on these results, participant gender was included as a control variable in the subsequent analyses for promotion, performance, and the effort attribution measure, but not for the ability attribution measure.

Main Analyses

Correlations and descriptive statistics for all study variables are reported in Table 2. See Table 3 for the means and standard deviations for all study conditions (physician gender and specialty), across all of the outcome variables. See Table 4 for the means, standard deviations, and cell sizes for study conditions by participant SDO (physician gender X specialty X participant SDO), across all of the outcome variables.

In order to examine our predictions and research questions, we conducted four univariate 2 (gender) x 2 (specialty) x 3 (participant social dominance orientation or SDO) ANOVAs separately for each of the study outcome variables, controlling for participant gender for promotion, performance, and the effort attribution measure but not for the ability attribution

measure. In order to test the main hypotheses, more focused hypothesis tests were examined. Similar to past research with specific a priori hypotheses (Cuddy, Fiske, & Glick, 2004), within gender comparisons (female pediatrician vs. female surgeon; male pediatrician vs. male surgeon) were conducted using the Fisher least significant difference (LSD) method, with the significance level set at $p < .05$. The description and results of each ANCOVA is included in the following sections, separated by outcome variable.

Lastly, for purposes of clarity and comprehensiveness, all study hypotheses are examined within the two-way and three-way interactions. In other words, to test the main hypotheses, the two-way is examined; and to explore research question one, significant three-ways are examined. It is important to note that the central hypothesized interaction of physician gender and specialty on study outcomes was significantly moderated by participant SDO. Therefore, when there is a significant three-way interaction on a particular outcome variable, study hypotheses will be examined within the three-way interaction as well as the lower-order two-way interaction. The examination of the hypotheses within the two-way interactions will be reported in these results, however the overall interpretation of the analysis will center on the highest order significant interaction, as interpreting lower order interactions is considered less useful (Braumoeller, 2004). See Table 5 for a summary of study hypotheses, research questions, and findings.

Promotion. A univariate 2 (gender) x 2 (specialty) x 3 (SDO) ANCOVA, controlling for participant gender was conducted on recommendations for promotion. As a control variable, participant gender was significant, $F(1, 194) = 9.57, p < .01$. Female participants ($M=6.54$) awarded targets significantly higher recommendations for promotion than male participants ($M=6.25$). There was a significant main effect only for specialty, $F(1, 194) = 6.27, p < .01$. Surgeons ($M=6.50$) received significantly higher recommendations for promotion than

pediatricians (M=6.23). There were no other significant main effects or interactions on promotion. See Table 6 for results of the ANCOVA.

Hypothesis 1a predicted that male targets would be more promotable than female targets. Because there was no significant main effect for gender, hypothesis 1a was not supported. Male and female targets received equally good promotion recommendations (male targets: M=6.42, female targets: M=6.37).

To fully examine hypotheses 1b-1d post-hoc t-tests and intercell contrasts using Fisher's LSD were conducted, despite the non-significant 2 (gender) x 2 (specialty) interaction (Cuddy et al., 2004).

With respect to within gender contrasts, hypothesis 1b predicted that female pediatricians would be more promotable than female surgeons; and hypothesis 1c predicted that male pediatricians and male surgeons would be equally promotable. Contrary to hypothesis 1b, female surgeons received significantly higher promotion recommendations (M=6.50) than female pediatricians (M=6.19). Hypotheses 1b was not supported. However, hypothesis 1c was supported, there was no significant difference in promotion recommendations between male pediatricians (M=6.32) and male surgeons (M=6.52).

With respect to within specialty gender comparisons, hypothesis 1d predicted that differences in promotion would be larger between male and female targets within the surgery condition than between male and female targets in the pediatrics condition. Since there were no significant differences in promotion recommendations between males and females in either surgery (male: M=6.52, female: M=6.50) or pediatrics (male: M=6.32, female: M=6.19), hypotheses 2d was not confirmed.

Lastly, because the 2 (gender) x 2 (specialty) x 3 (SDO) interaction was not significant for promotion, $F(2, 194)=1.80$, $p=n.s.$, no further analyses were conducted to explore whether SDO moderated any of these effects (research question 1).

Performance. A univariate 2 (gender) x 2 (specialty) x 3 (SDO) ANCOVA, controlling for participant gender was conducted on ratings of performance. As a control variable, participant gender was significant, $F(1, 194) = 4.26$, $p<.05$. Female participants ($M=6.43$) awarded targets significantly higher recommendations for promotion than male participants ($M=6.28$). There was a significant main effect for only specialty, $F(1, 194) = 5.80$, $p<.05$. Surgeons ($M=6.44$) received significantly higher performance ratings than pediatricians ($M=6.26$). There were no other significant main effects or two-way interactions. However, this main effect was qualified by a marginally significant three-way interaction, $F(2, 194) = 2.82$, $p = .06$. See Table 7 for results of the ANCOVA.

Hypothesis 2a predicted that female targets would receive higher performance ratings than male targets. Because there was no significant main effect for gender, hypothesis 2a was not supported. Male and female targets received equally good performance ratings (male targets: $M=6.33$, female targets: $M=6.37$).

To examine hypotheses 2b-2d, post-hoc t-tests and intercell contrasts using Fisher's LSD were conducted to examine study hypotheses, despite the non-significant 2 (gender) x 2 (specialty) interaction (Cuddy et al., 2004).

With respect to within gender contrasts, hypothesis 2b predicted that female surgeons would receive higher performance ratings than female pediatricians; and hypothesis 2c predicted that male pediatricians and male surgeons would receive equally good performance ratings. In support of hypothesis 2b, female surgeons ($M=6.47$) received significantly higher performance

ratings than female pediatricians (M=6.23). Further, in support of hypothesis 2c, there was no significant difference in performance ratings between male surgeons (M=6.42) and male pediatricians (M=6.25).

Further, because the above effects were qualified by a marginally significant three-way interaction, further analyses were conducted to explore whether SDO moderated any of these effects (research question 1). In other words, the a priori hypotheses were examined within each level of participant SDO (high, moderate, low). Post-hoc t-tests and intercell contrasts using Fisher's LSD were used to examine the research question and a priori hypotheses. See Figure 7 for graphs of performance means for the three-way interaction and Table 4 for the corresponding means table.

Participants with the lowest levels of SDO (e.g., those who are most egalitarian), rated female surgeons (M=6.72) significantly higher in performance than female pediatricians (M=6.16), supporting our previous findings for hypotheses 2b. However, participants with both moderate and high levels of SDO (e.g., those who are less egalitarian) rated female surgeons and female pediatricians as performing similarly well (*moderate* SDO: female surgeon, M=6.42 female pediatrician, M=6.55; *high* SDO: female surgeon, M=6.24 female pediatrician, M=5.99), failing to support hypothesis 2b. Thus, it appears that hypothesis 2b, predicting that female surgeons would be perceived as higher performing than female pediatricians, is only true for participants with the lowest levels of SDO.

Further, consistent with previous support for hypothesis 2c, there was no significant difference in performance ratings between male surgeons and male pediatricians within all three levels of SDO (*low*: male surgeon, M=6.33 male pediatrician: M=6.31; *moderate*: male surgeon, M=6.45 male pediatrician, M=6.21; and *high*: male surgeon, M=6.46 male pediatrician,

M=6.25). Thus, it appears that, as hypothesized, regardless of level of SDO, male surgeons and male pediatricians are rated as similarly performing.

Looking at within specialty gender differences, consistent with hypothesis 2d, we found that participants with the lowest levels of SDO (e.g., most egalitarian) rated female surgeons (M=6.72) significantly higher in performance than male surgeons (M=6.33). Conversely, participants with moderate levels of SDO (e.g., less egalitarian) rated female pediatricians (M=6.55) marginally higher in performance than male pediatricians (M=6.21). Those with high levels of SDO (e.g., least egalitarian), rated males and females as performing similarly in surgery (male: M=6.46, female: M=6.24) and pediatrics (male: M=6.25, female: M=5.99). Thus, it appears that hypothesis 2d, predicting performance differences as larger between male and female targets within surgery than pediatrics, is only true for participants with the lowest levels of SDO.

Lastly, another way to explore the three-way interaction is to examine the target gender and specialty across participant SDO level (e.g., low versus high SDO, low versus moderate SDO, moderate versus high SDO, etc.). See Figure 8 for graphs of the significant three-way interaction organized by target gender and Table 4 for the corresponding means table. There are no significant differences between all three levels of SDO for male pediatricians (low SDO: M=6.35, moderate SDO: M=6.2, high SDO: M=6.23) or for male surgeons (low SDO: M=6.35, moderate SDO: M=6.48, high SDO: M=6.41). Overall, for male targets, SDO level did not matter for ratings of performance.

However, for female targets, participant SDO did matter. When evaluating female pediatricians, participants with moderate SDO rated targets significantly higher in performance than participants low and high in SDO (low SDO: M=6.17, moderate SDO: M=6.53, high SDO:

M=5.99). When evaluating female surgeons, participants with low SDO rated targets significantly higher than participants with moderate and high SDO (low SDO: M=6.71, moderate SDO: M=6.41, high SDO: M=6.24). Overall, for female targets, participants high in SDO tended to rate female targets lower as compared to the more egalitarian participants (moderate and low SDO).

Ability Characterizations. A univariate 2 (gender) x 2 (specialty) x 3 (SDO) ANOVA was conducted on ability characterizations. There was a marginally significant main effect for participant SDO, $F(2, 195)=2.12, p=.05$. Targets' were described as more "naturally capable" by participants with the lowest levels of SDO (e.g., those who are most egalitarian) (M=5.79), followed by participants with moderate levels of SDO (e.g., those who are less egalitarian) (M=5.63), and least by participants with high levels of SDO (e.g., those who are least egalitarian) (M=5.45). There were no other significant main effects or two-way interactions. However, this main effect was qualified by a marginally significant three-way interaction, $F(2, 195)=2.87, p=.059$. See Table 8 for results of the ANOVA.

Hypothesis 3a predicted that ability would be attributed more to male targets than female targets. Because there was no significant main effect for gender, hypothesis 3a was not supported. Participants described male and female targets as equally capable (male targets: M=5.64, female targets: M=5.63).

To fully examine hypotheses 3b-3d, post-hoc t-tests and intercell contrasts using Fisher's LSD were conducted to examine study hypotheses, despite the non-significant 2 (gender) x 2 (specialty) interaction (Cuddy et al., 2004).

With respect to within gender contrasts, hypothesis 3b predicted that ability would be attributed more to female pediatricians than female surgeons; and hypothesis 3c predicted that

ability would be equally assigned to male pediatricians and male surgeons. Contrary to hypothesis 3b, female pediatricians (M=5.51) and female surgeons (M=5.69) were similarly perceived as capable. In support of hypothesis 3c, there was no significant difference in ability between male surgeons (M=5.72) and male pediatricians (M=5.57).

With respect to across gender comparisons, hypothesis 3d predicted that differences in ability would be larger between male and female targets within the surgery condition than between male and female targets in the pediatrics condition. Since there were no significant differences in ability characterizations between males and females in either surgery (male: M=5.72, female: M=5.69), or pediatrics (male: M=5.57, female: M=5.51), hypothesis 3d was not confirmed.

Further, because the above effects were qualified by a marginally significant 2 (gender) x 2 (specialty) x 3 (SDO) interaction, $F(2, 195) = 2.87, p = .059$, further analyses were conducted to explore whether SDO moderated any of these effects (research question 1). In other words, the a priori hypotheses were examined within each level of participant SDO (high, moderate, low). Post-hoc t-tests and intercell contrasts using Fisher's LSD were used to examine the research question. See Figure 9 for graphs of ability attribution means for the three-way interaction and Table 4 for the corresponding means table. Participants with the lowest levels of SDO (e.g., those who are most egalitarian), described female surgeons (M=6.01) as more capable than female pediatricians (M=5.50), reversing predictions in hypothesis 3b. Further, participants with both moderate and high levels of SDO (e.g. those who are less egalitarian) perceived female surgeons and female pediatricians as equally capable (*moderate* SDO: female surgeon, M=5.58 female pediatrician, M=5.83; and *high* SDO: female surgeon, M=5.44 female pediatrician, M=5.25), also failing to support Hypothesis 3b. Thus, it appears that hypothesis 3b, predicting

ability to be attributed more to female pediatricians than female surgeons, is not supported. However, the opposite pattern (female surgeons receive greater ability descriptions than female pediatricians) was found for participants with the lowest levels of SDO.

Further, consistent with previous support for hypothesis 3c, there was no significant difference in ability ratings between male surgeons and male pediatricians across all three levels of SDO (*low*: male surgeon, M=5.69 male pediatrician: M=5.85; *moderate*: male surgeon, M=5.67 male pediatrician, M=5.46; and *high*: male surgeon, M=5.81 male pediatrician, M=5.39). Thus, it appears that regardless of SDO level, male surgeons and male pediatricians are seen as similarly capable.

Looking at within specialty gender differences, in contrast with hypothesis 3d, we found that participants across all three levels of SDO perceived male and female physicians as equally capable within surgery (*low*: male, M=5.69 female: M=6.01; *moderate*: male, M=5.67 female, M=5.58; and *high*: male, M=5.81 female, M=5.44) and within pediatrics (*low*: male, M=5.85 female: M=5.50; *moderate*: male, M=5.46 female, M=5.83; and *high*: male, M=5.39 female, M=5.25). Thus, it appears that regardless of level of SDO, participants perceived male and female targets within the same specialty as equally capable.

Lastly, another way to explore the three-way interaction is to examine the target gender and specialty across participant SDO level (e.g., low versus high SDO, low versus moderate SDO, moderate versus high SDO, etc.). See Figure 10 for graphs of the significant three-way interaction organized by target gender and Table 4 for the corresponding means table. For male pediatricians, there is only one significant difference between all three levels of participant SDO; participants low in SDO perceived male targets as more capable than participants high in SDO (low SDO: M=5.88, moderate SDO: M=5.45, high SDO: M=5.37). Male surgeons were rated

similarly in ability across the three levels of SDO (low SDO: $M=5.70$, moderate SDO: $M=5.69$, high SDO: $M=5.77$). Overall, SDO level did not matter for ratings of ability for male surgeons.

However, for female targets in both specialties, participant SDO did matter. When evaluating female pediatricians, participants with moderate SDO rated targets significantly higher in ability than participants high in SDO (low SDO: $M=5.51$, moderate SDO: $M=5.81$, high SDO: $M=5.25$). When evaluating female surgeons, participants with low SDO rated targets significantly higher than participants with moderate and high SDO (low SDO: $M=6.01$, moderate SDO: $M=5.57$, high SDO: $M=5.45$). Overall, for female targets, participants high in SDO tended to rate female targets lower in ability as compared to the more egalitarian participants (moderate and low SDO).

Effort Characterizations. A univariate 2 (gender) x 2 (specialty) x 3 (SDO) ANCOVA, controlling for participant gender was conducted on effort characterizations. As a control variable, participant gender was significant, $F(1, 194) = 15.15, p < .0001$. Female participants ($M=6.28$) characterized targets significantly higher in effort than male participants ($M=5.91$). There was a significant main effect for specialty, $F(1, 194) = 15.15, p < .0001$, and for SDO, $F(2, 194) = 13.84, p < .05$. Surgeons' ($M=6.24$) were rated higher in effort characterizations than pediatricians ($M=5.93$). Further, targets were seen as hardest working (i.e., effort) by participants with the lowest levels of SDO (e.g., those who are most egalitarian) ($M=6.26$), followed by participants with moderate levels of SDO (e.g., those who are less egalitarian) ($M=6.12$), and least by participants with high levels of SDO (e.g., those who are least egalitarian) ($M=5.87$). There were no other significant main effects or two-way interactions. However, these main effects were qualified by a significant three-way interaction, $F(2, 194) = 4.10, p < .05$. See Table 9 for results of the ANCOVA.

Hypothesis 4a predicted that effort would be attributed more to female targets than male targets. Because there was no significant main effect for gender, hypothesis 4a was not supported. Participants rated male and female targets as similarly hard working (male targets: $M=6.09$, female targets: $M=6.10$).

To fully examine hypotheses 4b-4d, post-hoc t-tests and intercell contrasts using Fisher's LSD were conducted to examine study hypotheses, despite the non-significant 2 (gender) x 2 (specialty) interaction (Cuddy et al., 2004).

With respect to within gender contrasts, hypothesis 4b predicted that effort would be attributed more to female surgeons than female pediatricians; and hypothesis 4c predicted that effort would be equally assigned to male pediatricians and male surgeons. In support of hypothesis 4b, female surgeons ($M=6.24$) were seen as harder working than female pediatricians ($M=5.88$). In contrast to hypothesis 4c, male surgeons ($M=6.24$) were seen as harder working than male pediatricians ($M=5.93$).

With respect to across gender comparisons, hypothesis 4d predicted that differences in effort ratings would be larger between male and female targets within the surgery condition than between male and female targets in the pediatrics condition. Since there were no significant differences in effort ratings between males and females in either surgery (male: $M=6.24$, female: $M=6.23$) or pediatrics (male: $M=5.93$, female: $M=5.88$), hypotheses 4d was not confirmed.

Further, because the above effects were qualified by a significant 2 (gender) x 2 (specialty) x 3 (SDO) interaction, $F(2, 194) = 4.10$, $p < .05$, further analyses were conducted to explore whether SDO moderated any of these effects (research question 1). In other words, the a priori hypotheses were examined within each level of participant SDO (high, moderate, low). Post-hoc t-tests and intercell contrasts using Fisher's LSD were used to examine the research

question. See Figure 11 for graphs of means for the three-way interaction and Table 4 for the corresponding means table.

Participants with the lowest and highest levels of SDO (e.g., those who are most and least egalitarian, respectively), perceived female surgeons (low SDO: $M=6.51$, high SDO: $M=6.06$) as harder working than female pediatricians (low SDO: $M=5.91$, high SDO: $M=5.50$), supporting our previous findings for hypothesis 4b. However, participants with moderate levels of SDO perceived female surgeons ($M=6.12$) and female pediatricians ($M=6.27$) as equally hard working. Thus, it appears that Hypothesis 4b, predicting that female surgeons would receive higher effort ratings than female pediatricians, is supported by participants with the lowest and highest levels of SDO.

Further, in contrast to hypothesis 4c, participants with higher levels of SDO saw male surgeons (moderate SDO: $M=6.26$, high SDO: $M=6.28$) as harder working than male pediatricians (moderate SDO: $M=5.82$, high SDO: $M=5.77$). However, in support of hypothesis 4c, participants with the lowest levels of SDO rated male surgeons ($M=6.22$) and male pediatricians ($M=6.29$) as equally hard working. Thus, it appears that hypothesis 4c, predicting no difference in effort ratings between male surgeons and male pediatricians, is supported only by participants with the lowest levels of SDO.

Looking at within specialty gender differences, in support of hypothesis 4d, we found that participants with the lowest levels of SDO (e.g., most egalitarian) rated female surgeons ($M=6.50$) as harder working than male surgeons ($M=6.23$). Conversely, participants with moderate levels of SDO (e.g., less egalitarian) saw female pediatricians ($M=6.24$) as harder working than male pediatricians ($M=5.79$). There were no cross gender differences for those with the highest levels of SDO (e.g., least egalitarian) for targets within surgery (male, $M=6.18$

female, M=6.07) and within pediatrics (male, M=5.73 female, M=5.49). Thus, it appears that hypothesis 4d, predicting larger effort rating differences for targets within surgery than within pediatrics, is supported only by participants with the lowest levels of SDO.

Lastly, another way to explore the three-way interaction is to examine target gender and specialty across participant SDO level (e.g., low versus high SDO, low versus moderate SDO, moderate versus high SDO, etc.). See Figure 12 for graphs of the significant three-way interaction organized by target gender and Table 4 for the corresponding means table. For male pediatricians, participants low in SDO perceived targets as harder working than participants moderate and high in SDO (low SDO: M=6.29, moderate SDO: M=5.79, high SDO: M=5.73). Male surgeons were rated similarly in effort across the three levels of SDO (low SDO: M=6.22, moderate SDO: M=6.32, high SDO: M=6.18). Overall, SDO level did not matter for ratings of effort for male surgeons.

However, for female targets in both specialties, participant SDO did matter. When evaluating female pediatricians, participants with moderate SDO rated targets higher in effort than participants low and high in SDO; further, participants low in SDO rated targets higher in effort than those high in SDO (low SDO: M=5.94, moderate SDO: M=6.24, high SDO: M=5.49). When evaluating female surgeons, participants with low SDO rated targets significantly higher than participants with moderate and high SDO (low SDO: M=6.50, moderate SDO: M=6.11, high SDO: M=6.07). Overall, for female targets, participants high in SDO tended to rate female targets lower in effort as compared to the more egalitarian participants (moderate and low SDO).

Effort and Ability Characterizations as a Predictors. Hypotheses 5a predicted that effort characterizations would be a stronger predictor of performance than ability characterizations, and targets with higher effort ratings would be perceived as better performers than targets with lower

effort ratings. Conversely, hypothesis 5b predicted that ability characterizations would be a stronger predictor of promotion than effort characterizations, and targets with higher ability ratings would be perceived as more promotable than targets with lower ability ratings. To examine these hypotheses, two hierarchical multiple regression analyses were conducted with the ability and effort scales as continuous predictors (previously included as outcomes), controlling for participant gender. The first regression included performance as the outcome and the second included promotion as the outcome. For both regression models, step 1 included participant gender and step 2 included ability and effort attributions.

Results from the first hierarchical multiple regression support that the model (ability, effort, and participant gender) significantly predicted promotion, $F(3, 203) = 48.33, p < .001, R^2 = .42$. In step 1, participant gender contributed marginally significantly to the prediction, an examination of the means support that female participants rated targets higher on promotion as compared to male participants. In step 2, however, contrary to hypothesis 5b, only effort ratings ($p < .001$) added significantly to the prediction. Examination of the coefficients supported a significant positive relationship between effort ratings and promotion, such that targets with higher effort ratings have higher promotion recommendations. Further, predictor coefficients support that effort characterizations ($\beta = 0.63$) are a much stronger predictor of promotion than ability characterizations ($\beta = 0.01$). Hypothesis 5b was not supported. See Table 10 for the regression analysis on promotion.

Results from the second hierarchical multiple regression support that the model (ability, effort, and participant gender) significantly predicted performance, $F(3, 203) = 47.78, p < .0001, R^2 = .41$. In step 1, participant gender contributed marginally significantly to the prediction, an examination of the means support that female participants rated targets higher on

performance as compared to male participants. In step 2, both ability ($p < .05$) and effort ($p < .001$) ratings added significantly to the prediction. Examination of the coefficients support a significant positive relationship between both ability and effort ratings on evaluation of performance. Specifically, targets with higher ability and effort ratings have higher performance ratings. Further, in support of hypothesis 5a, predictor coefficients supported that effort characterizations ($\beta = 0.48$) are a stronger predictor of performance than ability characterizations ($\beta = 0.15$). See Table 11 for the regression analysis on performance.

Thus, it appears that characterizations of effort are a stronger predictor of both performance and promotion recommendations than characterizations of ability.

Mediational Analysis

Research question two proposed examining ability and effort as success attributions that mediate the relationship between study predictors (gender, specialty, and participant SDO) and outcomes (performance and promotion). To explore this, Baron and Kenny's (1986) three-step procedure for assessing mediated effects using regression was performed.

Linear regression analysis requires that all variables included in the model be either continuous or dichotomous (Jaccard & Turrisi, 2003; Stockburger, 1998). Statisticians support that the simplest and most interpretable way to handle a three level categorical variable in regression is to convert the variable into two dichotomous variables (Stockburger, 1998). "In general, a categorical variable with k levels should be transformed into $k-1$ variables each with two levels," (Stockburger, 1998, p.2). Accordingly, we converted the three level participant SDO variable into two dichotomous variables in order to include SDO in the regression analysis. Further, when a categorical variable has more than 2 levels, then more than one code variable must be built into the regression equation to fully represent the single categorical variable; this

includes all interaction variables contained in the equation (West, Aiken, & Krull, 1996). See Table M4 in Appendix M for the dummy coding scheme.

For the present set of regression analyses, the comparison group chosen was the low SDO group as prior analyses supported that many of the significant and meaningful differences involve participants low in SDO. Additionally, the low SDO group was selected because scholars recommend that the comparison group is the group expected to “score lowest or highest on the dependent variable,” (West et al., 1996). Based on this, in the following analysis, two dummy variables are built into the equation that comprise participant SDO: (1) dummy comparing low to high SDO participants; and (2) comparing low to moderate SDO participants.

The first condition for establishing mediation using the Baron and Kenny (1986) method is determining that the highest level interaction of the independent variables (in this case: (1) gender X specialty X low vs. high SDO; and (2) gender X specialty X low vs. moderate SDO) significantly predict the outcome variable. Since we previously found a significant interaction effect only for the performance rating outcome, we assessed a possible mediation effect for performance, but not promotion. Further, the test of mediation was conducted twice, first testing effort attributions as a possible mediator, then ability attributions.

Mediation analyses are presented in Tables 12 and 13. Baron et al.’s (1986) three step procedure is outlined here. First, the predictor variable ((1) gender X specialty X low vs. high SDO; and (2) gender X specialty X low vs. moderate SDO) should be significantly related to the outcome variable (performance rating). Second, the predictor variable ((1) gender X specialty X low vs. high SDO; and (2) gender X specialty X low vs. moderate SDO) should be significantly related to the mediator (effort, ability). Third, the mediating variable (effort, ability) should be

related to the outcome variable (performance) with the predictor ((1) gender X specialty X low vs. high SDO; and (2) gender X specialty X low vs. moderate SDO) included in the equation.

If these three conditions are met, partial mediation is present. Complete mediation is present if the predictor variable has a non-significant beta weight in the third step of the regression model. Lastly, the Sobel (1982) test was used to assess whether the effect of the predictor on the outcome variable was significantly reduced by the inclusion of the mediating variable (Kenny, 2009). The Sobel test was conducted using Preacher and Leonardelli's (2006) interactive calculation tool.

In order to test for mediation, first the performance ratings measure was simultaneously regressed on (1) *gender X specialty X low vs. high SDO*; and (2) *gender X specialty X low vs. moderate SDO*. The effect of the *gender X specialty X low vs. high SDO* term on performance ratings was significant, $\beta = .40$, $t(193) = 2.37$, $p < .05$. However, the *gender X specialty X low vs. moderate SDO* term was not, $\beta = .220$, $t(193) = 1.30$, $p = n.s.$ These findings satisfied the first condition of mediation for the interaction effect including the *gender X specialty X low vs. high SDO* predictor. Thus, only the term including *low vs. high SDO* was explored as a possible predictor in the following analysis.

For the second step, effort and ability attributions were separately regressed on *gender X specialty X low vs. high SDO*. The effect of *gender X specialty X low vs. high SDO* on effort and ability attributions was significant (effort: $\beta = .46$, $t(194) = 2.86$, $p < .01$; ability: $\beta = .39$, $t(194) = 2.24$, $p < .05$). However, the overall Model F at this step was significant only for effort attributions ($F(12, 194) = 4.55$, $p < .001$), and not for ability attributions ($F(12, 194) = 1.64$, $p = n.s.$), therefore satisfying the second condition of mediation for effort attributions and only for the

interaction effect including the *low vs. high SDO*. Thus, only the effort attributions measure was further explored as a possible mediator.

Third, performance ratings were regressed on *gender X specialty X low vs. high SDO*, with effort attributions included in the regression equation. Effort attributions also significantly predicted performance ratings (effort: $\beta = .62$, $t(193) = 2.86$, $p < .0001$); and *gender X specialty X low vs. high SDO* was rendered non-significant, $\beta = .12$, $t(193) = .85$, $p = n.s.$ The reduction in beta weight associated with *gender X specialty X low vs. high SDO* when effort attributions was not in the equation ($\beta = .461$, $t(206) = 2.86$, $p < .01$) compared to when it was ($\beta = .12$, $t(193) = .85$, $p = n.s.$), suggests a full mediation effect for effort attributions on the performance ratings outcome. Lastly, the Sobel test revealed a significant reduction in the effect of *gender X specialty X low vs. high SDO* on performance ratings, supporting a mediation role for effort attributions (Sobel = 2.75, $p < .01$).

The mediation analysis support that higher effort attributions lead to higher performance ratings, particularly for participants low in SDO compared to participants high in SDO when evaluating female surgeons. An examination of means support these findings. The only condition out of the four experimental conditions (female pediatrics, female surgery, male pediatrics, male surgery) that had significantly different means on the performance outcome ($p < .05$) between participants with low SDO and high SDO were female targets in the surgery condition. Results appear to indicate that the most egalitarian participants (low SDO) described the success of female targets in the surgery condition as significantly more due to her hard work (i.e., effort) as compared to the least egalitarian participants (low SDO: $M = 6.51$; high SDO: $M = 6.06$, $p < .05$). Further, characterizations of effort for female surgeons were not significantly different for participants low and moderate in SDO ($p = n.s.$) or for participants high and moderate in SDO

($p=n.s.$). For the most egalitarian participants (low SDO), these higher effort attributions lead to higher performance ratings for female surgeons.

In further support of this mediation effect, participants low in SDO rated female targets in the surgery condition as significantly higher in performance as compared to participants high in SDO (low SDO: $m=6.72$; high SDO: $m=6.24$, $p<.01$). Performance ratings for female surgeons were not significantly different for participants low and moderate in SDO ($p=n.s.$) or for participants high and moderate in SDO ($p=n.s.$).

Summary of Findings

In support of within gender hypotheses 2b and 4b, female surgeons were seen as better performers and had higher effort ratings than female pediatricians. Further in support of within gender hypotheses 1c, 2c, 3c, and 4c, male surgeons and male pediatricians were seen as equal on all outcome variables (promotion, performance, ability characterizations, and effort characterizations, respectively).

Contrary to hypotheses 1a, 2a, 3a, and 4a, male and female physicians were seen as equal on all outcome variables (promotion, performance, ability characterizations, and effort characterizations, respectively). Further, contrary to within gender hypotheses 1b and 3b, female surgeons were seen as more promotable and more capable than female pediatricians. Lastly, contrary to hypotheses 1d, 2d, 3d, and 4d, there were no within specialty gender differences; males and females within pediatrics as well as males and female within surgery were seen as equal on all outcome variables (promotion, performance, ability characterizations, and effort characterizations, respectively).

However, these findings appear to be moderated by participant SDO level (research question 1) for evaluations of performance, ability characterizations, and effort characterizations.

Participants low in SDO (i.e., most egalitarian) consistently saw female surgeons as better performing, more brilliant, and harder working than female pediatricians (hypotheses 2b, 3b, and 4b, respectively). Further, in support of hypotheses 2d and 4d, only those low in SDO (i.e., most egalitarian) saw female surgeons as better performers and harder working than male surgeons, but equally to female and males in pediatrics.

There were some, but fewer effects present for participants with moderate SDO levels (i.e., those less egalitarian). Those with moderate SDO levels (i.e., less egalitarian) had a significant and opposite result for performance and effort ratings as compared to those with low SDO. Contrary to within specialty hypotheses 2d and 4d, participants with moderate SDO saw female pediatricians as better performing and harder working than male pediatricians. Female and male surgeons were seen as equal on these outcomes.

The only significant effects for participants with high SDO levels (i.e., those least egalitarian) were on effort ratings. There was a significant main effect for target specialty, such that those high in SDO (i.e., least egalitarian) attributed effort more to males in surgery than males in pediatrics (contrasting hypothesis 4c), as well as females in surgery than females in pediatrics (supporting hypothesis 4b). It appeared that those with high SDO (i.e., least egalitarian) favored targets in surgery over pediatrics.

Overall, although most of the significant effects appear to be in the group with low SDO (i.e., those most egalitarian), it also appears that those high in SDO (i.e., least egalitarian) rated female targets consistently lower on all four of the study outcomes as compared to the other SDO groups. The graphs in Figures 8, 10, and 12 help elucidate this pattern. Specifically, (though not significantly different) the high SDO group (i.e., least egalitarian) was the only SDO group that consistently rated female pediatricians and female surgeons less well than male targets. The other

two SDO groups (low and moderate) consistently rated at least one female group (pediatricians or surgeons) as higher than the corresponding male group.

Lastly, in summary and in response to research question two (attributions as a mediator), results support that effort characterizations likely mediate the relationship between *gender X specialty X low vs. high SDO* and performance ratings. Particularly within the female surgery condition, participants low in SDO (most egalitarian) attributed higher levels of effort to targets as compared to participants high in SDO (least egalitarian). In turn, these higher effort attributions had positive implications for female surgeons' performance ratings.

CHAPTER FIVE: DISCUSSION

Overview

The primary goal of this study was to examine whether the gender of a successful physician leader impacted decision makers' allocation of workplace outcomes such as promotion and performance evaluation ratings. This study intended to supplement the discourse on gender bias because current explanations of the scarcity of female physicians in leadership positions and stereotypically highly masculine specialties (i.e., surgery) are largely limited to “blame-the-women” justifications, outlining either that women opt out of choosing more competitive positions due to lack of work-life balance or that women are deficient in competitive characteristics needed to succeed in those positions (Alers et al., 2014; Carli, 2006, 2016; Eagly et al, 2002; Ely et al., 2000, 2011). However, the surge of extant research and theory focused on gender and leadership have provided overwhelming evidence of another explanation, that is, gender bias resulting from interactions of gender stereotypes and the characteristics of social contexts where women seek leadership positions, namely in male-typed contexts (Eagly & Heliman, 2016). Surprisingly, despite this surge, few investigations include an examination of the extremely masculine context of physician leadership – the focal setting of this study. Uncovering the stereotypic process underlying gender bias is helpful in understanding what women face, because it reveals implicit patterns projected onto women simply because of their gender. If these patterns are understood, organizational members may become more likely to decrease engaging in inaccurate stereotype based judgments, and women may reduce their tendency to blame themselves for difficulty in obtaining leadership positions.

This study extends research demonstrating gender stereotyping as a barrier to female leaders by focusing on a specific case of male and female physician leaders who have clearly

exhibited success on the job. The specific goals of this work were to (1) examine whether the degree of “male-ness” of the context influenced the effect of physician gender on the allocation of workplace outcomes; (2) explore whether different types of success attributions influenced the allocation of workplace outcomes; and (3) investigate whether participants’ level of egalitarianism impacted these relationships. Correspondingly, the study design manipulated physician gender and degree of “male-ness” of specialty to examine the effects of these variables on promotion recommendations, evaluation of performance, and ratings of the physician’s brilliance and hard work.

Overall, the magnitude of gender stereotyping should still be considered in light of the identical nature of stimulus materials provided to participants to make decisions. In other words, each physician profile contained exactly the same credentials and were identically successful in their jobs. The only differences aside from participant traits were target gender and the degree of masculinity of the job. Therefore, there is evidence that participants relied on gender stereotypes to inform their decisions above and beyond the balanced set of materials provided. What follows is a more detailed review of the study findings in consideration of prior research, a discussion of study limitations, and an exploration of future research directions and practical implications of the work.

Discussion and Interpretation of Results

Influence of Target Specialty and Participant SDO on Target Gender and Study Outcomes. Target specialty and participant SDO were examined as moderators to the relationship between target gender on promotion recommendations, performance ratings, ability characterizations, and effort characterizations. In line with the theoretical link predicting that (1) a higher degree of gender congruence to a male-typed job will result in increased promotion

recommendations and more attributions of ability; and that (2) a lower degree of gender congruence to a male-typed job will result in increased performance ratings and more attributions of effort; results pertaining to promotion and ability are jointly discussed followed by a joint discussion of performance and effort. The findings are presented in this way because this study posited an identical pattern of results for the effect of study predictors on promotion and ability outcomes and a countering but still identical pattern of results for the effect of study predictors on performance and effort outcomes. As previously discussed, these patterns were based on extant theory and research on gender/role congruity and attribution theory (i.e., Lyness et al., 2006; Swim et al, 1996).

Promotion and Ability Outcomes. The results for promotion and ability largely countered study predictions, such that instead of receiving penalties due to heightened incongruity to the job, female surgeons received a boost in both promotion recommendations and characterizations of brilliance. All participants, regardless of SDO level, awarded significantly higher promotion recommendations to female surgeons as compared to female pediatricians; these promotion recommendations were equally as good as those awarded to male surgeons. Similarly, regardless of SDO level, participants rated female surgeons as equally brilliant as male surgeons; and female pediatricians as equally brilliant as male pediatricians. However, it is important to note that only the most egalitarian participants perceived female surgeons as more brilliant than female pediatricians; those less egalitarian perceived female physicians as equal in ability. These results contrasted previous research finding that male leaders were awarded with higher rates of promotions than female leaders (Lyness et al., 2006). These results also contrasted meta-analytic findings that attributions of ability were reserved for successful male targets (Swim et al., 1996).

The only set of findings that were consistent with promotion and ability predictions was that the degree of masculinity of the job would not matter for male physicians in terms of promotion allocation and characteristics of brilliance. In other words, male physicians would receive equally good promotion and ability outcomes no matter what specialty they were in. As predicted, hypotheses were supported across all levels of participant SDO, such that male surgeons and male pediatricians were seen as equally promotable and equally as brilliant. Specific interpretations of these outcomes are jointly discussed with the outcomes of the following section.

Performance and Effort Outcomes. Results confirmed performance and effort predictions only for the most egalitarian participants (those lowest in SDO), indicating that target specialty moderated the relationship between target gender and evaluations of performance and effort, such that female physicians were evaluated as higher performing and harder working than male physicians in the surgery condition, but not in the pediatrics condition. Consistent with study predictions, female surgeons were evaluated as significantly higher performing and harder working than female pediatricians, while male physicians were evaluated as equivalent on performance and effort across both specialties. In other words, participants awarded female physicians a boost in evaluations of performance and effort when females were surgeons, and male physicians were evaluated equally well as surgeons and pediatricians. Further, male and female pediatricians were equally evaluated on performance and effort; while female surgeons were evaluated as significantly higher performing and harder working than male surgeons. Lastly, although it is clear that females receive a boost when in the surgery condition, there was no significant main effect for gender such that male and female physicians overall were evaluated as equally good performers and equally hard working.

These results for participants low in SDO are consistent with prior findings that identified the phenomenon in which successful women were awarded with higher evaluations of performance than men in male-typed jobs. Research conducted by Lyness et al. (2006) and Eagly et al. (2002) found that successful female leaders were assigned higher performance evaluation ratings as compared to equally successful male leaders in male-typed jobs. These results are also in line with previous research that women's success in male-typed jobs are characterized by how hard they worked, thus receiving more effort attributions for success than equivalent men (Swim et al., 1989, 1996).

Taken together, these studies suggest that for women, the *less congruent* a target's gender is to the sex-type of the job, the higher the evaluations of performance and attributions of effort are for the successful target. Conversely, *more congruent* the target's gender is to the sex-type of the job, the higher the recommendations for promotion and attributions of ability for the successful target. Adding to lack of fit predictions (Heilman, 1983) and gender congruity theory (Eagly et al., 2002) that outline an automatic stereotype-based preference for men in male-typed jobs, these studies provide evidence of a boost for successful female leaders in jobs that have a lower degree of congruency to their gender, especially for those low in SDO. In line with previous research, the present study found evidence of a boost for successful female targets in the less congruent sex-typed job (i.e., surgery), but not in the more congruent sex-typed job (i.e., pediatrics). In other words, for participants low in SDO, because female targets were highly incongruent to the highly masculine job of physician leader in the surgery department, they were awarded a boost in the form of higher evaluations of performance and characterized as a harder worker as compared to those more congruent to sex-type of their jobs, in this case females in pediatrics and males in surgery.

Results for less egalitarian participants (those with moderate SDO) failed to support study predictions on evaluations of performance and effort. Participants moderate in SDO evaluated male and female physicians overall and as surgeons as equally well performing and hard working. However, results revealed with marginal significance that female pediatricians were evaluated as better performing and harder working than male pediatricians, suggesting that participants moderate in SDO gave a slight penalty only to male pediatricians.

Results for the least egalitarian participants (those with high SDO) similarly failed to support study predictions on evaluations of performance and effort. Specifically, for performance ratings, male physicians were evaluated as better performing, but equally as hard working as compared to female physicians; this was true for both pediatricians and surgeons. For effort ratings, female surgeons and male surgeons were evaluated as significantly harder working than female pediatricians and male pediatricians, respectively. In other words, participants high in SDO showed a preference for male physicians when evaluating performance as well as a preference for surgeons when determining effort.

While study predictions for gender and specialty on evaluations of performance and effort were supported largely by the results of most egalitarian participants, they were generally not supported by the less egalitarian participants. The most egalitarian participants (low SDO) awarded female surgeons with a boost in performance ratings and characterizations of hard work. On the other hand, the least egalitarian participants (high SDO) reserved a slight boost for male physicians overall in terms of higher evaluations of performance. Those highest in SDO also awarded a significant boost for targets in the surgery condition overall, with higher evaluations of performance and higher characterizations of effort. Consequently, these high SDO participants assigned lower performance scores to female physicians and lower performance scores and effort

characterizations to targets in the pediatrics condition, even though all targets were identically successful in job status and achievement.

Considering findings for all four outcomes, the results for participants high in SDO are consistent with prior scholarship that identified trait level antecedents when examining people's responses to gender and sex-type job incongruities. Social dominance theory outlines that those high in SDO hold a strong desire to maintain current power hierarchies based on social identity group membership (Sidanius et al., 1999). Previous research has shown that participants higher in SDO are more likely than participants lower in SDO to demonstrate a preference for higher status groups such as men, and bias against low-status groups such as women (Crandall, 1994; Eisenman, 1991; Federico & Sidanius, 2002; Hoyt, 2012; Sidanius et al., 1996). For example, Hoyt (2012) found that for equivalent men and women in a male-typed job, participants low in SDO demonstrated a “pro-female” bias while participants high in SDO demonstrated an “anti-female” bias. Further, Pratto et al. (1997) found these biases extend to value judgements in job roles, such that jobs seen as more masculine (i.e., police officers and business executives) hold more value than jobs seen as more feminine (i.e., social workers and charity workers). These scholars assert that individual level traits such as participant SDO can strongly influence the typical “gender-matching heuristic that matches the gender-type of the applicant with that of the job,” (Pratto et al., 1997, p. 47). Thus, the interpretation of the boost in performance and effort outcomes for male physicians and surgeons can be explained by the desire for participants high in SDO to maintain power for more dominant social groups (men) and for more valued jobs (surgeon).

Lastly, these findings support that participant SDO and the degree of job mattered more in the evaluation of female physicians than of male physicians, especially on evaluations of

performance and characterizations of being a hard worker. Recall the overall results for male targets: across all levels of participants SDO, male surgeons and male pediatricians were evaluated as equally promotable, appraised as equally good performers, seen as equally brilliant, and as equivalently hard working. Conversely, female physicians received significantly better performance and effort outcomes when they were in the surgery condition and only by participants low in SDO. Moreover, these findings provide support for the proposed conceptual link between performance and effort characterizations, especially for female physicians in the job with a higher degree of masculinity.

In summary, results generally support that dispositional traits (i.e., participant SDO) as well as contextual traits (i.e., target specialty) mattered when examining evaluators' responses to target gender and job-type incongruities. Overall, participants awarded a boost to female surgeons across all four outcomes, assigning them higher promotion recommendations, higher evaluations of performance, and described them as more capable and harder working than female pediatricians and male physicians overall. However, this boost for female surgeons are largely limited to the most egalitarian participants (i.e., those low in SDO). One interpretation of this unexpectedly positive set of results for females in the most masculine specialty, is perhaps due to a heightened awareness of the difficulties women face by those low in SDO. Individuals low in SDO may be more aware of the gender based struggles that women encounter especially the more masculine a context becomes. Accordingly, low SDO participants awarded a boost to females in surgery because they associated more gender incongruity to the job with more gender bias on the job, thus overly compensating to offset the typical bias against women in male-typed contexts. This interpretation is consistent with extant research on gender and social dominance orientation, such that people low in social dominance orientation are more politically liberal

(Pratto et al., 1997b), hold less rigid ideologies about social hierarchy (Pratto et al., 1997a), and tend to behave consistently with egalitarian values (Dasgupta, 2004). Further, this interpretation is consistent with Hoyt's (2012) finding that those low in SDO exhibited a "pro-female" bias.

Additionally, Biernat and Mannis' (1994) shifting standards model provides two possible explanations for these results. First, the scholars suggest that when making judgments on stereotyped social groups, people tend to make *within-category* comparisons. Specifically, target females are compared to stereotypes of "women in general" and target males are compared to stereotypes of "men in general". Given the widely held stereotype of men as better leaders than women; when faced with making judgements for a target female in a male-typed leadership job, people will compare the female target to "women in general," thus a lower standard of leadership competence relative to "men in general" because of higher competence standards for men (Biernat et al., 1994; Biernat & Fuegen, 2001). This results in higher leadership competence ratings for female targets as compared to male targets in the same male-typed leadership job.

Further, Biernat et al. (1994) suggest that in male-typed jobs, people assume that it is easier for men to be successful at the job than it is for women (i.e., women have to work harder to achieve the same level of success). Therefore, when a woman occupies a leadership position in a male-typed job, she may be automatically seen as highly competent as a result of having met the high standard to occupy the masculine role. Further, this effect could be heightened with a higher degree of masculinity of the male-typed job (i.e., physician leader in the surgery department vs. physician leader in the pediatrics department) (Perry et al., 1994). Based on shifting standards theory, occupying and being successful in the highly male-typed job of a physician leader in the surgery department, was more than enough to be an exception to the typical gender congruity rule. Given this line of reasoning, female surgeons were seen as

exceptional, and thus awarded with higher promotion recommendations, evaluated as higher performing as well as more brilliant and harder working than female pediatricians and for some outcomes, male physicians. However, it is critical to remember that only the most egalitarian participants perceived female surgeons as exceptional women.

Although, the least egalitarian participants demonstrated a bias for male physicians and surgeons overall, awarding them higher evaluations of performance and describing them as harder working than female physicians and pediatricians overall; dispositional traits and contextual traits mattered less for male targets than female targets. In general, results support no differences between male pediatricians and male surgeons across all four outcomes. These findings were supported by all participants, across all levels of SDO. In other words, male physicians were seen as equally promotable, well performing, hard-working, and brilliant despite being in contexts more or less masculine and despite being evaluated by participants who were more or less egalitarian.

The greater influence of contextual level and trait level moderators for female physicians provide evidence that decision makers rely on stereotypes in the evaluation of women, even when the evaluator is egalitarian. The similar nature of the of pattern of outcomes, especially for performance evaluations and characterizations of hard work provide evidence of the theoretical link between conceptions of hard work and evaluations of performance. The next section examines this link further, discussing the results and interpretation of ability and effort characterizations first as predictors of promotion and performance, then as mediators of the relationship between target gender, specialty, and participant SDO on the allocation of promotion recommendations and performance evaluation.

The Role of Characterizations of Ability and Effort

Ability and Effort Characterizations as Predictors. Characterizations of brilliance (ability) and hard work (effort) were examined as predictors of promotion recommendations and evaluations of performance through hierarchical multiple regression analyses. Although results support that both ability and effort characterizations were predictive of both outcomes, characterizations of effort was the stronger predictor on both promotion and performance outcomes. These results are consistent with hypothesis 5a predicting a relationship between effort characterizations and evaluations of performance, but not 5b which predicted a stronger relationship between ability characterizations and promotion recommendations.

Consistent with the current findings, Heilman and colleagues (1978) found that the less favorable outcome (i.e., higher evaluations of performance) was awarded to those whose success was due to effort. However, contrary to the current findings, Heilman and colleagues (1978) found that the more favorable outcome (i.e., higher promotion ratings) was reserved only for those whose success was due to ability. Obtaining a promotion is regarded as a more favorable outcome than receiving good evaluations of performance because promotions provide upward mobility, prestige, and more access to power and influence in organizations; while good performance evaluations are not guaranteed to be tied to rewards (Joshi et al., 2015). Using attribution theory to describe their findings, the scholars explain that the more desirable reward (promotion) was awarded more often when success was due to stable (i.e., ability) than to unstable (i.e., effort) personal characteristics (Deaux et al., 1974; Heilman et al., 1978). In this study, while higher evaluations of performance and higher rates of promotion recommendations were awarded to both those perceived as hard workers (effort) and to those perceived as highly capable workers (ability), it was descriptions of hard work that better predicted outcomes.

There are two main explanations for this outcome, both rooted in differences in study design. Heilman et al. (1978) manipulated cause of success, while the current study measured perceived characteristics of a successful target. Thus, in Heilman et al.'s (1978) design, there could only be one cause of success at a time; while in the current study, targets could be perceived simultaneously as both hard workers and highly capable. Still, effort characterizations was the stronger predictor of both outcomes, lending support to the second explanation, which is the nature of the jobs under investigation. Although the job was male-typed in nature (a management position), Heilman et al. (1978) did not specify the actual job. The current study focused clearly on two very specific jobs: professor and director of the residency program either in the pediatrics or surgery department. Medical literature supports that especially in the U.S., patients and people in general, perceive doctors as experts and expect high technical competence from them (Emanuel & Emanuel, 1992; Lings, Evans, Seamark, Seamark, Sweeney, Dixon, & Gray, 2003). Further, the occupation of physician is seen as high status, high value, and high prestige due in part the extremely high level of skills, knowledge, and abilities associated with success on the job (Jozefowicz, Barber, Eccles, 1993). Thus, one interpretation of these findings is that the participants perceived successful targets as deserving of promotions and high evaluations of performance because to be a successful physician leader, one has to be incredibly capable because the nature of the job is inherently technical and difficult, but especially very hard working because on top of the job itself, the target has a leadership role and is in charge of an entire residency program.

Lastly, in light of results from the prior section, these findings provide further evidence of the linkage between characterizations hard work and evaluations of performance. Which, together strongly suggest meditation, especially for effort characterizations.

Ability and Effort Characterizations as Mediators. Ability and effort characterizations were tested as mediators by using Baron and Kenny's (1986) regression based procedure for assessing mediated effects. This study found evidence of a full mediation for female surgeons, such that the characterization of hard work fully mediated the relationship between *gender X specialty X and participant SDO* and the evaluations of performance she received. With regards to providing support that effort characterizations acted as a mediator, the current study's results are consistent with Heilman & Guzzo's (1978) finding that causal explanations typically attributed to women (i.e., effort) were strongly associated to the less good outcome (i.e., performance). Similar to what they found, evidence of mediation was supported, but only for females in the surgery condition, such that higher characterizations of effort were awarded with higher evaluations of performance. One interpretation could be a detrimental one for successful women; in their study, Heilman et al., (1978, p. 354) concluded that "women, who more often than men are assumed to be successful because of effort not ability, are apt to be subtly bypassed when the most meaningful organizational rewards are distributed." However, these prior findings support only a part of the overall story.

Another interpretation of the current results could be one more positive in nature, providing a better outlook for successful women. The current study supported full mediation, particularly within the female surgery condition, such that participants low in SDO (most egalitarian) attributed higher levels of effort to targets as compared to participants high in SDO (least egalitarian). In turn, these higher effort attributions had positive implications for female surgeons' performance ratings. The current mediation findings imply a more positive result for females in male-typed jobs than those found by Garcia-Retamero and Lopez-Zafra (2006). Consistent with lack of fit and role congruity predictions, participants in their study assigned

lower performance scores to female leaders in an industry incongruent with her gender role (auto-manufacturing) as compared to female leaders in an industry more congruent to her gender role (clothing manufacturing) (Garcia-Retamero et al., 2006).

However, the mediation claims of these comparable studies (Garcia-Retamero et al., 2006; Heilman et al., 1978) should be extrapolated to the current mediation results with caution, as these two studies failed to actually perform statistical analyses specifically testing for mediation effects. For example, Heilman et al. (1978) first tested the impact of gender on attributions, then separately tested attributions as predictors of workplace outcomes, missing a full test of mediation effects. There is a scarcity of scholarly work that examine causal attributions as a mediator, especially in gender and leadership literature (Martinko et al., 2007). One recent study conducted by Brescoll and Uhlmann (2008) found support that internal causal attributions for anger can harm women more than men in a leadership context. When expressing anger at work, females were more likely to be penalized (lower status, lower wages) than males because female targets' emotions were attributed to her "being out of control" while male targets' emotions are attributed to the "situation being frustrating". Their central contribution was finding evidence that it was the internal attribution that lead to penalties over and above gender itself. Although Brescoll and colleagues (2008) focused on attributions for emotion displays versus attributions for reasons for success, their study highlights the importance of examining the cognitive processes perceivers undergo to make sense of others' behaviors at work.

The joint interpretation of these mediation effects and with the findings supporting a three-way interaction of target gender, target specialty, and participant SDO on evaluations of performance, provide evidence of a mediated boost specifically for female surgeons.

Specifically, when evaluating a successful female physician leader in a highly incongruous job (i.e., surgery), evaluators first make sense of her success by characterizing her as a very hard worker, then in turn award her hard work with higher evaluations of her performance. This boost, however, is awarded only by the most egalitarian evaluators. Although the results of the mediation analysis on performance is significant, it is important to remember that the results of the 3-way interaction on performance was marginally significant ($p=.06$), thus any subsequent interpretation should be made with caution. Still, many researchers argue that failing to include marginally significant results risks a false acceptance of the null hypothesis, thus marginally significant results should be included and interpreted with some caution (Pritschet, Powell, & Horne, 2016).

Overall, it seems that perceivers undergo a longer, more complicated sensemaking process when faced with evaluating a highly incongruous target. Support for mediation of effort characterizations provide evidence of an extra cognitive step for decision makers, in addition to the typical gender role matching process, when evaluating female surgeons. Moreover, this complicated path leads to an outcome of less value (good performance evaluations), even when evaluated by the most egalitarian of participants.

Limitations and Future Directions

This study has several limitations to consider. First, the stimulus materials used in the study may have been too distracting or too complicated for participants who are not nurses, doctors, or professors themselves, because they likely do not have the technical knowledge and understanding of healthcare or academic medical settings. For example, when reading the summary performance evaluation with all the technical performance indicators, participants may have been overly impressed, thus skewing the results of study measures in an overly positive

way. Additionally, because this research was conducted with an MTurk sample including participants claiming current employment in a fulltime job and not with a sample of actual physicians in an academic medical institution, another major limitation is the generalizability of the results. Despite this, researchers have exhibited considerable external validity of experimental work in organizational behavior (Stone-Romero, 2002). Specifically, there exists “strong and convincing evidence of the similarity of laboratory and field research” in areas including decision making and employee evaluations and the “criticisms of laboratory research are often based upon stereotypes about such research, not on objective evidence of its supposed deficiencies” (Stone-Romero, 2002, p. 79). Still, future research should test the current predictions in the field. Further, recent studies have demonstrated that data obtained through online sources such as MTurk are valid and generalizable (e.g., Buhrmester, Kwang, & Gosling, 2011; Holden et al., 2013; Shapiro et al., 2013). Additionally, studies support that in many cases, online sampling methods and research performed online can yield either high quality data or data of similar quality to traditional study designs (Goodman, Cryder, & Cheema, 2012; Hauser & Schwarz, 2015). However, due to the novel nature of online recruitment methods, more research is needed to understand the differences in responses between MTurk workers and other sample populations.

The use of a cross-sectional design, whereby the sample population was exposed to all proposed measures at a single point in time, may have been a potential limitation. Thus, readers should interpret any causality found between the study constructs with some caution (Tippins & Sohi, 2003). Further, all data were obtained via an online questionnaire, exposing the study to common method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). These two reasons could have either inflated causal relationships that were discovered in this study or contributed to

a failure to find or link predicted causal relationships. Further, these causes could have contributed to the high correlations among the four outcome variables, in particular performance evaluations with promotion recommendations and characterizations of effort with characterizations of ability. While the study and measurement of stereotypes and attributions is difficult without questionnaires, future research should attempt other designs (e.g., a longitudinal study in which participants respond to measures at multiple points in time) or methods of data collection such that limitations such as these would not inflate or neglect predicted relationships between constructs.

Another potential set of limitations is the study's measurement of attributions and use of a research question (versus a hypothesis) to examine attributions as a mediator. With regards to the measurement, the decline of attribution research in the leadership domain in the 1980's contributed to the current need for more refined and robust measures of attribution (Martinko, 2007). This need was acknowledged by Swim and Sanna (2002) and Martinko, Douglas, and Harvey (2006, 2007), particularly in measuring the attribution of others' as opposed to one's own behavior. The current study attempted to address this need by creating a set of items mirroring "effort" and "ability" attributions. The resulting ability and effort scales used in this study incorporated the two most commonly mentioned traits describing men and women in a qualitative study that content analyzed over 300 letters of recommendation (Trix & Psenka, 2003). In these letters, men were consistently described using "standout" or brilliance adjectives, while women were consistently described using "grindstone" or adjectives indicative of working extremely hard. Similar to the organizational context and promotion scenario of the current study, Trix et al.'s (2003) study investigated medical faculty at a large academic medical institution who submitted their materials as part of their promotion application. Future studies of

causal attributions should systematically use a consistent measure of effort and ability, such as the newly formulated brilliance and grindstone scales, thereby contributing to its robustness and utility in survey methods.

Additionally, future studies of causal attributions should examine the mediational role of attributions with causal hypotheses rather than with a research question. Due to the gap in prior research testing attributions as mediators in a “lack of fit” context, the current research addressed this gap first by testing hypotheses examining attributions as a predictor and an outcome in the same study, then by exploring (through a research question) whether attributions could mediate the relationship between study predictors and outcomes. The current study provided the necessary empirical step that supported the mediational role of attributions. Future research can build on this research by constructing hypotheses that directly examine attributions as mediators.

The use of social dominance orientation may also be a limitation due to measurement and construct validity concerns. With regards to the measurement of social dominance orientation, the current study used the shortest form of the instrument, containing four items (Pratto et al., 2013). The use of fewer items may have decreased the reliability of the SDO scale. However, the internal consistency of the current study’s SDO scale was not low (Cronbach’s alpha = .88). Also, the study intentionally included the measure of SDO with fewer items to ensure the greatest number of respondents. Further, as Pratto and colleagues (2013) pointed out in their robust validation study of this four item SDO measure, “there may be a trade-off between the number of items and the number of participants in producing reliable results. For studies with fewer available participants, researchers may opt to use longer measures of SDO, and/or longer measures of criterion variables” (p. 592). However, the researchers state that construct validity results reach a satisfactory level of robustness with samples of 100 or more. As such, the current

study met that criteria with a final sample of N=209. Still, future studies should include more participants or longer measures of SDO.

Further, the overall construct of SDO has been criticized by some researchers. They argue that because the meaning and role of SDO has vastly shifted over the years, SDO as a distinct construct is either weak, explained by more robust constructs such as social identity theory, or even false (Hewstone & Miles; 2004; Turner & Reynolds, 2003). According to these scholars, one of the most problematic issues is the claim that SDO as a trait, therefore dispositional and unchanging. In their (2003) paper, Turner and Reynolds reference research findings that indicate a person's level of SDO is context specific because, "SDO is a measure of specific forms of inequality that are relevant to identity concerns and the social context which SDO is being assessed" (p. 202). In other words, because a person's primary social identity is often made salient by the surrounding context, identity concerns will shift depending on the environmental cues (Tajfel & Turner, 1986; Turner & Reynolds, 2010). Because the present research explicitly considers study predictions within the specific context of a highly masculine organizational context, where women are vastly underrepresented especially in leadership roles, there is an assumption that SDO remains relatively stable within this contextual boundary. In support of this assumption, prior research supports that male-dominated workplaces with highly prestigious occupations such as physician, are much more likely to employ individuals who are high in social dominance orientation (Pratto, et al., 1997; Joshi et al., 2015). With regards to the present study, this critique could be applied as an external validity concern. Therefore, the results of the current study could be generalized to healthcare professionals, specifically for those within the occupation of physician. However, future research should test SDO in other highly masculine contexts to continue to evaluate the construct validity and external validity of SDO.

Moreover, Joshi et al., (2015) explain that, “Research based on social dominance theory suggest that formal and informal practices in these occupational settings constitute a hierarchy-enhancing context...” and “tend to support and facilitate the disproportionate allocation of things with positive social value (e.g., high income, prestigious jobs, good education) to dominant social groups” such as men (p. 1520). However, the underlying nature of these hierarchy enhancing processes associated with individuals who are high in SDO is largely unknown. Specifically, are those high in SDO motivated to maintain the dominance of men or do they tend to maintain social hierarchy because of their disposition? A group of scholars indirectly address this question by breaking down the dispositional nature of SDO and the motivational nature behind biased behavior (i.e., disproportionately allocating valued outcomes to men). These researchers who study both SDO and implicit bias explain that what is most damaging is not a person’s disposition, but rather the discriminatory behavior towards the non-dominant group member resulting from implicit bias. These scholars outline that those high in SDO can mitigate discriminatory behavior if the person (1) is aware of the bias, (2) is in control of the discriminatory behavior due to the bias, and (3) is motivated to correct the bias (Blair, 2003; Dasgupta, 2004; Towles-Schwen & Fazio, 2003). Further, the researchers outline this behavior “correction” will take less effort if practiced over time, even if a person’s SDO itself remains unchanged over time (Sindanius, et al., 2004). Future studies should directly address this dispositional or motivational question behind SDO.

Additionally, future research should extend the current study predictions to other dimensions of social identity other than gender and in other highly masculine contexts other than in an academic medical institution. Prior scholarship on social dominance orientation and workplace discrimination claim that bias based on gender can easily translate into and explain

biased based on other low-status groups such as racial and ethnic minorities, older workers, and LGBTQ individuals (Chin, 2010; Dasgupta, 2014; Eagly & Chin, 2010; Pratto et al., 1994, 2014). Further, because the current study supported a “pro-female” bias for those low in SDO, future research in other highly masculine contexts should examine if and why those low in SDO exhibit a bias for members belonging to the less dominant social identity group (i.e., women, racial and ethnic minorities, older workers, and LGBTQ identified people). Empirical and anecdotal evidence has supported that organizations, especially ones dominated by men, benefit from employing a more diverse workforce because diversity brings innovation, a competitive advantage with novel thinking, and positive impact on employee satisfaction as well as the bottom line (Herring, 2009; Joshi & Roh, 2009; Robinson & Dechant, 1997). Further, due to the lack of studies that examine contexts with varying degrees of high masculinity, the current study offers a valuable contribution. Still, the current findings highlight the need for these predictions to be investigated in other highly masculine contexts because these findings may help further the understanding and alleviation of bias notorious in other industries (e.g., science, technology, engineering, and math fields).

Lastly, the inclusion of only very highly successful targets may have contributed to study limitations. Although the study intentionally focused on perceptions of successful targets, perhaps the level of success was too high, preventing a larger range of perceptions, impressions, and attributions to occur. This could have restricted the way people responded because, for example, target gender was no longer salient or mattered much less, rendering the implicit process of gender-based stereotyping and attribution based sensemaking irrelevant. Prior research supports that if a target’s performance is clearly and incontrovertibly successful, people rely less on gender based stereotypes and more on objective information (Heilman et. al, 1978;

Heilman et al., 2008). Despite this potential limitation, the current study found that target gender did matter, especially when considering participant SDO. However, in order to address this issue of the target being too successful and to allow stereotypic processes to occur, future research could include successful targets who are not highly successful in every dimension presented. Related to this issue, future research could consider highly successful targets who have also failed. Perhaps by incorporating instances where the target makes a mistake, target gender could become more salient to the perceiver. Extant research supports that even when successful, men and women are evaluated at different standards (Biernat & Kobrynowicz, 1997; Biernat & Fuegen, 2001; Foschi, 1989), thus men are afforded more leniency to fail and make mistakes in leadership roles as compared to equivalent women. This results in organizational penalties for female leaders (Rudman, 1998; Rudman & Glick, 1999). Additionally, prior research supports that causes ascribed to failure (i.e., internal or external attributions) can further influence the organizational outcomes (Deaux et al., 1974; Feldman-Summers et al., 1974; Weiner, 1985). Based on this, future research should examine if and why highly successful women suffer the same penalties as men when they fail.

Implications for Theory

This research highlights the complexity and subtlety of the effect of gender stereotypes in highly male-typed workplaces and has a number of implications for psychological theory. Based on lack of fit theory and role congruity theory (Eagly et al., 2002; Heilman, 1986), we tested if the degree of incongruence between the leadership role and the gender role could affect people's allocation decisions and characterizations about a successful physician leader. Contrary to these theories, our study instead provided support for a shifting standards model (Biernat et al., 1998), where the study's most incongruous target (female surgeon) consistently received more

favorable outcomes and characterizations. Perhaps due to the high degree of masculinity of the context, perceivers lacked an adequate cognitive frame or prototype in which to accurately evaluate the female surgeon, consequently defaulting to a comparison with stereotypes of “women in general”. There are remarkably few studies that test the lack of fit theory, role congruity theory, and attribution theory (Eagly et al., 2002; Heider, 1959; Heilman, 1986) in extremely high masculine contexts like a leadership role in a surgery department. Perhaps at a certain degree of incongruity, perceivers switch from a heuristically based gender/role matching process to a shifting standards comparison process to make sense of a highly incongruous target. This conjecture rather seamlessly complements, extends, and hopefully helps revive the study of attribution theory in leadership contexts.

The decline in attribution research in the leadership domain in the 1980s was largely due to two frequently cited criticisms: (1) that there are too many factors that can affect workplace behavior, such that causal attributions explain too small a portion of the variance in leader behavior (Mitchell, 1982); and (2) that the cognitive labor of the attributional process is too laborious, such that it is unlikely to occur day to day, thus reserved mainly for disappointing or surprising occurrences (Cronshaw & Lord, 1987; Lord & Maher, 1990). With regards to attributions accounting for unsubstantial proportions of variance, the criticizers themselves presented a solution outlined by Martinko and colleagues in 2007, “Green and Mitchell (1979) expressed doubt that attributions would account for major portions of the variance in leader behavior *unless* moderators were controlled or accounted for,” (p. 565). Accordingly, the current study addresses this criticism by identifying and accounting for multiple moderators, both contextual (specialty) and dispositional (participant SDO) - in addition to other more obvious variables such as target and participant gender.

Regarding the second criticism, outlining that attributional processes are most likely to occur when outcomes are disappointing or surprising (Martinko et al., 2007), the current study may fall well within these boundary conditions. Given that characterizations of effort fully mediated the relationship between study predictors and outcomes particularly for female physicians in the surgery condition, this finding could be evidence of a “surprising” enough event for participants, therefore triggering a more laborious cognitive sensemaking process. As previously outlined, based on degree of role congruity, female surgeons were the most incongruous target, thus female surgeons’ success in a highly masculine context very well could have surprised participants, inciting an attributional process in an attempt to help explain the rare occurrence. Criticizers of the attributional process who claim it too laborious suggest that “people are most likely to attend to and use the most salient and immediate cues (Taylor & Fiske, 1978) or use scripts and heuristics (e.g., Schank & Abelson, 1977) when forming behavioral responses to everyday events,” (Martinko et al., 2007, p. 565). This reasoning complements the previous conjecture that perhaps at a high degree of incongruity, people switch from the typical heuristically based gender/role matching process to a different sense making process. Based on the current findings, this study extends theory by proposing that if a target is incongruous or “surprising” enough, perceivers will turn to a more laborious process to help make sense of the dissonant target by using a shifting standards comparison (Biernat et al., 1998) and attributional process (Heider, 1959; Weiner, 1990). Because attribution theory has fallen out of favor and given the findings of the current study, there is a substantiation of an enormous opportunity to revive and standardize the study of the causal attributional process.

Implications for Practice

In addition to theoretical contributions, this research has several important practical implications. Although this study demonstrated that evaluations of female physicians were impacted by the degree of masculinity of the context (specialty) and the evaluator's level of egalitarianism (SDO), evaluations of male physicians largely were not. These results provide evidence of prejudice grounded in gender stereotypes, otherwise widely known in organizations as *unconscious* or *implicit* gender bias (Dasgupta, 2004). Following the lead of the tech giant Google, practically all Fortune 500 companies and new and existing technology companies have integrated diversity training, rebranded as unconscious bias training, to address the lack of women in their male-dominated industries, especially in leadership roles. According to a recent article in *Forbes*, the diversity training industry has quickly burgeoned to a \$8 billion-a-year industry, owing its newly found popularity to the blameless nature of unconscious bias, "everyone is told they have it and can't avoid it, so no one is singled out or gets defensive" (Huet, 2015, p.12).

Despite the tremendous expansion of diversity training and other related programming within male-dominated businesses and organizations, the actual numbers still reflect a preference for white men (Ely et al., 2011). Even with the shift away from a "blame-the-woman" approach, there is considerable proof that unconscious bias still exists - systematically and inconspicuously showing up in performance evaluation and promotion processes within organizations. In a recent article in *Harvard Business Review*, Dobbin and Kalev (2016) affirmed that, "more than 90% of midsize and large companies use annual performance ratings to ensure that managers make fair pay and promotion decisions," however, "When companies introduce them, there's no effect on minority managers over the next five years, and the share of white women in management *drops* by 4%, on average," (p. 5).

The current study's findings could be applied here, offering a more nuanced approach to unconscious gender bias trainings, which rely mainly on lack of fit and role congruity theories to explain the scarcity of women and minorities in male-dominated fields (Hoyt, 2012). One application follows the results of the current research underscoring the crucial role social dominance orientation had on influencing the gender/role congruity process. Similar to previous findings (e.g., Biernat & Malin, 2008; Hoyt, 2012), the current findings supported that people high in SDO tended to demonstrate a bias for high status groups (males), while people low in SDO tended to demonstrate a bias in favor on low status groups (females). However, bias is not inevitable. Dasgupta (2004) carefully reviews scholarship outlining that when people become aware of their dispositional bias (SDO), the consciousness provides an opportunity to control it, sometimes preventing their bias from manifesting behaviorally. Perhaps by integrating a measure of SDO in diversity training programs, people can understand how personality traits such as SDO can influence gender bias over and above the gender/role matching process. In fact, one main critique of unconscious gender bias trainings is that it lowers people's accountability because of its blameless nature (Huet, 2015). By becoming more aware of one's own values and dispositional traits through a SDO measure, people may be able to better understand how they see the world. This could provide people with more of an opportunity to clearly understand, then control their personal biases, thus hopefully preventing bias from manifesting in performance evaluation and promotion processes.

Another application follows the overall finding that participants used stereotypes to inform their promotion and performance evaluation decisions, despite the identical nature of the profiles. This finding is consistent with previous studies, for example, Moss-Racusin, Dovidio, Brescoll, Graham, and Hadelsman (2012) found that male *and* female senior scientists were

more inclined to hire, mentor, and propose higher pay, when the same candidate for a vacancy was identified as John rather than Jennifer. Further, the current study provides evidence that people will create stories or make causal attributions to make sense of surprising behavior. These findings point to the importance of developing better evaluation criteria such that they are more objective, otherwise the subjective nature of evaluation processes helps perpetuate a bias for male physicians. Supporting this recommendation, Ellemers (2014) stated that “Organizations can counter the impact of implicit bias by developing clear criteria for employment, wages, and promotion, instead of relying on subjective impressions, ambiguous prototypes, or existing selection practices” (p. 52).

Perhaps if hospitals invested in technological systems that track objective criteria, such as clinical outcomes, promotions can be awarded in a data-based and meritocratic way. For example, one recent study provided clear and objective evidence that female physicians were better performers than male physicians. Tsugawa, Jena, Figueroa, Orav, Blumenthal and Jha (2016) found that patients were significantly better off when treated by female versus male internists. Specifically, the patients of female internists had “lower 30- day readmissions (15.02% vs 15.57%) and lower 30-day mortality (11.07% vs 11.49%)” as compared to patients of male internists (Parks & Redberg, 2016, p. 1). The authors attributed the success of female internists to their more attentive bedside manner as well as their stricter adherence to clinical protocol as compared to male internists (Tsugawa et al., 2016). News outlets quickly spread these findings, declaring the superiority of female doctors. For example, *The Atlantic* quickly released an article announcing that, “if all physicians were female, 32,000 fewer Americans would die every year” (Hamlin, 2016). If female internists objectively provide a higher quality of care, why then are these physicians supported, promoted, and paid less than their male peers?

It is difficult to make the case in light of Tsugawa et al.'s (2016) findings. In direct response to Tsugawa et al.'s (2016) findings, the academic medical community is similarly calling for a push in creating objective systems that promote equity for all physicians; further highlighting that, "such equity promises to result in better professional fulfillment for all physicians as well as improved patient satisfaction and outcomes," (Parks et al., 2016).

These practical applications provide some evidence that the path from unconscious bias to discriminatory action is not always inevitable. These practical implications are important for men and women to consider as research supports that behaviors driven by implicit gender stereotyping are similarly carried out by both genders. This is critical in the context of medicine, especially because much of the medical literature in its current state use a "blame-the-woman" approach to explain the gender gap and glass ceiling. A growing contingent of scholars and practitioners within medicine recognize this opportunity. For example, in a recent article in *Medical Education*, Bleakley (2013) argued that, "...gender issues tend to be restricted to discussions of demographic changes and structural inequalities based on a biological reading of gender" (p. 59). These medical scholars cite a need for analysis outside of demography and biology that incorporates knowledge from social science to help understand why and how gender bias occurs in medicine. Social psychologists echo this need in more practical terms, stating that "Interventions to increase women's representation as leaders are more likely to be effective if they are guided by sound social science," (Eagly & Heilman, 2016). The current study hopes to begin to address these calls.

Conclusion

This study considered contextual factors, individual factors, and sensemaking processes that may perpetuate and explain a strong preference for men in physician leadership roles.

Counter to explanations that blame women, results of this study provided evidence of stereotype based gender bias. Even when provided with identically successful profiles, evaluators used stereotypes associated with the target's gender and job context to inform decisions. In addition, characterizations of hard work were highly important in evaluations of performance, specifically for female surgeons. The path to promotion, the more valued outcome, was less complicated for male physicians. Lastly, evaluator's level of egalitarianism played a vital role; with those high in SDO exhibiting a bias for men, and those low in SDO exhibiting a bias for women, perhaps overcompensating because of a heightened awareness of the increased barriers for women in surgical specialties. Hopefully this initial investigation, by taking a combined contextual and dispositional approach and merging gender role congruity and attributional perspectives within an employment context will foster future research to facilitate much needed practical applications.

REFERENCES

- Alers, M., van Leerdam, L., Dielissen, P., & Lagro-Janssen, A. (2014). Gendered specialities during medical education: a literature review. *Perspectives on medical education, 3*(3), 163-178.
- American College of Physician Executives (2009). Physician executive compensation survey. *American College of Physician Executives (ACPE) and Cejka Executive Search*.
- Ash, A.S., Carr, P.L., Goldstein, R. (2004). Compensation and advancement of women in academic medicine: is there equity? *Annals of Intern Medicine, 141*(3):205–212.
- Association of American Medical Colleges. (2006). Women in U.S. Academic Medicine: Statistics and Medical School Benchmarking 2005–2006. Washington, DC: Association of American Medical Colleges, 2006; Table 2.
- Association of American Medical Colleges. (2014). Physician Specialty Data Book. Center for Workforce Studies, 2013-2014. Washington, DC: Association of American Medical Colleges, November 2014; Table 1.3.
- Bartol, K. M. (1999). Gender influences on performance evaluations. In G. Powell (Ed.), *Handbook of gender and work* (pp. 165-178). Thousand Oaks, CA: Sage.
- Bickel, J. (2001). Gender equity in undergraduate medical education: a status report. *Journal of women's health & gender-based medicine, 10*(3), 261-270.
- Bickel, J., Wara, D., Atkinson, B.F. (2002). Increasing women's leadership in academic medicine: Report of the AAMC Project Implementation Committee. *Academic Medicine, 77*(10):1043–1061.
- Biernat, M., Crandall, C. S., Young, L. V., Kobrynowicz, D., & Halpin, S. M. (1998). All that you can be: Stereotyping of self and others in a military context. *Journal of Personality and Social Psychology, 75*(2), 301-317
- Biernat, M., & Fuegen, K. (2001). Shifting Standards and the Evaluation of Competence: Complexity in Gender-Based Judgment and Decision Making. *Journal of Social Issues, 57*(4), 707-724.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of personality and social psychology, 66*(1), 5. Biernat, M., & Malin, M. (2008). Political ideology and labor arbitrators' decision making in work–family conflict cases. *Personality and Social Psychology Bulletin, 34*(7), 888–899.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242-261.

- Bleakley, A. (2013). Gender matters in medical education. *Medical education*, 47(1), 59-70.
- Boulis, A. K., & Jacobs, J. A. (2010). *The changing face of medicine: women doctors and the evolution of health care in America*. Cornell University Press.
- Bollen, K. A., & Stine, R., (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115-40.
- Booth, A. L., Francesconi, M., & Frank, J. (2003). A sticky floors model of promotion, pay, and gender. *European Economic Review*, 47, 295–322.
- Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1), 84-103.
- Bowman, G. W., Worthy, N. B., & Greyser, S. A. (1965). Are women executives people? *Harvard Business Review*, 43(4), 14–28, 164–178.
- Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological science*, 19(3), 268-275.
- Buddeberg-Fischer, B., Stamm, M., Buddeberg, C., Bauer, G., Hämmig, O., Knecht, M., & Klaghofer, R. (2010). The impact of gender and parenthood on physicians' careers- professional and personal situation seven years after graduation. *BMC Health Services Research*, 10(1), 40.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, 6(1), 3-5.
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692.
- Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and social Psychology*, 5(1), 82.
- Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, 57(4), 725-741.
- Carli, L. L. (2006, July). Gender and social influence: Women confront the double bind. Paper presented at the 26th International Conference of Applied Psychology, Athens, Greece.
- Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology*, 68(6), 1030.

- Cancian, F. M., & Oliner, S. J. (2000). *Caring and gender*. Rowman & Littlefield.
- Caricati, L., Mancini, T., & Marletta, G. (2016). The role of ingroup threat and conservative ideologies on prejudice against immigrants in two samples of Italian adults. *The Journal of social psychology*, 1-12.
- Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. *Financial review*, 38(1), 33-53.
- Casciaro, T., & Lobo, M. S. (2005). Competent jerks, lovable fools, and the formation of social networks. *Harvard Business Review*, 83(6), 92-99.
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and social psychology bulletin*, 25(4), 413-423.
- Chin, J. L. (2010). Introduction to the special issue on diversity and leadership. *American Psychologist*, 65(3), 150.
- Cleveland, J. N., & Hollmann, G. (1990). The effects of the age-type of tasks and incumbent age composition on job perceptions. *Journal of Vocational Behavior*, 36(2), 181-194.
- Cokley, K. O., Tran, K., Hall-Clark, B., Chapman, C., Bessa, L., Finley, A., & Martinez, M. (2010). Predicting student attitudes about racial diversity and gender equity. *Journal of Diversity in Higher Education*, 3(3), 187-199.
- Committee on Maximizing the Potential of Women in Academic Science, Engineering (US), Committee on Science, & Public Policy (US). (2007). *Beyond bias and barriers: fulfilling the potential of women in academic science and engineering*. National Academy Press.
- Cook, C., Heath, F., & Thompson, R. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and Psychological Measurement*, 60 (6), 812-836.
- Crandall, C. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, 66, 882-894.
- Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology*, 72, 97-106.
- Darves, B. (2012). Women physicians in the specialties: Making gains. *The New England Journal of Medicine*, Career Center. Retrieved from: <http://www.nejmcareercenter.org/article/women-physicians-in-the-specialties-making-gains/> on September 8, 2015.

- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research, 17*(2), 143-169.
- Deaux, K. (1976). Sex: A perspective on the attribution process. *New directions in attribution research, 1*, 335-352.
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology, 29*, 80-85.
- DeNavas-Walt, C., Proctor, B. D., & Smith, J. C. (2010). US Census Bureau, current population reports, P60-238. Income, poverty, and health insurance coverage in the United States: 2009.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Dillman, D., Smyth, J., & Christian, L. (2007). *Mail and internet surveys: The tailored design method*. Hoboken: John Wiley and Sons.
- Dister, L. (2009). Outlook is bright for physician executive leadership compensation. *Phys Executive J Med Manage, 35*(6), 60-66.
- Dobbin, F., & Kalev, A. (2016). DIVERSITY Why Diversity Programs Fail And what works better. *Harvard Business Review, 94*(7-8), 52-60.
- Drinkwater, J., Tully, M. P., & Dornan, T. (2008). The effect of gender on medical students' aspirations: a qualitative study. *Medical education, 42*(4), 420-426.
- Dubno, P., Costas, J., Cannon, H., Wanker, C., & Emin, H. (1979). An empirically keyed scale for measuring managerial attitudes toward women executives. *Psychology of Women Quarterly, 3*, 357-364.
- Dunning, D., & Sherman, D. A. (1997). Stereotypes and tacit inference. *Journal of Personality and Social Psychology, 73*, 459-471.
- Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Harvard Business Press.
- Eagly, A. H., & Heilman, M. E. (2016). Gender and leadership: Introduction to the special issue. *The Leadership Quarterly, 3*(27), 349-353.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological bulletin, 108*(2), 233.

- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*, 2–33.
- Eisenman, R. (1991). Gender and racial prejudice of conservative college women. *Psychological Reports*, *68*, 450.
- Ellemers, N. (2014). Women at Work How Organizational Features Impact Career Development. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 46-54.
- Emanuel, E. J., & Emanuel, L. L. (1992). Four models of the physician-patient relationship. *Jama*, *267*(16), 2221-2226.
- Elsesser, K. M. (2015). Gender Bias Against Female Leaders: A Review. In *Handbook on Well-Being of Working Women* (pp. 161-173). Springer Netherlands.
- Ely, R. J. (1995). The power in demography: Women's social constructions of gender identity at work. *Academy of Management journal*, *38*(3), 589-634.
- Ely, R. J., Ibarra, H., & Kolb, D. M. (2011). Taking gender into account: Theory and design for women's leadership development programs. *Academy of Management Learning & Education*, *10*(3), 474-493.
- Ely, R. J., & Meyerson, D. E. 2000. Theories of gender: A new approach to organizational analysis and change. *Research in Organizational Behavior*, *22*: 103–153.
- Emanuel, E. J., & Emanuel, L. L. (1992). Four models of the physician-patient relationship. *Jama*, *267*(16), 2221-2226.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.
- Feldman-Summers, S., & Kiesler, S. B. (1974). Those who are number two try harder: The effect of sex on attributions of causality. *Journal of Personality and Social Psychology*, *30*(6), 846.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1991). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins.. *American Psychologist*, *46*(10), 1049.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, *82*(6), 878.

- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*, 2nd. NY: McGraw-Hill, 16-15.
- Federico, C., & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: The antecedents and consequences of “principled objections” to affirmative action. *Journal of Personality and Social Psychology*, 82(4), 488–502.
- Garcia-Retamero, R., & López-Zafra, E. (2006). Prejudice against women in male-congenial environments: Perceptions of gender role congruity in leadership. *Sex roles*, 55(1-2), 51-61.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26(3), 213-224.
- Gedeon, J. A., & Rubin, R. E. (1999). Attribution theory and academic library performance evaluation. *The Journal of academic librarianship*, 25(1), 18-25.
- Green, S. G., & Mitchell, T. R. (1979). Attributional processes of leaders in leader—member interactions. *Organizational Behavior and Human Performance*, 23(3), 429-458.
- Greenhaus, J. H., & Parasuraman, S. (1999). Research on work, family, and gender: Current status and future directions.
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451-470.
- Heider, F. (1959). On perception, event-structure and psychological environment. *Psychological Issues*, 1, 1-123.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In Staw, B., & Cummings, L. Eds. *Research in organizational behavior* (Vol. 5). Greenwich, CT: JAI.
- Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior & Personality*.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57(4), 657–674.
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, 32, 113-135.
- Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74, 935–942.

- Heilman, M. E., & Guzzo, R. A. (1978). The perceived cause of work success as a mediator of sex discrimination in organizations. *Organizational Behavior and Human Performance*, 21(3), 346-357.
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, 92, 81–92.
- Heilman, M. E., & Parks-Stamm, E. J. (2007). Gender stereotypes in the workplace: Obstacles to women's career progress. In Correll, S. J. (Ed.). *Social Psychology of Gender: Advances in Group Processes*. Vol. 24 (pp.47–77). Elsevier Ltd., JAI Press.
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89(3), 416.
- Henderson, A. C., Grappendorf, H., & Burton, L. (2011). Attributions for success and failure in athletic administration positions. *Journal of Issues in Intercollegiate Athletics*, 4, 257-270.
- Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2), 208-224.
- Hewstone, M. (1989). *Causal attribution: From cognitive processes to collective beliefs*. Basil Blackwell.
- Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2), 208-224.
- Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's Mechanical Turk. *Computers in Human Behavior*, 29(4), 1749-1754.
- Hoyt, C. L. (2012). Gender bias in employment contexts: A closer examination of the role incongruity principle. *Journal of Experimental Social Psychology*, 48(1), 86-96.
- Huet, E. (2015). Rise of the Bias Busters: How Unconscious Bias Became Silicon Valley's Newest Target. *Forbes/Tech*.
- Huffman, M. L. (2016). Does the Presence of Women in Management Impact Gender Inequality? In *Handbook on Well-Being of Working Women* (pp. 175-187). Springer Netherlands.
- Icahn School of Medicine at Mount Sinai. (2016). The Office of Academic Development and Enrichment Mentor Guide. [pdf file] Retrieved from <https://icahn.mssm.edu/files/ISMMS/Assets/About%20the%20School/For%20Faculty/OADE/mentor-guide.pdf>

- Index, C. P. (2011). US Bureau of labor statistics. Washington, DC.
- Jackson, L. A., Sullivan, L. A., & Hodge, C. N. (1993). Stereotype effects of attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology*, 65(1), 69.
- Jacobsen, J. P. (1998). *The economics of gender* (2nd ed.). Malden, MA: Blackwell.
- Jozefowicz, D. M., Barber, B. L., & Eccles, J. S. (1993). Adolescent work-related values and beliefs: Gender differences and relation to occupational aspirations. In *biennial meeting of the Society for Research on Child Development*.
- Jones, E.E., & Davis, K.E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219-266). New York: Academic.
- Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, 52(3), 599-627.
- Joshi, A., Son, J., & Roh, H. (2015). When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, 58(5), 1516-1545.
- Jozefowicz, D. M., Barber, B. L., & Eccles, J. S. (1993, March). Adolescent work-related values and beliefs: Gender differences and relation to occupational aspirations. In biennial meeting of the Society for Research on Child Development.
- Judge, T. A., Livingston, B. A., & Hurst, C. (2012). Do nice guys—and gals—really finish last? The joint effects of sex and agreeableness on income. *Journal of personality and social psychology*, 102(2), 390.
- Kanter, R. M. (1977a). *Men and Women of the Corporation* (Vol. 5049). Basic books.
- Kanter, R. M. (1977b). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American journal of Sociology*, 965-990.
- Kilminster, S., Downes, J., Gough, B., Murdoch-Eaton, D., & Roberts, T. (2007). Women in medicine— is there a problem? A literature review of the changing gender composition, structures and occupational cultures in medicine. *Medical education*, 41(1), 39-49.
- Krefting, L. A., Berger, P. K., & Wallace, M. J. (1978). The contribution of sex distribution, job content, and occupational classification to job sextyping: Two studies. *Journal of Vocational Behavior*, 13(2), 181-191.
- Kulik, C. T., & Bainbridge, H. T. (2006). Psychological Perspectives on Workplace Diversity. *Handbook of Workplace Diversity*, 25.

- Lautenberger, D. M., Dandar, V. M., & Raezer, C. L. (2014). *The State of Women in Academic Medicine: The Pipeline and Pathways to Leadership, 2013-2014*.
- Langston, E. 2008. Report 19 of the Board of Trustees: Gender Disparities in Physician Income and Advancement. American Medical Association.
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of applied psychology*, 86(1), 114.
- Lings, P., Evans, P., Seamark, D., Seamark, C., Sweeney, K., Dixon, M., & Gray, D. P. (2003). The doctor—patient relationship in US primary care. *Journal of the Royal Society of Medicine*, 96(4), 180-184.
- Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual review of Sociology*, 363-385.
- Lord, R. G., & Maher, K. J. (1990). Alternative information-processing models and their implications for theory, research, and practice. *Academy of Management Review*, 15, 9–28.
- Lyness, K. S. & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4), 777-785.
- Lyness, K. S., & Judiesch, M. K. (1999). Are women more likely to be hired or promoted into management positions? *Journal of Vocational Behavior*, 54(1), 158-173.
- Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66(1), 37.
- Martinko, M. J. (2004). Parting thoughts: Current issues and future directions. In M. J. Martinko (Ed.), *Attribution theory in the organizational sciences: theoretical and empirical contributions* (pp. 297–305). Greenwich, CT: Information Age Publishing.
- Martinko, M. J., Douglas, S. C., & Harvey, P. (2006). Attribution theory in industrial and organizational psychology: A review. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology*, Vol. 21 (pp. 127–187). Chichester, UK: Wiley.
- Martinko, M. J., Harvey, P., & Douglas, S. C. (2007). The role, function, and contribution of attribution theory to leadership: A review. *The Leadership Quarterly*, 18(6), 561-585.
- Maume, D. J., Jr. (1999). Glass ceilings and glass escalators: Occupational segregation and race and sex differences in managerial promotions. *Work and Occupations*, 1999, 483–509.

- Matschiner, M., & Murnen, S. K. (1999). Hyperfemininity and influence. *Psychology of Women Quarterly*, 23(3), 631-642.
- Meyers, D. G. (1999). *Social psychology*. Holland, MI: McGraw-Hill.
- Mitchell, T. (1982). Attributions and actions: A note of caution. *Journal of Management*, 8, 65-74.
- Mueller, D. C. (1997). First-mover advantages and path dependence. *International Journal of Industrial Organization*, 15(6), 827-850.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479.
- Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, 11(2), 140.
- Nadler, J. T., & Stockdale, M. S. (2012). Workplace gender bias: Not just between strangers. *North American Journal of Psychology*, 14(2), 213.
- Niemann, Y. F., & Dovidio, J. F. (1998). Relationship of solo status, academic rank, and perceived distinctiveness to job satisfaction of racial/ethnic minorities. *Journal of Applied Psychology*, 83(1), 55.
- Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 558-625.
- Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. *Academy of Management Review*, 5(2), 267-276.
- Nieva, V. F., & Gutek, B. A. (1981). *Women and work: A psychological perspective*. Praeger Publishers.
- Nonnemaker, L. (2000). Women physicians in academic medicine—new insights from cohort studies. *New England Journal of Medicine*, 342(6), 399-405.
- Novielli, K., Hojat, M., Park, P. K., Gonnella, J. S., & Veloski, J. J. (2001). Change of interest in surgery during medical school: a comparison of men and women. *Academic Medicine*, 76(10), S58-S61.
- Operario, D., & Fiske, S. T. (2001). Effects of trait dominance on powerholders' judgments of subordinates. *Social Cognition*, 19(2), 161-180.

- Paolillo, J. G. (1981, August). Role Profiles for Manager's at Different Hierarchical . In *Academy of Management Proceedings* (Vol. 1981, No. 1, pp. 91-94). Academy of Management.
- Park, J., Minor, S., Taylor, R. A., Vikis, E., & Poenaru, D. (2005). Why are women deterred from general surgery training?. *The American journal of surgery*, 190(1), 141-146.
- Parks, A. L., & Redberg, R. F. (2016). Women in Medicine and Patient Outcomes: Equal Rights for Better Work?. *JAMA Internal Medicine*.
- Pazy, A., & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior*, 22, 689–702.
- Perry, E. L., & Bourhis, A. C. (1998). A closer look at the role of applicant age in selection decisions. *Journal of Applied Social Psychology*, 28(18), 1670-1697.
- Peters, L. H., Terborg, J. R., & Taynor, J. (1974). Women as Managers Scale:(WAMS): a Measure of Attitudes Toward Women in Management Positions. *Journal Supplement Abstract Service of the American Psychological Association*.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879.
- Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: Have the times changed? *Journal of Management*, 28, 177–193.
- Pratto, F., Çidam, A., Stewart, A. L., Zeineddine, F. B., Aranda, M., Aiello, A., & Eicher, V. (2013). Social dominance in context and in individuals: Contextual moderation of robust effects of social dominance orientation in 15 languages and 20 countries. *Social Psychological and Personality Science*, 4(5), 587-599.
- Pratto, F., Saguy, T., Stewart, A. L., Morselli, D., Foels, R., Aiello, A., Aranda, M., Cidam, A., Chrysoschoou, X., Durrheim, K. & Eicher, V. (2014). Attitudes Toward Arab Ascendance Israeli and Global Perspectives. *Psychological science*, 25(3), 85-94.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, 67(4), 741.
- Pratto, F., Stallworth, L., & Sidanius, J. (1997). The gender gap: Differences in political attitudes and social dominance orientation. *British Journal of Social Psychology*, 36, 49–68.
- Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological methods*, 10(2), 178.

- Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological science*, 27(7), 1036-1042.
- Ragins, B. R. & Sundstrom, E. (1989). Gender and power in organizations: A longitudinal perspective. *Psychological Bulletin*, 105(1), 51-88.
- Ratcliff, J. J., Lassiter, G. D., Markman, K. D., & Snyder, C. J. (2006). Gender differences in attitudes toward gay men and lesbians: The role of motivation to respond without prejudice. *Personality and Social Psychology Bulletin*, 32(10), 1325-1338.
- Robinson, G., & Dechant, K. (1997). Building a business case for diversity. *The Academy of Management Executive*, 11(3), 21-31.
- Roman, R. J., Moskowitz, G. B., Stein, M. I., & Eisenberg, R. F. (1995). Individual differences in experiment participation: Structure, autonomy, and the time of the semester. *Journal of Personality*, 63(1), 113-138.
- Rubin, M., Hewstone, M. (2004). "Social Identity, System Justification, and Social Dominance: Commentary on Reicher, Jost et al., and Sidanius et al". *Political Psychology*. 25 (6): 823–44.
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, 74(3), 629.
- Sapiro, V. (1983). *The political integration of women: Roles, socialization, and politics*. University of Illinois Press.
- Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Earlbaum Assoc.
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, 57, 95–100.
- Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, 57, 675–688.
- Schmitt, M. T., Branscombe, N. R., & Kappen, D. M. (2003). Attitudes toward group-based inequality: Social dominance or social identity?. *British Journal of Social Psychology*, 42(2), 161-186.
- Shah, J., & Higgins, E. T. (1997). Expectancy \times value effects: Regulatory focus as determinant of magnitude and direction. *Journal of personality and social psychology*, 73(3), 447.

- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science*, 1(2), 213-220.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422-445.
- Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.
- Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology*, 70, 476-490.
- Simpson, D., Hafler, J., Brown, D., & Wilkerson, L. (2004). Documentation systems for educators seeking academic promotion in US medical schools. *Academic Medicine*, 79(8), 783-790.
- Simpson, R. (2000). Gender mix and organisational fit: How gender imbalance at different levels of the organisation impacts on women managers. *Women in Management Review*, 15(1), 5-18.
- Sinclair, S., Sidanius, J., & Levin, S. (1998). The Interface Between Ethnic and Social System Attachment: The Differential Effects of Hierarchy-Enhancing and Hierarchy-Attenuating Environments. *Journal of Social Issues*, 54(4), 741-757.
- Smith, B. (2007). *The psychology of sex and gender*. Needham Heights, MA: Allyn & Bacon.
- Snyder, T. D., & Dillow, S. A. (2012). *Digest of education statistics 2011*. National Center for Education Statistics.
- Stoller, J. K. (2008). Developing physician-leaders: key competencies and available programs. *Journal of Health Administration Education*, 25(4), 307-328.
- Stoller, J. K. (2009). Developing physician-leaders: a call to action. *Journal of general internal medicine*, 24(7), 876-878.
- Stone-Romero, E. F. (2002). The relative validity and usefulness of various empirical research designs. *Handbook of research methods in industrial and organizational psychology*, 77-98.
- Stuhlmacher, A., & Walters, A. (1999). Gender differences in negotiation outcome: A meta-analysis. *Personnel Psychology*, 52(3), 653-677.
- Sutton, C. D., & Moore, K. K. (1985). Executive women—20 years later. *Harvard Business Review*, 63(6), 42-66.

- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, *105*, 409–429.
- Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, *22*, 507–522.
- Tajfel, H., & Turner, J. C. (1986). "The social identity theory of intergroup behaviour". In S. Worchel & W. G. Austin. *Psychology of Intergroup Relations*. Chicago, IL: Nelson-Hall. pp. 7–24.
- Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in experimental social psychology*, *11*, 249-288.
- Thomson, M. F., & Martinko, M. J. (2004). Social Attributional Style. *Attribution Theory in the Organizational Sciences: Theoretical and Empirical Contributions*, 173.
- Tippins, M. J., & Sohi, R. S. (2003). IT competency and firm performance: is organizational learning a missing link?. *Strategic management journal*, *24*(8), 745-761.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, *14*(2), 191-220.
- Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin*, *29*(2), 170-182.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, *14*(2), 191-220.
- Tsugawa, Y., Jena, A. B., Figueroa, J. F., Orav, E. J., Blumenthal, D. M., & Jha, A. K. (2016). Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians. *JAMA Internal Medicine*.
- Tsui, A. S. (1982). A role set analysis of managerial reputation. In *Academy of Management Proceedings* (Vol. 1982, No. 1, pp. 265-269). Academy of Management.
- Tsui, A. S., & Gutek, B. A. (1984). A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. *Academy of Management Journal*, *27*(3), 619-635.
- Turner, J. C., & Reynolds, K. J. (2003). Why social dominance theory has been falsified. *British Journal of Social Psychology*, *42*(2), 199-206.
- Turner, J. C. & Reynolds, K. J. (2010). "The story of social identity". In T. Postmes & N. Branscombe. *Rediscovering Social Identity: Core Sources*. Psychology Press.

- Umphress, E. E., Smith-Crowe, K., Brief, A. P., Dietz, J., & Watkins, M. B. (2007). When birds of a feather flock together and when they do not: Status composition, social dominance orientation, and organizational attractiveness. *Journal of Applied Psychology, 92*, 396-409.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573.
- Weiner, B. (1990). Searching for the roots of applied attribution theory. *Attribution theory: Applications to achievement, mental health, and interpersonal conflict*, 1-13.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). Perceiving the causes of success and failure. Morristown, NJ: General Learning Press, 1971.
- Weiner, B., & Kukla, A. (1970). An attributional analysis of achievement motivation. *Journal of personality and Social Psychology, 15*(1), 1.
- West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of personality, 64*(1), 1-48.
- Wheeler, S. C., Petty, R. E., & Bizer, G. Y. (2005). Self-schema matching and attitude change: Situational and dispositional determinants of message elaboration. *Journal of Consumer Research, 31*(4), 787-797.
- Whitley, B. E. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology, 77*(1), 126-134.
- Williams, C. L. (1992). The glass escalator: Hidden advantages for men in the “female” professions. *Social problems, 39*(3), 253-267.
- Williams, C. L. (1995). Still a man's world: Men who do women's work (Vol. 1). Univ of California Press.
- Williams, J. C., & Dempsey, R. (2014). What Works for Women at Work.
- Williams, S. J., Pecenco, L., and Blair-Loy, M. (2013). Medical professions: The status of women and men. Center for Research on Gender in the Professions, UC San Diego.
- Wosinska, W., Dabul, A. J., Whetstone-Dion, R., & Cialdini, R. B. (1996). Self-presentational responses to success in the organization: The costs and benefits of modesty. *Basic and Applied Social Psychology, 18*, 229-242.
- Wright, A. L., Schwindt, L. A., Bassford, T. L., Reyna, V. F., Shisslak, C. M., Germain, P. A. S., & Reed, K. L. (2003). Gender differences in academic advancement: patterns, causes,

and potential solutions in one US College of Medicine. *Academic Medicine*, 78(5), 500-508.

Yedidia, M. J., & Bickel, J. (2001). Why aren't there more women leaders in academic medicine? The views of clinical department chairs. *Academic Medicine*, 76(5), 453-465.

Zhuge, Y., Kaufman, J., Simeone, D. M., Chen, H., & Velazquez, O. C. (2011). Is there still a glass ceiling for women in academic surgery? *Annals of surgery*, 253(4), 637-643.

Table 1

Number and Percentage of Active Physicians by Sex and Specialty: 2014

<i>Specialty</i>	<i>Total Active Physicians</i>	<i>Male</i>		<i>Female</i>	
		<i>Number</i>	<i>Percent</i>	<i>Number</i>	<i>Percent</i>
All Specialties	828,869	558,794	67.4	270,075	32.6
Allergy & Immunology	4,502	2,978	66.1	1,524	33.9
Anatomic/Clinical Pathology	13,701	8,825	64.4	4,876	35.6
Anesthesiology	40,714	30,814	75.7	9,900	24.3
Cardiovascular Disease	22,001	19,300	87.7	2,701	12.3
Child & Adolescent Psychiatry	8,232	4,085	49.6	4,147	50.4
Critical Care Medicine	8,816	6,676	75.7	2,140	24.3
Dermatology	11,353	6,277	55.3	5,076	44.7
Emergency Medicine	37,210	27,713	74.5	9,497	25.5
Endocrinology, Diabetes & Metabolism	6,509	3,658	56.2	2,851	43.8
Family Medicine/General Practice	108,751	68,647	63.1	40,104	36.9
Gastroenterology	13,607	11,562	85	2,045	15
General Surgery	25,164	20,727	82.4	4,437	17.6
Geriatric Medicine	4,827	2,445	50.7	2,382	49.3
Hematology & Oncology	13,755	9,600	69.8	4,155	30.2
Infectious Disease	7,938	4,887	61.6	3,051	38.4
Internal Medicine	110,823	71,096	64.2	39,727	35.8
Internal Medicine/Pediatrics	4,394	2,158	49.1	2,236	50.9
Interventional Cardiology	2,704	2,509	92.8	195	7.2
Neonatal-Perinatal Medicine	4,882	2,549	52.2	2,333	47.8
Nephrology	9,381	6,943	74	2,438	26
Neurological Surgery	5,168	4,790	92.7	378	7.3
Neurology	13,142	9,601	73.1	3,541	26.9
Neuroradiology	2,911	2,358	81	553	19
Obstetrics & Gynecology	40,790	19,669	48.2	21,121	51.8
Ophthalmology	18,308	14,211	77.6	4,097	22.4
Orthopedic Surgery	19,372	18,483	95.4	889	4.6
Otolaryngology	9,314	7,966	85.5	1,348	14.5
Pain Medicine	4,002	3,280	82	722	18
Pediatrics	56,210	22,266	39.6	33,944	60.4
Physical Medicine & Rehabilitation	8,879	5,796	65.3	3,083	34.7
Plastic Surgery	6,939	5,954	85.8	985	14.2
Preventive Medicine	6,602	4,536	68.7	2,066	31.3

Table 1 (Continued)

Psychiatry	37,277	23,535	63.1	13,742	36.9
Pulmonary Disease	5,724	5,100	89.1	624	10.9
Radiation Oncology	4,680	3,443	73.6	1,237	26.4
Radiology & Diagnostic Radiology	27,553	20,987	76.2	6566	23.8
Rheumatology	5,345	3,162	59.2	2,183	40.8
Thoracic Surgery	4,520	4,272	94.5	248	5.5
Urology	9,767	9,066	92.8	701	7.2
Vascular & Interventional Radiology	2,580	2,349	91	231	9
Vascular Surgery	3,128	2,827	90.4	301	9.6

Table 2

Means, Standard Deviations, and Intercorrelations for All Study Variables

	Mean	SD	1	2	3	4	5	6	7	8
<u>Controls</u>										
1. Participant Gender	1.51	.50	--							
<u>Predictors</u>										
2. Target Gender	1.48	.50	-.02	--						
3. Target Specialty	1.53	.50	-.02	-.03	--					
4. Participant SDO	1.96	.81	.14*	-.04	-.07	.88				
<u>Outcomes</u>										
5. Promotion	6.40	.70	-.21**	.03	.16*	-.14*	.83			
6. Performance	6.35	.58	-.13	-.04	.15*	-.14*	.77**	.90		
7. Ability Attribution	5.63	.75	-.10	.01	.10	-.18**	.49**	.57**	.87	
8. Effort Attribution	6.09	.68	-.27**	-.01	.23**	-.23**	.64**	.63**	.77**	.84

Note. $N = 207$. * $p < .05$ ** $p < .01$. Internal consistency reliabilities (Cronbach's alpha coefficient) for each variable are indicated in bold on the diagonal. Participant gender code dummies include 1=female 2=male. Target gender code dummies include 1=female 2=male. Target specialty code dummies include 1=pediatrics 2=surgery. Social dominance orientation (participant SDO), promotion, performance, ability attribution, and effort attribution were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table 3

Means and Standard Deviations for Outcome Measures by Condition

Outcome Measures	Female Pediatrician (n=49)		Female Surgeon (n=59)		Male Pediatrician (n=48)		Male Surgeon (n=51)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Promotion	6.19 _a	0.87	6.50 _b	0.63	6.32 _{a,b}	0.74	6.52 _b	0.53
Performance	6.23 _a	0.71	6.47 _b	0.50	6.25 _{a,b}	0.61	6.42 _b	0.49
Ability	5.51 _a	0.84	5.69 _a	0.58	5.57 _a	0.67	5.72 _a	0.58
Effort	5.88 _a	0.78	6.24 _b	0.81	5.93 _a	0.75	6.24 _b	0.67

Note. $N=207$. Condition means within rows not sharing a subscript differ significantly at the $p<.05$ level. Promotion, performance, ability, and effort were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table 4

Means and Standard Deviations for Outcome Measures by Condition and Participant SDO

Target Gender	Predictors		Promotion Recommendations		Performance Ratings		Ability Attributions		Effort Attributions	
	Target Specialty	Participant SDO	M	SD	M	SD	M	SD	M	SD
Female	Pediatrics	Low (n=14)	6.19	0.84	6.17	0.65	5.51	0.84	5.94	0.76
		Moderate (n=20)	6.50	0.51	6.53	0.56	5.81	0.51	6.24	0.46
	Surgery	High (n=15)	5.91	1.16	5.99	0.86	5.25	0.93	5.49	1.11
		Low (n=21)	6.73	0.45	6.71	0.36	6.01	0.86	6.50	0.50
		Moderate (n=19)	6.39	0.59	6.41	0.54	5.57	0.78	6.11	0.59
		High (n=19)	6.33	0.77	6.24	0.50	5.45	0.70	6.07	0.57
Male	Pediatrics	Low (n=15)	6.42	0.84	6.35	0.80	5.88	0.86	6.29	0.65
		Moderate (n=17)	6.25	0.72	6.20	0.52	5.45	0.69	5.79	0.68
	Surgery	High (n=16)	6.29	0.70	6.23	0.51	5.37	0.62	5.73	0.56
		Low (n=22)	6.50	0.56	6.35	0.56	5.70	0.70	6.22	0.58
		Moderate (n=15)	6.60	0.47	6.48	0.45	5.69	0.78	6.32	0.48
		High (n=14)	6.45	0.56	6.41	0.43	5.77	0.53	6.18	0.70

Note. N=207. Promotion, performance, ability, and effort were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table 5

Summary of Research Questions, Hypotheses, and Findings

<i>Hypotheses and Research Questions</i>		<i>Results</i>	
<i>Gender x Specialty</i>	<i>Participant SDO → Outcome Variables</i>	<i>Two-way Findings (gender X Specialty)</i>	<i>Three-way Findings (gender X Specialty X SDO)</i>
Research Question 1	Will participant SDO moderate the effects of the study predictors (gender X specialty) on the outcome variables?	--	Supported for performance, ability, and effort.
<i>Gender x Specialty → Promotion</i>			
Hypothesis 1a	Successful male physicians will be perceived to be more promotable than successful female physicians, no matter which specialty.	Not Supported: M=F	--
Hypothesis 1b	Successful female physicians in pediatrics will be perceived to be more promotable than successful female physicians in surgery.	Opposite Result: FP<FS**	--
Hypothesis 1c	Successful male physicians will be perceived to be equally promotable in pediatrics and surgery.	Supported: MP=MS	--
Hypothesis 1d	Differences in promotion between successful male and female physicians will be more amplified in surgery than in pediatrics.	Not supported: MP=FP, MS = FS	--
<i>Gender x Specialty → Performance</i>			
Hypothesis 2a	Successful female physicians will be perceived to be better performing than successful male physicians, no matter which specialty.	Not supported: M = F	Opposite Result: M > F High SDO†

Table 5 (Continued)

<i>Hypotheses and Research Questions</i>		<i>Results</i>	
Hypothesis 2b	Successful female physicians in surgery will be perceived to be better performing than successful female physicians in pediatrics.	Supported: FP < FS*	Supported: FP < FS Low SDO**
Hypothesis 2c	Successful male physicians will be perceived to be equally performing in pediatrics and surgery.	Supported: MP = MS	Supported: MP = MS All 3 levels
Hypothesis 2d	Differences in performance ratings between successful male and female physicians will be more amplified in surgery than in pediatrics.	Not supported: MP = FP, MS = FS	Supported: MP = FP, MS < FS Low SDO*
Gender x Specialty → Ability			
Hypothesis 3a	Participants will attribute ability more to successful male physicians than successful female physicians.	Not Supported: M = F	Not Supported: M = F All 3 Levels
Hypothesis 3b	Participants will attribute ability more to successful female pediatricians than successful female surgeons.	Opposite Result: FP < FS**	Opposite Result: FP < FS Low SDO*
Hypothesis 3c	Participants will attribute ability similarly to successful males in pediatrics and surgery.	Supported: MP = MS	Supported: MP = MS All 3 levels
Hypothesis 3d	Differences in ability attributions between successful male and female physicians will be more amplified in surgery than pediatrics.	Not Supported: MP = FP, MS = FS	Not Supported: MP = FP, MS = FS All 3 Levels

Table 5 (Continued)

<i>Hypotheses and Research Questions</i>		<i>Results</i>	
Gender x Specialty → Effort			
Hypothesis 4a	Participants will attribute effort more to successful female physicians than successful male physicians.	Not Supported: M = F	Not Supported: M = F All 3 Levels
Hypothesis 4b	Participants will attribute effort more to successful female surgeons than successful female pediatricians.	Supported: FP < FS**	Supported: FP < FS Low SDO**, High SDO**
Hypothesis 4c	Participants will attribute effort similarly to successful males in pediatrics and surgery.	Not Supported: MP < MS	Supported: MP = MS Low SDO
Hypothesis 4d	Differences in effort attributions between successful male and female physicians will be more amplified in surgery than pediatrics.	Not Supported: MP = FP, MS = FS	Supported: MP = FP, MS < FS Low SDO†
Ability & Effort → Promotion & Performance			
Hypothesis 5a	Effort attributions will be a stronger predictor of performance than ability attributions, and targets with higher effort attributions will be perceived as better performers than targets with lower effort attributions.	Supported: Effort attributions significantly predicted performance, more strongly than ability attributions.	--
Hypothesis 5b	Ability attributions will be a stronger predictor of promotion than effort attributions, and targets with higher ability attributions will be perceived as more promotable than targets with lower ability attributions.	Not Supported: Effort attributions significantly predicted promotion more strongly than ability attributions.	--

Table 5 (Continued)

<i>Hypotheses and Research Questions</i>		<i>Results</i>	<i>Hypotheses and Research Questions</i>
<i>Gender x Specialty → Ability & Effort → Promotion & Performance</i>			
Research Question 2	Will effort and ability attributions act as mediators to the relationship of gender X specialty X SDO on performance and promotion?	Supported: Effort attributions (but not ability) mediated the relationship between gender X specialty and performance (but not promotion).	--

Note. Bolded text in the "Results" columns indicate a significant finding consistent with the hypothesis.

SDO: Participant Social Dominance Orientation (three levels are low, moderate, and high)

FP: Female Pediatrician, FS: Female Surgeon, MP: Male Pediatrician, MS: Male Surgeon

<: Less than

>: Greater than

=: Equal to

† p< .10, *p<.05, **p<.01

Table 6

Univariate ANCOVA Results for Study Predictors on Promotion

	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Participant Gender	9.57	1	194	0.00
Target Gender	0.61	1	194	0.44
Target Specialty	6.27	1	194	0.01
Participant SDO	1.15	2	194	0.32
Target Gender x Target Specialty	0.17	1	194	0.68
Target Gender x Participant SDO	1.45	2	194	0.24
Target Specialty x Participant SDO	0.84	2	194	0.43
Target Gender x Target Specialty x Participant SDO	1.80	2	194	0.17

Note. $N = 207$. Participant gender was included as a control variable in the ANCOVA.

Table 7

Univariate ANCOVA Results for Study Predictors on Performance

	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Participant Gender	4.26	1	194	0.04
Target Gender	0.02	1	194	0.90
Target Specialty	5.79	1	194	0.02
Participant SDO	1.72	2	194	0.18
Target Gender x Target Specialty	0.16	1	194	0.69
Target Gender x Participant SDO	2.31	2	194	0.10
Target Specialty x Participant SDO	0.82	2	194	0.44
Target Gender x Target Specialty x Participant SDO	2.82	2	194	0.06

Note. $N = 207$. Participant gender was included as a control variable in the ANCOVA.

Table 8

Univariate ANOVA Results for Study Predictors on Ability Characterizations

	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Target Gender	0.19	1	195	0.66
Target Specialty	2.12	1	195	0.15
Participant SDO	3.00	2	195	0.05
Target Gender x Target Specialty	0.00	1	195	1.00
Target Gender x Participant SDO	0.90	2	195	0.41
Target Specialty x Participant SDO	0.66	2	195	0.52
Target Gender x Target Specialty x Participant SDO	2.87	2	195	0.06

Note. $N = 207$. Participant gender was not included as a control variable in the ANOVA.

Table 9

Univariate ANCOVA Results for Study Predictors on Effort Characterizations

	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Participant Gender	15.15	1	194	0.00
Target Gender	0.08	1	194	0.77
Target Specialty	13.48	1	194	0.00
Participant SDO	4.03	2	194	0.02
Target Gender x Target Specialty	0.02	1	194	0.89
Target Gender x Participant SDO	1.72	2	194	0.18
Target Specialty x Participant SDO	1.65	2	194	0.19
Target Gender x Target Specialty x Participant SDO	4.10	2	194	0.02

Note. $N = 207$. Participant gender was included as a control variable in the ANCOVA.

Table 10

Results of Hierarchical Regression of Effort and Ability Characterizations on Promotion

	β	R^2	R^2_{change}	df
<i>Step 1</i>		.04	--	205
Participant Gender	-.21**			
<i>Step 2</i>		.41	.37**	203
Participant Gender	-.04			
Ability	.01			
Effort	.63**			
Overall Model F	48.33**	--	--	203

Note. $N = 207$. † $p < .10$, * $p < .05$, ** $p < .01$ Standardized regression coefficients are reported. Participant gender code dummies include 1=female 2=male. Performance, ability, and effort were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table 11

Results of Hierarchical Regression of Effort and Ability Characterizations on Performance

	β	R^2	R^2_{change}	df
<i>Step 1</i>		.02	--	205
Participant Gender	-.13†			
<i>Step 2</i>		.41	.40**	203
Participant Gender	.02			
Ability	.21*			
Effort	.48**			
Overall Model F	47.78**	--	--	203

Note. $N = 207$. † $p < .10$, * $p < .05$, ** $p < .01$ Standardized regression coefficients are reported. Participant gender code dummies include 1=female 2=male. Performance, ability, and effort were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table 12

Mediational Analyses: Effort Attributions as a Mediator

	<i>Criterion 1</i> <i>Gender X Specialty</i> <i>X SDO (low vs. high) → Performance Ratings</i>	<i>Criterion 2</i> <i>Gender X Specialty</i> <i>X SDO (low vs. high) → Effort Attributions</i>	<i>Criterion 3</i> <i>Gender X Specialty X SDO (low vs. high) → Effort Attributions → Performance Ratings</i>
Gender X Specialty X SDO (low vs. high) (β)	.40*	.46**	.12 .62**
Overall Model <i>F</i>	2.21*	4.55**	10.60**
<i>R</i> -Square	.12	.22	.42

Note. Table entries are standardized β coefficients.

p*<.05, *p*<.01

Table 13

Mediational Analyses: Ability Attributions as a Mediator

	<i>Criterion 1</i> <i>Gender X Specialty</i> <i>X SDO (low vs. high) → Performance</i> <i>Ratings</i>	<i>Criterion 2</i> <i>Gender X Specialty</i> <i>X SDO (low vs. high)</i> <i>→ Ability</i> <i>Attributions</i>	<i>Criterion 3</i> <i>Gender X Specialty X</i> <i>SDO (low vs. high) →</i> <i>Ability Attributions →</i> <i>Performance Ratings</i>
Gender X Specialty X SDO (low vs. high) (β)	.40*	.39*	.20 .52**
Overall Model <i>F</i>	2.21*	1.64	8.60**
<i>R</i> -Square	.12	.09	.37

Note. Table entries are standardized β coefficients.

* $p < .05$, ** $p < .01$.

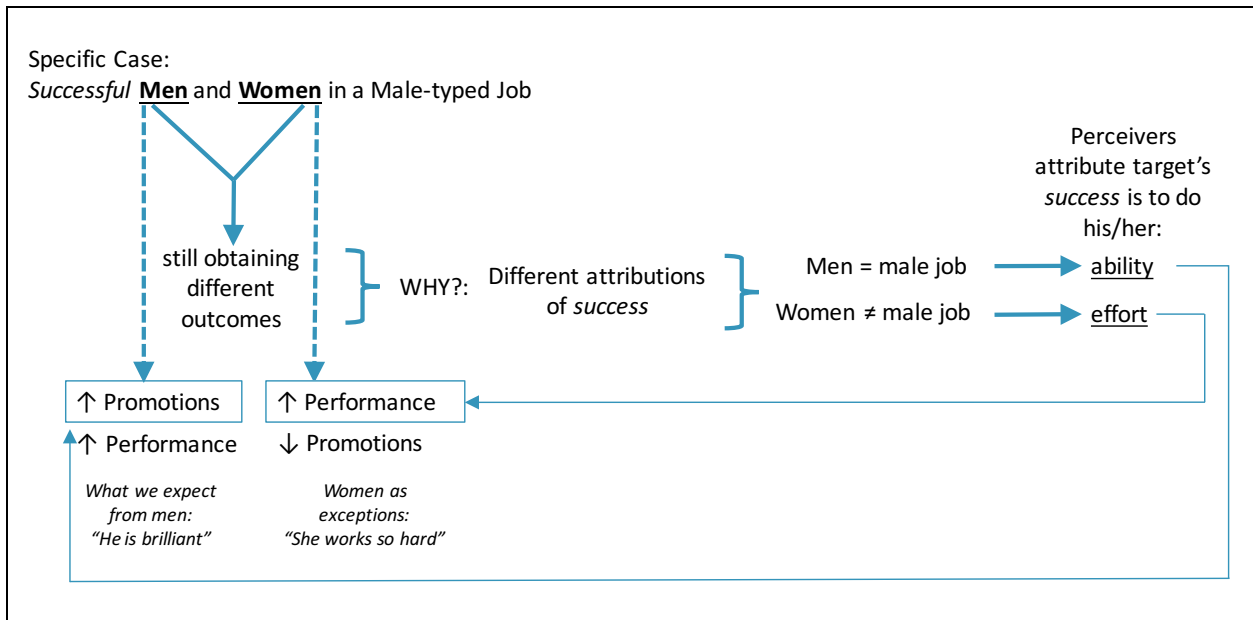


Figure 1. Theoretical process map of the central purpose of this study.

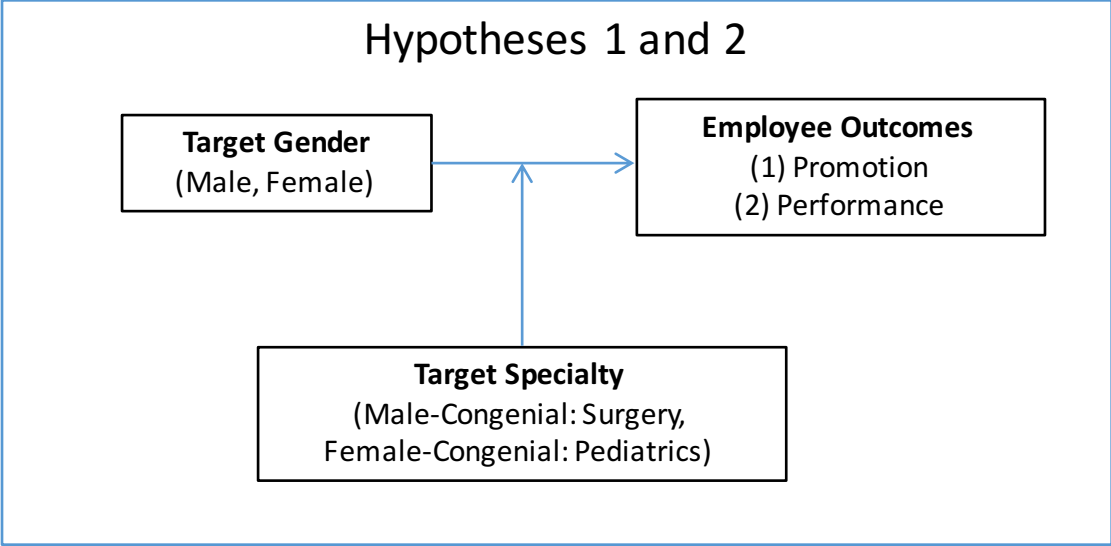


Figure 2. Model of moderation of specialty to gender on promotion and performance.

		Locus of Control	
		Internal	External
Stability	Stable	Ability (no control)	Task Difficulty (control)
	Unstable	Effort (control)	Luck (no control)

Controllability
(in parentheses)

Figure 3. Attributional explanations categorized by attribution dimension (Weiner, 1985)

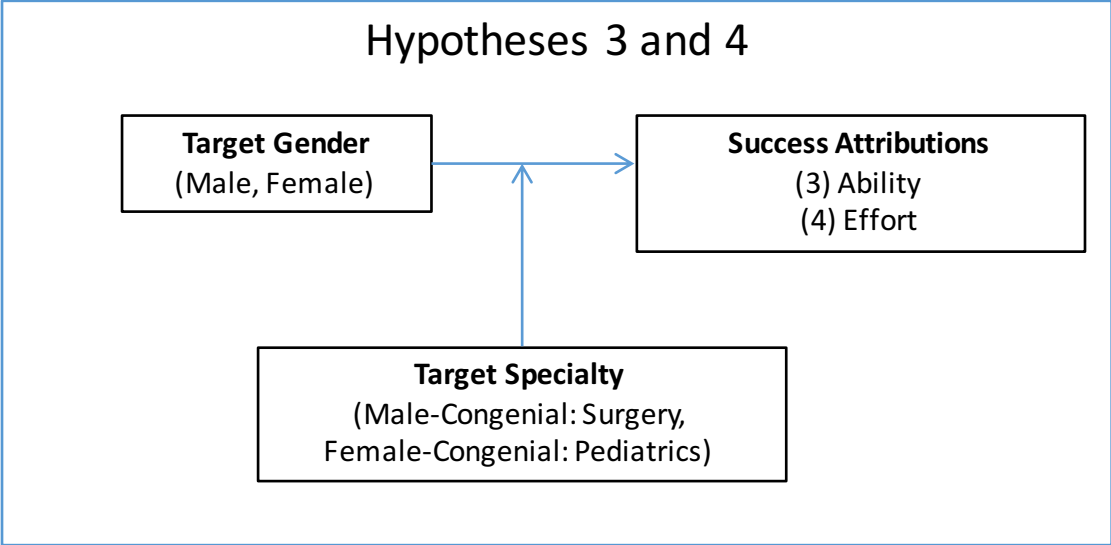


Figure 4. Moderation of specialty to gender on ability and effort.

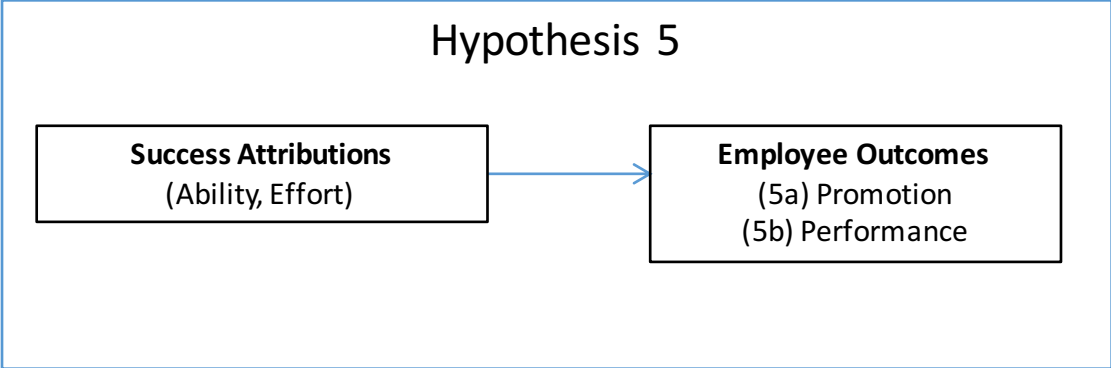


Figure 5. The impact of success attributions on employee outcomes.

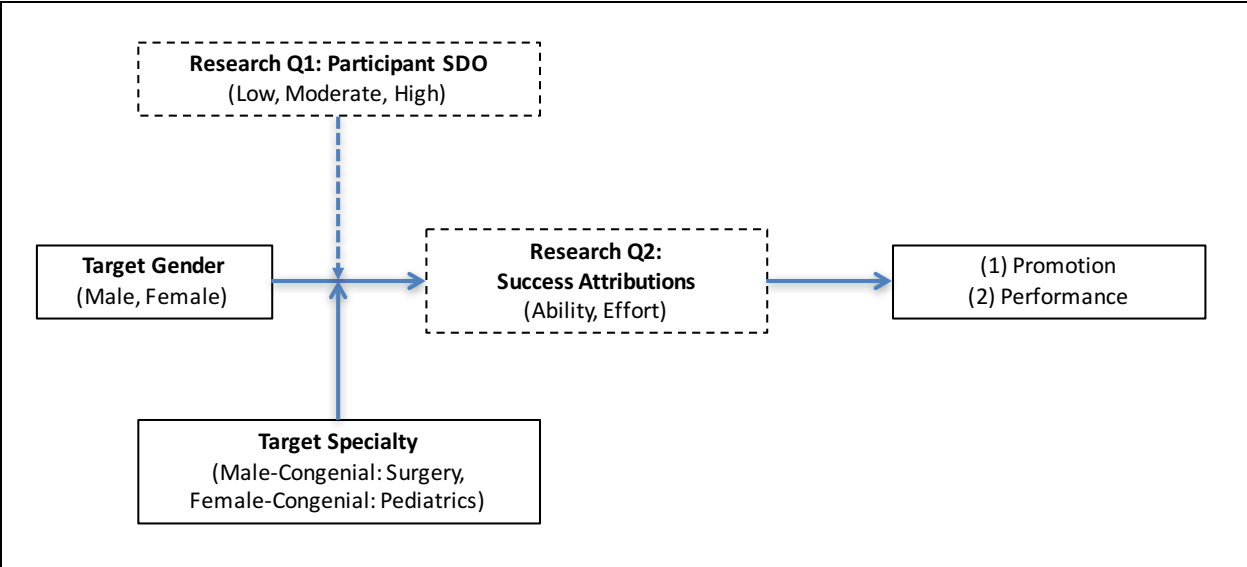


Figure 6. Full mediated moderation model of specialty congruency and success attribution on work outcomes.

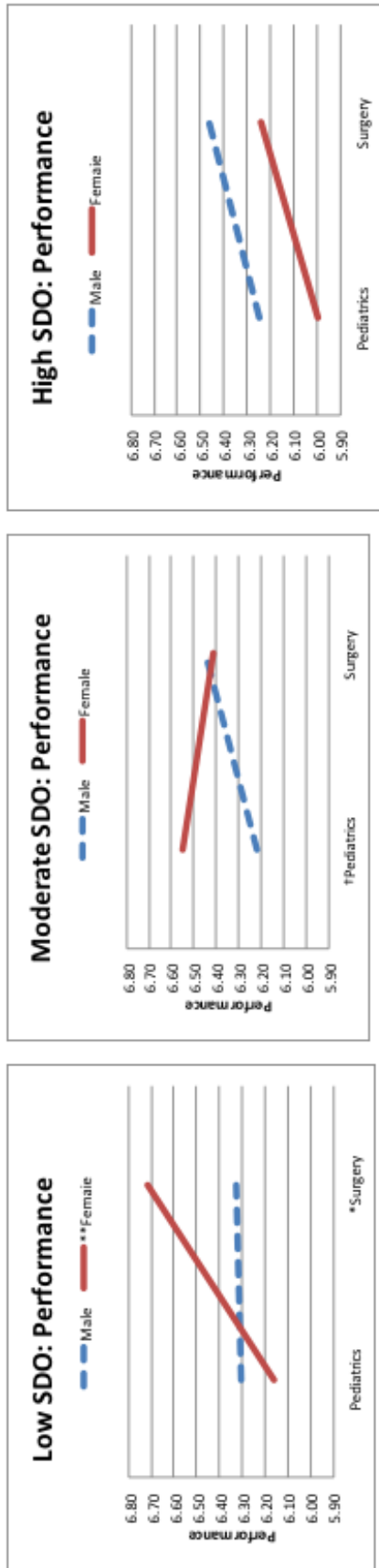


Figure 7. Graphs of performance means for the three-way interaction by participant SDO Level (low, moderate, and high).
 Note. † $p < .10$, * $p < .05$, ** $p < .01$. A significance symbol by a target gender indicates a significant within gender difference across specialty. A significance symbol by a target specialty indicates a significant across gender difference within the same specialty.

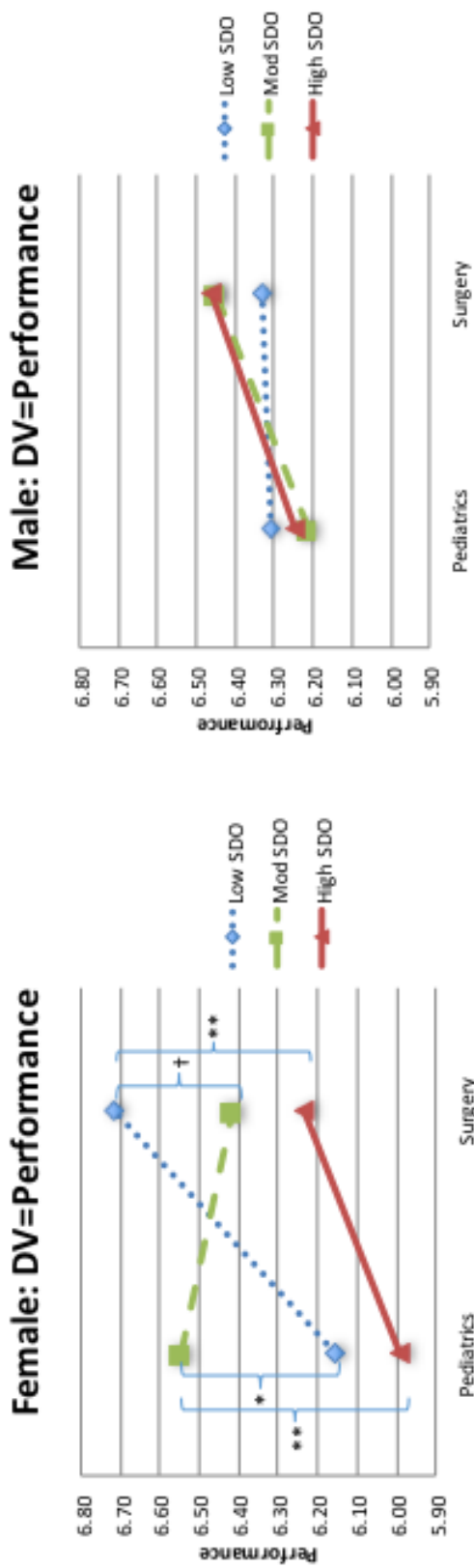


Figure 8. Graphs of performance means for the significant three-way interaction by target gender (male, female). Note. † $p < .10$, * $p < .05$, ** $p < .01$.

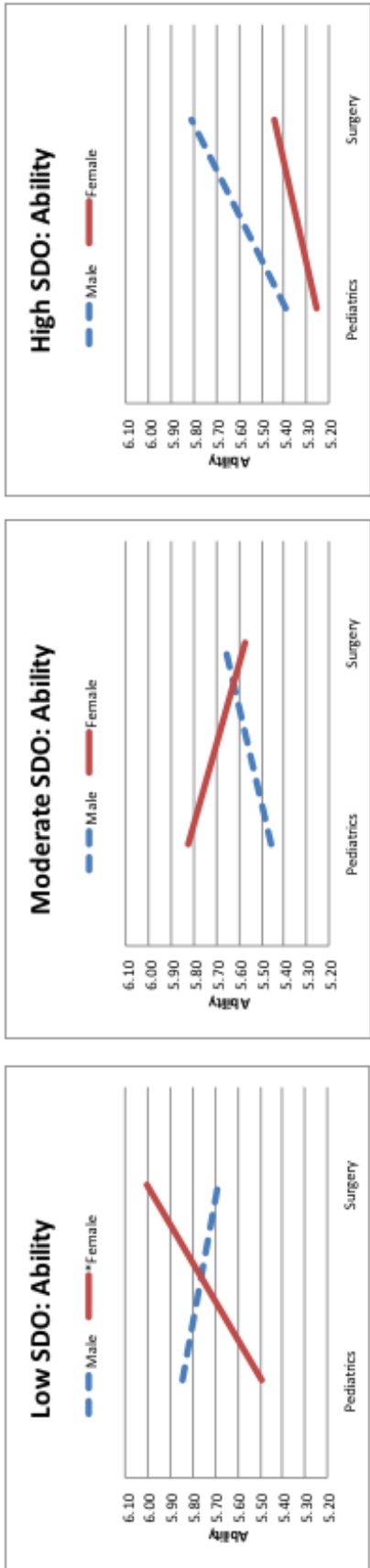


Figure 9. Graphs of ability characterization means for the three-way interaction by participant SDO Level (low, moderate, and high). Note. * $p < .05$. A significance symbol by a target gender indicates a significant within gender difference across specialty.

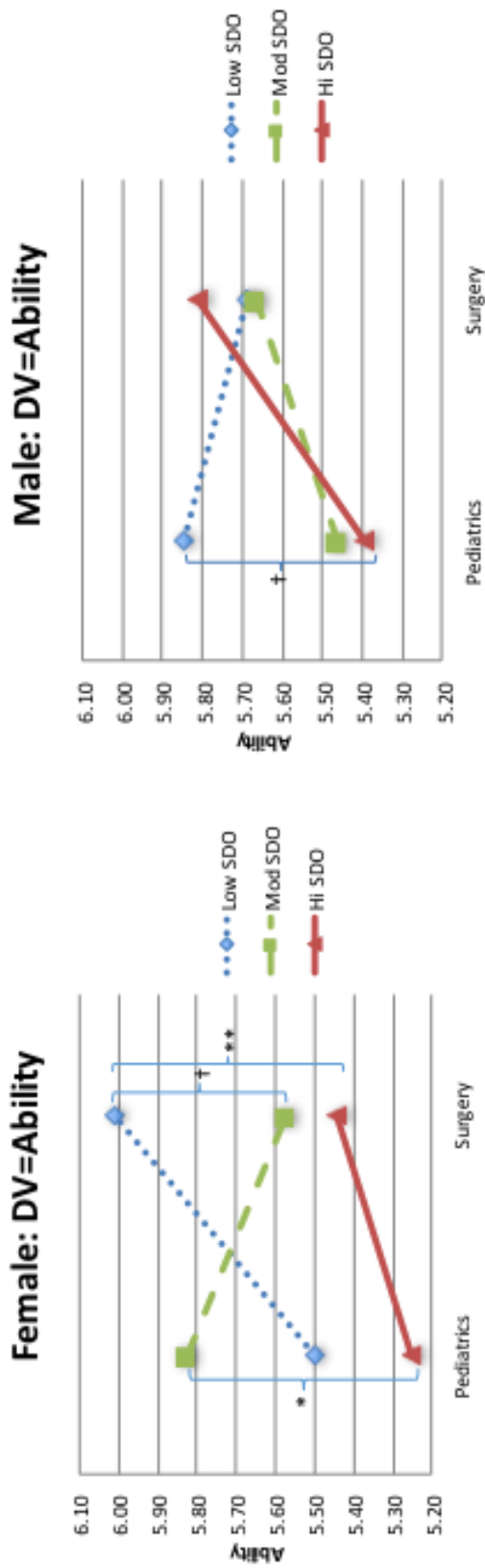


Figure 10. Graphs of ability means for the significant three-way interaction by target gender (male, female).
 Note. † $p < .10$, * $p < .05$, ** $p < .01$.

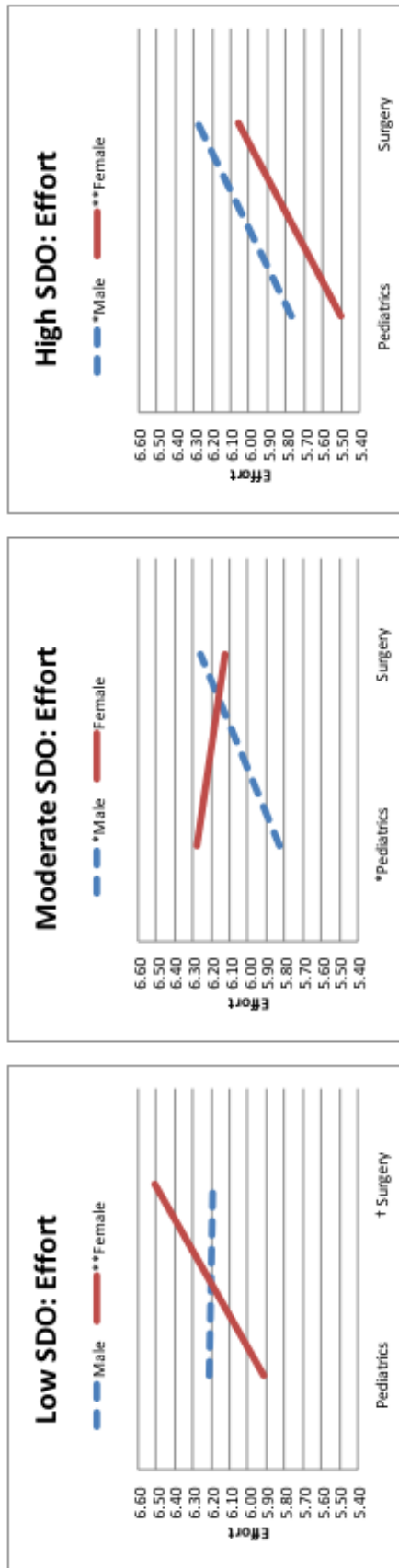


Figure 11. Graphs of effort characterization means for the three-way interaction by participant SDO Level (low, moderate, and high). Note. † p<.10, *p<.05, **p<.01. A significance symbol by a target gender indicates a significant within gender difference across specialty. A significance symbol by a target specialty indicates a significant across gender difference within the same specialty.

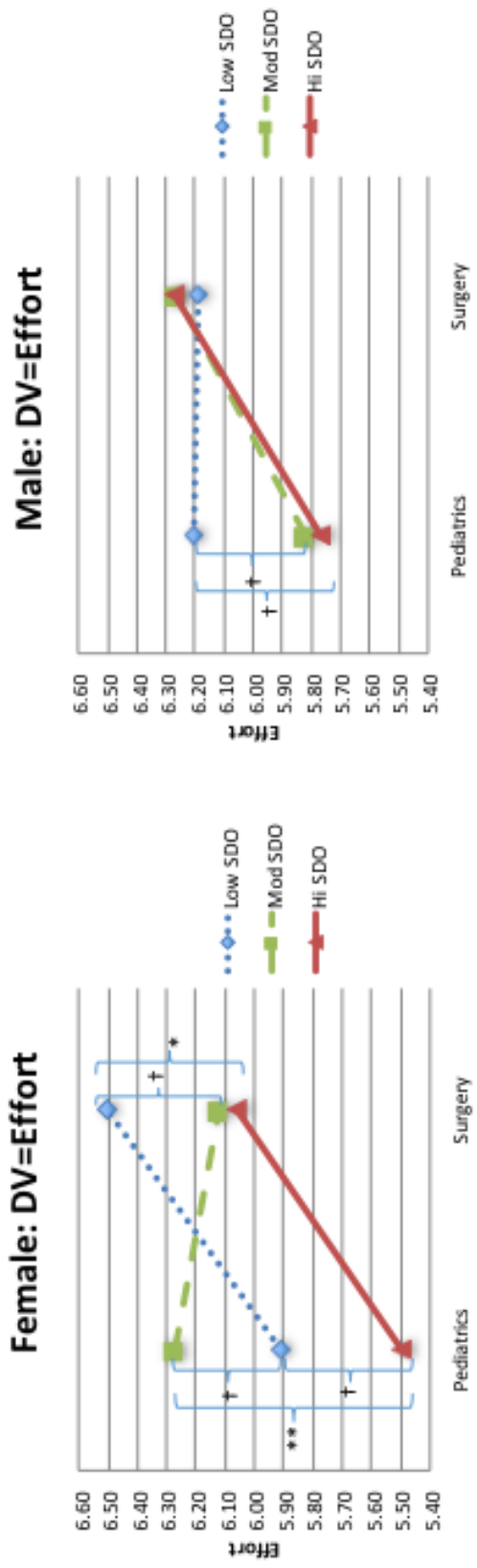


Figure 12. Graphs of effort means for the significant three-way interaction by target gender (male, female).
 Note. † $p < .10$, * $p < .05$, ** $p < .01$.

APPENDIX A1: COVER STORY

Study Title: IMPROVING THE PERFORMANCE EVALUATION OF PHYSICIANS IRB

Protocol ID:16-292

Principal Investigator: Dyan Ferraris, M.A.: daf2113@columbia.edu

The overall goal of this study is to help improve performance evaluation tools and processes in academic medical institutions. We hope to better understand how people form impressions of faculty in academic medical institutions and how they make employment related decisions about them (e.g., reappointment, promotion, and compensation).

Specifically, we are interested in addressing the balance in providing enough but not too much information to reviewers of performance evaluations. Not enough information makes it difficult for the reviewer to get a sense of the person and too much information makes it challenging for the reviewer to attend to all the details provided.^{[1],[2]}

Given this, we are interested in how the amount of information provided in a performance evaluation influences people's decision-making processes. Because of this, providing **your immediate and general impression is most useful for this study**.

Your participation is greatly appreciated and should take a total of 20-30 minutes.

[1] Kreber, C., Brook, P., & Policy, E. (2001). Impact evaluation of educational development programmes. *International Journal for Academic Development*, 6(2), 96-108.

[2] McLean, M., Cilliers, F., & Van Wyk, J. M. (2008). Faculty development: Yesterday, today and tomorrow. *Medical teacher*, 30(6), 555-584.

Please click ">>" below to review the Participant Informed Consent information.

APPENDIX A2: INFORMED CONSENT

INFORMED CONSENT

Protocol Title: Improving the Performance Evaluation of Physicians

IRB Protocol ID:16-292

Principal Investigator: Dyan Ferraris, M.A.: daf2113@columbia.edu

INTRODUCTION

You are being invited to participate in this research study called "Improving the Performance Evaluation of Physicians."

You must meet the following to qualify to take part in this research study:

- **18 years old or above**
- **You are located in the United States**
- **You agree to participate in this study**

Approximately 200 people will participate in this study and it will take 30 minutes of your time to complete.

WHY IS THIS STUDY BEING DONE?

This study is being done to determine how people form impressions of faculty in academic medical institutions and how they make employment related decisions about them (e.g., reappointment, promotion, and compensation). The practical significance of the research is to better understand how to improve performance evaluation tools and processes in academic medical institutions.

WHAT WILL I BE ASKED TO DO IF I AGREE TO TAKE PART IN THIS STUDY?

If you decide to participate, you will review the results of a completed performance evaluation of a faculty member. After that, you will be asked to complete a questionnaire, asking about your thoughts, perceptions, and employment related decisions of the faculty member you reviewed. All participation will occur online and at your convenience. You will not be asked to provide any identifying information.

WHAT POSSIBLE RISKS OR DISCOMFORTS CAN I EXPECT FROM TAKING PART IN THIS STUDY?

The risks associated with this study are minimal. Any felt discomfort would be no greater than what is typically encountered when anonymously reviewing the performance of a colleague. In order to mitigate this risk, please note that the profile you will view is hypothetical. Other minimal risks may include boredom and fatigue. To

mitigate these risks, please note that you are able to take a break during the course of participating in the study, as needed. Further, participation is voluntary; therefore, participation may be stopped at any time without penalty.

The principal investigator is taking precautions to keep your information confidential and prevent anyone from discovering or guessing your identity, as not collecting any identifying information and keeping all information on a password protected computer and locked in a file drawer.

WHAT POSSIBLE BENEFITS CAN I EXPECT FROM TAKING PART IN THIS STUDY?

There is no direct benefit to you for participating in this study. Participation may improve the efficacy of performance evaluation of faculty in academic medical institutions.

WILL I BE PAID FOR BEING IN THIS STUDY?

You will be paid to participate if you complete the entire survey and meet the qualification criteria.

WHEN IS THE STUDY OVER? CAN I LEAVE THE STUDY BEFORE IT ENDS?

The study is over when you have filled out the questionnaire. However, you can stop the study at any time even if you haven't finished. If you do not complete the entire questionnaire, it will not be considered a completed task.

PROTECTION OF YOUR CONFIDENTIALITY

The investigator will keep all written materials locked in a desk drawer in a locked office. Any electronic or digital information (including audio recordings) will be stored on a computer that is password protected. Regulations require that research data be kept for at least three years. No identifying information will be collected from you and your identity will remain anonymous.

HOW WILL THE RESULTS BE USED?

The results of this study will be published in journals and presented at academic conferences if accepted. Your name or any identifying information about you will not be published. This study is being conducted as part of the dissertation of the principal investigator.

WHO MAY VIEW MY PARTICIPATION IN THIS STUDY (please select one option)

I consent to allow written materials viewed at an educational setting or at a conference outside of Teachers College, Columbia University.

I **do not** consent to allow written materials viewed outside of Teachers College, Columbia University.

WHO CAN ANSWER MY QUESTIONS ABOUT THIS STUDY?

If you have any questions about taking part in this research study, you should contact the principal investigator, Dyan Ferraris at daf2113@columbia.edu.

If you have questions or concerns about your rights as a research subject, you should contact the Institutional Review Board (IRB) (the human research ethics committee) at 212-678-4105 or email

IRB@tc.edu. Or you can write to the IRB at Teachers College, Columbia University, 525 W. 120th Street, New York, NY 1002. The IRB is the committee that oversees human research protection for Teachers College, Columbia University.

Please click ">>" below to view the Participant's Rights page.

APPENDIX A3: PARTICIPANT'S RIGHTS

PARTICIPANT'S RIGHTS

I have read the informed consent. I have had ample opportunity to ask questions about the purposes, procedures, risks and benefits regarding this research study. I understand that my participation is voluntary. I may refuse to participate or withdraw participation at any time without penalty.

The researcher may withdraw me from the research at his or her professional discretion. If, during the course of the study, significant new information that has been developed becomes available which may relate to my willingness to continue my participation, the investigator will provide this information to me.

Any information derived from the research study that personally identifies me will not be voluntarily released or disclosed without my separate consent, except as specifically required by law. I can receive a copy of the Research Description and this Participant's Rights document by emailing the principal investigator. **Clicking the “>>” button below means that I agree to participate in this study.**

Thank you in advance for your participation!

APPENDIX B: STUDY INSTRUCTIONS

Instructions:

In the following section, you will be presented with information adapted from an actual promotion scenario. Specifically, you will be provided with:

A background information page of the faculty member's specialty, so you can get some idea of the type of work these individuals do and the types of people who occupy the role. A completed set of performance evaluation materials for a faculty member who recently went through a performance review.

Please read the scenario ***as if you were a faculty member in the individual's department.*** Following your review of the materials, you will be asked a series of questions including promotion related decisions and perceptions of the individual.

Do not worry about memorizing specific information on the materials provided. We are interested in the *general impression* the materials had on you and understanding the variety of responses that people have in these types of situations. There are no right or wrong answers.

Your responses will be confidential and anonymous. Thank you in advance for your help.

Click ">>" below to begin.

APPENDIX C1: PEDIATRICS BACKGROUND INFORMATION

Background Information on the specialty of Pediatrics in the United States¹

Definition:

- **Pediatrics** is a medical specialty that is primarily concerned with the physical, emotional, and social health of infants, children, and young adolescents (from birth to young adulthood).

Statistics of All Specialties compared to Pediatrics:

Table 1. Number and Percentage of Active Physicians

Specialty	Total Active Physicians	
	<i>Number</i>	<i>Percent</i>
All Specialties	829,962	100%
Pediatrics	56,262	6.8%

Table 2. Numbers and Percentage of Active Physicians by Gender

Specialty	Male		Female	
	<i>Number</i>	<i>Percent</i>	<i>Number</i>	<i>Percent</i>
All Specialties	558,794	67.4%	270,075	32.6%
Pediatrics	22,266	39.6%	33,944	60.4%

Table 3. Numbers and Percentage of Active Physicians by Age

Specialty	Under Age 55		Age 55 or Older	
	<i>Number</i>	<i>Percent</i>	<i>Number</i>	<i>Percent</i>
All Specialties	479,070	57.4%	349,248	42.6%
Pediatrics	2,712	57.9%	1,970	42.1%

Click “>>” below to begin your role as a *pediatrics faculty member*.

¹ Association of American Medical Colleges (AAMC), published in 2014.

APPENDIX C2: SURGERY BACKGROUND INFORMATION

Background Information on the specialty of General Surgery in the United States²

Definition:

- **General surgery** is a surgical specialty having a central core of knowledge common to all surgical specialties--anatomy, physiology, metabolism, immunology, nutrition, pathology, wound healing, shock and resuscitation, intensive care, and neoplasia.

Statistics of All Specialties compared to Surgery:

Table 1. Number and Percentage of Active Physicians

Specialty	Total Active Physicians	
	<i>Number</i>	<i>Percent</i>
All Specialties	829,962	100%
General Surgery	25,187	3.0%

Table 2. Numbers and Percentage of Active Physicians by Gender

Specialty	Male		Female	
	<i>Number</i>	<i>Percent</i>	<i>Number</i>	<i>Percent</i>
All Specialties	558,794	67.4%	270,075	32.6%
General Surgery	20,727	82.4%	4,437	17.6%

Table 3. Numbers and Percentage of Active Physicians by Age

Specialty	Under Age 55		Age 55 or Older	
	<i>Number</i>	<i>Percent</i>	<i>Number</i>	<i>Percent</i>
All Specialties	479,070	57.4%	349,248	42.6%
General Surgery	13,970	55.5%	11,206	44.5%

Click “>>” below to begin your role as a *surgery faculty member*.

² Association of American Medical Colleges (AAMC), published in 2014.

APPENDIX D1: PEDIATRICS DEAN'S LETTER (FEMALE CONDITION)



To: Faculty of the Pediatrics Department
From: Office of the Dean

Memo Re: Promotion Review for **Dr. Michelle Anderson**, Associate Professor of Pediatrics and Director of the Pediatrics Residency Program

As you know, **Dr. Michelle Anderson is up for promotion from Associate Professor to Full Professor of Pediatrics.** Her complete dossier of promotion materials is on file with the department.

For your review, in the following section you will find Dr. Michelle Anderson's summary performance evaluation. This summary evaluation contains average ratings from all the departmental faculty members.

Thank you for your continued commitment to the Faculty within the Department of Pediatrics.

APPENDIX D2: SURGERY DEAN'S LETTER (MALE CONDITION)



To: Faculty of the Surgery Department
From: Office of the Dean

Memo Re: Promotion Review for **Dr. Michael Anderson**, Associate Professor of Surgery and Director of the Surgery Residency Program

As you know, **Dr. Michael Anderson is up for promotion from Associate Professor to Full Professor of Surgery**. His complete dossier of promotion materials is on file with the department.

For your review, in the following section you will find Dr. Michael Anderson's summary performance evaluation. This summary evaluation contains average ratings from all the departmental faculty members.

Thank you for your continued commitment to the Faculty within the Department of Surgery.

APPENDIX E1: PEDIATRICS SUMMARY PERFORMANCE EVALUATION (FEMALE CONDITION)



**ANNUAL PEDIATRICS FACULTY PERFORMANCE EVALUATION:
Summary Evaluation for Dr. Michelle Anderson**

DEPARTMENT OF: Pediatrics	YEAR: 2016
NAME: Michelle Anderson	DEGREE: MD
RANK: Associate Professor	ROLE(S): Director of Residency Program
YEARS AT RANK: 7	TERM END DATE: 12/31/2016

Rating Scale:
 1=below expectations
 2=marginally meets expectations
 3=meets expectations
 4=exceeds expectations
 5=markedly exceeds expectations

N/A=not applicable

APPENDIX E1 (Continued)

Performance Indicators	Average Ratings	
	Individual's Average	Department Faculty Norm
Scholarship		
Peer-Reviewed Publications - quality and quantity; personal impact factor	4.6	4.2
Invited Presentations - regional/national/int'l; quality and quantity	4.5	4.2
Other Evidence of Scholarship, e.g., innovation, web-based materials	4.5	4.0
Teaching		
University Medical/Graduate Students – lectures, course development/directorship	4.5	4.0
Postdoctoral Fellows, House Staff/Clinical Fellows - #, venues, outcomes	4.4	3.8
Regional, National, International Teaching – quantity/quality	4.4	4.2
Teaching/Mentoring excellence, e.g., course evaluations, teaching awards mentee evaluations, productivity	4.6	4.2
Research		
Extramural and Internal Funding - type, \$, R dollar generation, renewal prospects, research density, ISMMS pilot funding	4.6	4.2
Innovations - patents, patent applications, licenses	4.5	4.2
Clinical Trials - scope, patient enrollment, financials	4.4	4.2
Clinical		
Patient Care, e.g., quality, malpractice, certification, patient satisfaction	4.5	4.2
Work Relative Value Units (wRVUs) -- performance against benchmarks	4.4	3.8
Financial Results – clinical receipts, expenses, deficits	4.4	4.0
Service		
Service/Leadership, e.g., institutional committees	4.5	4.2
Service to Primary Dept/Institute, e.g., committees, administration	4.6	4.1
Mentoring, Active participation as a mentor or mentee	4.5	4.2
External Service/Leadership, e.g., study sections, professional society committees, editorial boards, public advocacy (e.g. testifying before Congress, lab tours to policy makers, advocating Federal \$ for science), public education/ talks to lay audiences/ high school science outreach	4.4	4.0
Professionalism towards faculty, trainees, staff	4.5	4.1
FINAL OVERALL RATING:	4.49	4.1

APPENDIX E2: SURGERY SUMMARY PERFORMANCE EVALUATION (MALE CONDITION)



**ANNUAL SURGERY FACULTY PERFORMANCE EVALUATION:
Summary Evaluation for Dr. Michael Anderson**

DEPARTMENT OF: Surgery	YEAR: 2016
NAME: Michael Anderson	DEGREE: MD
RANK: Associate Professor	ROLE(S): Director of Residency Program
YEARS AT RANK: 7	TERM END DATE: 12/31/2016

Rating Scale:
 1=below expectations
 2=marginally meets expectations
 3=meets expectations
 4=exceeds expectations
 5=markedly exceeds expectations
 N/A=not applicable

APPENDIX E2 (Continued)

Performance Indicators	Average Ratings	
	Individual's Average	Department Faculty Norm
Scholarship		
Peer-Reviewed Publications - quality and quantity; personal impact factor	4.6	4.2
Invited Presentations - regional/national/int'l; quality and quantity	4.5	4.2
Other Evidence of Scholarship, e.g., innovation, web-based materials	4.5	4.0
Teaching		
University Medical/Graduate Students – lectures, course development/directorship	4.5	4.0
Postdoctoral Fellows, House Staff/Clinical Fellows - #, venues, outcomes	4.4	3.8
Regional, National, International Teaching – quantity/quality	4.4	4.2
Teaching/Mentoring excellence, e.g., course evaluations, teaching awards mentee evaluations, productivity	4.6	4.2
Research		
Extramural and Internal Funding - type, \$, R dollar generation, renewal prospects, research density, ISMMS pilot funding	4.6	4.2
Innovations - patents, patent applications, licenses	4.5	4.2
Clinical Trials - scope, patient enrollment, financials	4.4	4.2
Clinical		
Patient Care, e.g., quality, malpractice, certification, patient satisfaction	4.5	4.2
Work Relative Value Units (wRVUs) -- performance against benchmarks	4.4	3.8
Financial Results – clinical receipts, expenses, deficits	4.4	4.0
Service		
Service/Leadership, e.g., institutional committees	4.5	4.2
Service to Primary Dept/Institute, e.g., committees, administration	4.6	4.1
Mentoring, Active participation as a mentor or mentee	4.5	4.2
External Service/Leadership, e.g., study sections, professional society committees, editorial boards, public advocacy (e.g. testifying before Congress, lab tours to policy makers, advocating Federal \$ for science), public education/ talks to lay audiences/ high school science outreach	4.4	4.0
Professionalism towards faculty, trainees, staff	4.5	4.1
FINAL OVERALL RATING:	4.49	4.1

APPENDIX F: QUESTIONNAIRE INTRODUCTION

You have completed the review of promotion materials.

In the following sections, you will be asked to respond to a series of questions. As a reminder, we are particularly interested in ***first impressions formed on the basis of minimal information***.

Please answer every question ***even if you feel you do not have enough information to respond*** – remember, this study is about first impressions, so it is ***your immediate reaction to this individual that is of interest to us***.

Before you begin, please type in the name of the Faculty Member you are reviewing:

Click ">>" below to respond to the questions.

APPENDIX G: STUDY DEBRIEF

Study Debrief:

Thank you for participating in this research. The specific purpose of the current study is to investigate differences in perceptions of physician leaders and how it impacts employment related decisions. The researcher is investigating whether there are differences in perceptions of male and female leaders within and between male-dominated and less male-dominated specialties (see Eagly & Karau, 2002, and Heilman, 2012, for further information).

Further, the researcher hypothesizes that differential explanations or attributions for someone's success may influence the allocation of organizational rewards. According to attribution theory, when an individual performs outside of what is expected, the results are often credited to factors beyond the individual's ability (see Deaux & Emswiller, 1974). Thus, because women are expected to fail or underperform as compared to men in a male-typed job (i.e., physician leadership), when a woman is instead successful, her achievements are viewed as a result of working especially hard (Deaux et al., 1974; Heilman, 1983; Swim & Sanna, 1996). Such attributions support the perpetuation of descriptive stereotypes of women and inhibit the use of objective evidence of her performance.

The study's goal is to use the data to examine the social cognitive processes that accompany performance evaluation in order to alleviate possibilities of unintended and unconscious gender bias.

Please note that all of your responses to the questions in the questionnaire will be kept anonymous and confidential and will only be used for this research.

If you have any questions or concerns about this study, please feel free to contact the primary investigator, Dyan Ferraris at daf2113@columbia.edu.

Thank you again for your participation!

Click ">>" below to finalize your participation.

References mentioned above:

- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29, 80-85.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3), 573.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In Staw, B., & Cummings, L. Eds. *Research in organizational behavior* (Vol. 5). Greenwich, CT: JAI.
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, 32, 113-135.
- Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, 22, 507-522.

APPENDIX H1: STUDY MEASURES – PROMOTION

Instructions: Please read each of the prompts carefully and rate the individual on a scale from 1 to 7.

- 1. How likely would you be to recommend that this institution invest in leadership development for this individual?**

not at all
likely 1 2 3 4 5 6 7 extremely
likely

- 2. How likely would you be to recommend this person to be placed in a more senior position in the department?**

not at all
likely 1 2 3 4 5 6 7 extremely
likely

- 3. How likely would you be to recommend this person for a promotion?**

not at all
likely 1 2 3 4 5 6 7 extremely
likely

APPENDIX H2: STUDY MEASURES – PERFORMANCE

Instructions: Please read each of the prompts carefully and rate the individual on a scale from 1 to 7.

1. How successful do you think this individual is in his/her role?

not at all
successful 1 2 3 4 5 6 7 extremely
successful

2. Overall, how would you rate this individual's performance in his/her role?

extremely
low 1 2 3 4 5 6 7 extremely
high

3. Please indicate your overall satisfaction with this individual in his/her role.

not at all
satisfied 1 2 3 4 5 6 7 extremely
satisfied

4. What is your personal view of this individual's overall effectiveness?

not at all
effective 1 2 3 4 5 6 7 extremely
effective

5. Rate this individual's potential to excel in his/her career.

extremely
low 1 2 3 4 5 6 7 extremely
high

APPENDIX I1: STUDY MEASURES – ABILITY ATTRIBUTIONS

Instructions: For each adjective, please select one number between 1 and 7.

6. For the following items, please indicate to what extent you think the individual is...

Excellent	not at all	1	2	3	4	5	6	7	very much
Superb	not at all	1	2	3	4	5	6	7	very much
Outstanding	not at all	1	2	3	4	5	6	7	very much
Unique	not at all	1	2	3	4	5	6	7	very much
Exceptional	not at all	1	2	3	4	5	6	7	very much
Unparalleled	not at all	1	2	3	4	5	6	7	very much
Brilliant	not at all	1	2	3	4	5	6	7	very much
Gifted	not at all	1	2	3	4	5	6	7	very much

APPENDIX I2: STUDY MEASURES – EFFORT ATTRIBUTIONS

Instructions: For each adjective, please select one number between 1 and 7.

1. For the following items, please indicate to what extent you think the individual is...

Hardworking	not at all	1	2	3	4	5	6	7	very much
Conscientious	not at all	1	2	3	4	5	6	7	very much
Dependable	not at all	1	2	3	4	5	6	7	very much
Meticulous	not at all	1	2	3	4	5	6	7	very much
Thorough	not at all	1	2	3	4	5	6	7	very much
Diligent	not at all	1	2	3	4	5	6	7	very much
Dedicated	not at all	1	2	3	4	5	6	7	very much
Careful	not at all	1	2	3	4	5	6	7	very much
Perfectionistic	not at all	1	2	3	4	5	6	7	very much

APPENDIX J1: SOCIAL DOMINANCE ORIENTATION MEASURE

Instructions: Please answer some questions about yourself.

There are many kinds of groups in the world: men and women, ethnic and religious groups, nationalities, political factions. *How much do you support or oppose the following ideas about groups in general?*

1. In setting priorities, we must consider all groups.

Extremely Oppose 1 2 3 4 5 6 7 Extremely Favor

2. We should not push for group equality.

Extremely Oppose 1 2 3 4 5 6 7 Extremely Favor

3. Group equality should be our ideal.

Extremely Oppose 1 2 3 4 5 6 7 Extremely Favor

4. Superior groups should dominate inferior groups.

Extremely Oppose 1 2 3 4 5 6 7 Extremely Favor

APPENDIX J2: Distribution of SDO Scores by Tertiary Split Category

	N	Mean	SD	Range	Minimum	Maximum
Low SDO	72	1.00	0.00	0.00	1.00	1.00
Moderate SDO	71	1.90	0.46	1.25	1.25	2.50
High SDO	64	3.90	0.97	4.25	2.75	7.00
Total	207	2.21	1.34	6.00	1.00	7.00

APPENDIX K: MANIPULATION CHECK QUESTIONS

Instructions: Please tell us about the individual you reviewed.

2. Please indicate the gender of the person you evaluated: [pull down menu]

- Male
- Female
- Other (Please specify _____)

3. Please indicate the person's specialty: [pull down menu]

- _____.
- Pediatrics
 - Surgery
 - Other (Please specify _____)

4. In general, do you think more men or more women occupy this individual's specific role?

more men 1 2 3 4 5 6 7 more women

5. Please provide comments, questions, or suggestions regarding this study (optional):

APPENDIX L: DEMOGRAPHIC QUESTIONNAIRE

Instructions: Please provide some demographic information about yourself.

5. Please indicate your gender: [pull down menu]

- Male
- Female
- Other (Please specify _____)

6. Please indicate your race/ethnicity (select all that apply):

- African American or Black
- Asian or Asian American
- Hispanic or Latino
- White
- American Indian or Alaska Native
- Native Hawaiian or other Pacific Islander
- Other (Please specify _____)

7. Please indicate your country of origin (i.e., where you were born): _____

8. What year were you born? _____.

9. How many years of full time work have you completed? _____

10. What was your total household income before taxes during the past 12 months (select one)?

- Less than \$25,000
- \$25,000 to \$49,999
- \$50,000 to \$74,999
- \$75,000 to \$99,999
- \$100,000 to \$149,999
- \$150,000 to \$199,999
- \$200,000 or more

11. What is your marital status?

- Single (never married)
- Married
- Separated
- Widowed
- Divorced

12. What is the highest degree or level of education you have completed?

- Some college, no degree
- Associate's degree
- Bachelor's degree
- Graduate or professional degree
- Other

13. In which country did you receive the above education? [pull-down menu]:_____.

14. What industry do you work in? [Drop down menu]

- Accommodations
- Accounting
- Advertising
- Aerospace
- Agriculture & Agribusiness
- Air Transportation
- Apparel & Accessories
- Auto
- Banking
- Beauty & Cosmetics
- Biotechnology
- Chemical
- Communications
- Computer
- Construction
- Consulting
- Consumer Products
- Education
- Electronics
- Employment
- Energy
- Entertainment & Recreation
- Fashion
- Financial Services
- Food & Beverage
- Health
- Information
- Information Technology
- Insurance
- Journalism & News
- Legal Services
- Manufacturing
- Media & Broadcasting
- Medical Devices & Supplies
- Motion Pictures & Video
- Music
- Pharmaceutical
- Public Administration
- Public Relations
- Publishing
- Real Estate
- Retail

- Service
- Sports
- Technology
- Telecommunications
- Tourism
- Transportation
- Travel
- Utilities
- Video Game
- Web Services
- Other

15. What level are you in your career?

- Entry Level
- Professional Level
- First-Level Management
- Mid-Level Management
- Technical & Engineering
- Senior-Level Management
- Executive-Level
- N/A

APPENDIX M: Additional Results

Table M1

Correlations Among all Demographic Variables and Independent and Dependent Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<u>Demographic variables</u>														
1 Gender	--													
2 Ethnicity	-.01	--												
3 Marital Status	.08	-.13	--											
4 Occupation	-.03	.03	.11	--										
5 Occupation Level	.14	.05	.01	-.14*	--									
6 Age	-.04	.06	-.28**	-.11	.18**	--								
7 Years of FT Work	.03	.13	-.11	-.07	.20**	.48**	--							
8 Income	.20**	-.07	-.16*	-.20**	-.02	.03	-.03	--						
<u>Independent variables</u>														
9 Target Gender	-.02	.02	-.02	-.03	.01	-.06	-.01	-.01	--					
10 Target Specialty	-.02	.14*	.05	.08	.06	.01	.00	-.12	-.03	--				
11 SDO	.14*	.15*	-.05	-.06	.08	.04	-.04	.11	-.04	-.07	--			
<u>Dependent variables</u>														
12 Promotion	-.21**	-.02	.03	.04	-.04	.03	.00	.06	.03	.16*	-.14*	--		
13 Performance	-.13	-.03	.08	.03	-.00	.06	-.00	.05	-.04	.15*	-.14*	.77**	--	
14 Ability	-.10	-.03	.03	.05	-.01	-.00	-.00	.05	.01	.10	-.18**	.49**	.57**	--
15 Effort	-.27**	-.05	-.07	.06	-.04	.10	.02	.04	-.01	.23**	-.23**	.64**	.63**	.77**

Note. $N = 207$. * $p < .05$ ** $p < .01$. Participant gender code dummies include 1=female 2=male. Participant ethnicity code dummies include 1=Black/African-American 2=Asian/Asian-American 3=Hispanic/Latino(a) 4=White 5=Native American or Alaskan 6=Hawaiian or Pacific Islander. Marital status code dummies include 1=Married 2=Widowed 3=Divorced 4=Separated 5=Never married. Occupation Level code dummies include 1=Entry Level 2=Professional Level 3= First-Level Management 4=Mid-Level Management 5= Technical & Engineering 6= Senior-Level Management 7= Executive-Level 8=N/A. Target gender code dummies include 1=female 2=male. Target specialty code dummies include 1=pediatrics 2=surgery. Social dominance orientation (participant SDO), promotion, performance, ability attribution, and effort attribution were measured on a 7-point scale. Higher means indicate that the target was rated as higher on the variable of interest.

Table M2

Multivariate MANOVA Results for Participant Gender on Dependent Variables

	<i>Value</i>	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Participant Gender	.90	9.57	1	194	0.00

Table M3

Univariate ANOVA Results for Participant Gender on Dependent Variables

Dependent Variables	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>Sig.</i>
Promotion	9.47	1	205	.002
Performance	3.69	1	205	.056
Ability	2.20	1	205	.14
Effort	15.74	1	205	.000

Table M4

Dummy Coding Scheme for Multiple Regression and Mediation Analysis

Group	Condition and Participant SDO Level	Dummy Codes and Regression Variables										
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
		gender	specialty	high SDO	moderate SDO	gender X specialty	gender X high SDO	gender X moderate SDO	specialty X high SDO	specialty X moderate SDO	gender X specialty X high SDO	gender X specialty X moderate SDO
1	Female Pediatrics, Low SDO	0	0	0	0	0	0	0	0	0	0	0
2	Female Pediatrics, Moderate SDO	0	0	0	1	0	0	0	0	0	0	0
3	Female Pediatrics, High SDO	0	0	1	0	0	0	0	0	0	0	0
4	Female Surgery, Low SDO	0	1	0	0	0	0	0	0	0	0	0
5	Female Surgery, Moderate SDO	0	1	0	1	0	0	0	0	1	0	0
6	Female Surgery, High SDO	0	1	1	0	0	0	0	1	0	0	0
7	Male Pediatrics, Low SDO	1	0	0	0	0	0	0	0	0	0	0
8	Male Pediatrics, Moderate SDO	1	0	0	1	0	0	1	0	0	0	0
9	Male Pediatrics, High SDO	1	0	1	0	0	0	0	0	0	0	0
10	Male Surgery, Low SDO	1	1	0	0	1	0	0	0	0	0	0
11	Male Surgery, Moderate SDO	1	1	0	1	1	0	1	0	1	0	1
12	Male Surgery, High SDO	1	1	1	0	1	1	0	1	0	1	0

Note. $N = 207$. Dummy codes were recoded for the regression analysis for interpretation purposes. Participant gender code dummies include 0=female 1=male. Target specialty code dummies include 0=pediatrics 1=surgery. Participant high SDO dummies include 0=moderate, low 1=high. Participant moderate SDO dummies include 0=low, high 1=moderate. Participant low SDO is the reference group. Dummies 5-11 were calculated by multiplying the dummy variables included in the interaction.