EXAMINING THE ROLE OF FOLLOWERS' LEADER BEHAVIOR EXPECTATIONS ON EVALUATIONS OF MEN AND WOMEN LEADERS

By

Connor Eichenauer

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Master of Arts

2021

ABSTRACT

EXAMINING THE ROLE OF FOLLOWERS' LEADER BEHAVIOR EXPECTATIONS ON EVALUATIONS OF MEN AND WOMEN LEADERS

By

Connor Eichenauer

Descriptive and prescriptive gender stereotypes research suggests that men are expected to engage in more agentic behaviors and women in more communal behavior as leaders. However, gender and leadership research has not explicitly measured expectations of men and women leaders nor considered how followers evaluate leaders who fail to fulfill or exceed expectations for agentic and communal behaviors. This vignette study sought to accomplish both by measuring follower expectations for a communal and an agentic leader behavior, manipulating these behaviors, and measuring follower perceptions and evaluations to investigate whether congruence between followers' expectations and supervisors' subsequent behavior produces similar evaluations of men and women leaders. Results indicate followers expected higher levels of communal behavior from the female than the male supervisor, but no differences were found in expectations for agentic behavior, suggesting a double standard in gender role-congruent behavior expectations. Regardless of whether expectations were exceeded or unmet, supervisor gender did not moderate effects of agentic or communal behavior expectations-perceptions incongruence on evaluations of effectiveness or liking in polynomial regression analyses. Implications and future research directions are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

On the surface, bias against women in organizational leadership seems to have steadily decreased from the overt discrimination women used to consistently face. For example, while a 1953 public opinion poll found two-thirds of Americans said they would prefer having a male boss if given the choice, a 2017 poll asking the same question found the majority now say they have no preference when it comes to the gender[1] of their boss (Gallup, 2017). Empirical findings of gender bias in leadership effectiveness ratings also appear to be subsiding; compared to past meta-analytic evidence that found substantial gender differences in ratings of effectiveness (Eagly et al., 1992), a more recent meta-analysis found no gender differences in effectiveness ratings across contexts (Paustian-Underdahl et al., 2014). This more recent meta-analysis concluded that study publication date predicted findings of bias such that recent studies tend to not find as much gender bias in leader effectiveness ratings as was found in older research. These indicators seem to suggest that women are being increasingly accepted as leaders.

Despite this progress, women are still underrepresented in organizational leadership positions today, especially in the top echelons of organizations. According to 2020 data from S&P 500 companies, women hold 45 percent of all jobs but make up only 37 percent of first- and mid-level managers, 27 percent of senior- and executive-level managers, and less than six percent of CEOs (Catalyst, 2020). This disparity could be due to what Ely and colleagues describe as a "second-generation" of bias against women in leadership where organizational structures and societal beliefs about gender and leadership combine to help maintain a status quo that favors men (Ely et al., 2011). Uncovering how this subtle bias impacts women requires consideration of the process by which leaders are evaluated.

[1] While gender is not considered to be a dichotomous concept, for the purposes of this paper, I will be referring to "male and female" or "men and women" as the two predominant gender groups.

The most popular theoretical explanations for the existence of bias against women in leadership have been offered by Role Congruity Theory (Eagly & Karau, 2002) and Heilman's Lack of Fit model (1983, 1995, 1997, 2001). However, much of what we know about how these theories explain gender bias is arguably outdated. Since these theories were proposed, modern conceptualizations of effective leadership have evolved from formerly favoring agentic qualities (e.g., confidence, assertiveness) to now more commonly emphasizing the importance of communal leadership attributes (e.g., humility, empathy; Avolio et al., 2009; Koenig et al., 2011). This is contrary to the original assumptions underlying Role Congruity Theory and the Lack of Fit Model that men's traditional gender role is more closely aligned with the role of leaders. Because less emphasis is being placed on agentic leader behaviors (i.e., role congruent for men) and more on communal leader attributes (i.e., role congruent for women), some have suggested this increased alignment between the stereotypical gender role of women and that of leaders could (or perhaps *should*) result in women having an advantage over men as leaders in modern times (Paustian-Underdahl et al., 2014). If communal behaviors are now more tied to effective leadership than agentic ones, and if men engage in fewer communal behaviors, it follows that male leaders should receive lower leader evaluations than women.

Additionally, past research has documented how similar leader behaviors are often differently rewarded and penalized for male and female leaders. For example, studies have demonstrated that women are often penalized for acting in ways that are perceived as agentic, even when men are rewarded for engaging in similar behaviors (e.g., Rudman & Glick, 2001). At this time, there is less evidence to suggest that evaluations of male leaders might similarly suffer from displaying communal (i.e., role incongruent) behaviors. Under conditions in which communal leadership is considered to be most effective, are men similarly penalized for

violating gender norms? Literature in other areas of organizational research also suggests expectations can play a critical role in evaluative bias against women. For example, research on organizational citizenship behaviors (OCBs) has demonstrated that higher amounts of workplace helping behaviors are expected of women than men, and failure to engage in helping OCBs results in a more severe penalty for women in performance evaluations (Heilman & Chen, 2005).

Thus, explicitly measuring follower expectations for specific leader behaviors could lend greater insight into whether bias against female leaders exists and if changing conceptualizations of leadership result in similar evaluative biases against men. However, empirical work on gender and leadership has not considered several aspects of the role of follower expectations in how men and women are perceived and evaluated as leaders. While implicit leadership theories like Leader Categorization Theory (Lord et al., 1984) do suggest leadership is evaluated based on prototypes of follower expectations, this line of research tends to a) simultaneously assess expectations and behaviors when examining differences in how male and female leaders are evaluated, b) assess only broad categories of leader behavior or leadership styles, and c) only consider effects for women.

To examine these ideas further, the present research has two primary goals: 1) to consider the role of followers' expectations for specific leader behaviors in how they evaluate leaders, and 2) to further investigate how leader behaviors exhibited by male and female leaders are rewarded and penalized under different levels of follower expectations and expectation fulfillment. Additionally, a methodological contribution will be made by assessing expectation-behavior congruence using polynomial regression, which has been used to measure congruence in other literatures but is not typically utilized in research on gender and leadership.

This paper is structured as follows. First, I will briefly review the literature on women in leadership, including the most popular theories relevant to the study of gender differences and gender bias in leadership. I will then describe Leader Categorization Theory and how leader prototypes might disadvantage women in spite of evolving conceptualizations of leadership. Finally, a series of hypotheses will be presented to test whether follower expectations differ as a function of leader gender and how expectation-incongruent behaviors are evaluated for men and women leaders.

Women and Leadership

Women hold nearly half of all jobs in the United States and 51.8% of all "management, professional, and related occupations" (U.S., 2020) but continue to be underrepresented in leadership positions. Among S&P 500 companies, women make up 44.7% of all employees, but only 36.9% of first and mid-level managers, 26.5% of executive and senior-level managers, and a mere 5.8% of CEOs (Catalyst, 2020). Progress is evident but a gap in representation remains, and data indicate only modest progress towards achieving parity has been made in the last five years (McKinsey, 2020).

Some of the progress that has been made can be attributed to improving societal attitudes towards women in leadership. Gallup began polling Americans 68 years ago about their attitudes towards women in management roles by asking, "If you were taking a new job and had your choice of a boss, would you prefer to work for a man or a woman?" In 1953, 66 percent of respondents reported preference for a male boss, while only five percent said they would prefer having a female boss–a 61 percentage point difference. At that time, only one in four reported no preference in their choice of gender of boss. These attitudes stand in stark contrast to 2017 poll results that showed 23 percent said they would prefer working for a man, 21 percent would

prefer a woman, and over half (55%) had no preference in the gender of their boss (Gallup, 2017). However, there is still a small subset of people who hold openly hostile beliefs about women leaders; this is evident in a poll that showed nearly one in ten Americans said they would not vote for a well-qualified woman for President of the United States (McCarthy, 2020). Further, poll results indicating progress might lead to overly optimistic conclusions; these statistics should be interpreted with caution as socially desirable responding bias might be skewing results to appear more favorable to women compared to the internal beliefs people actually hold.

Evaluative bias against women, stemming from both prejudicial beliefs and more subtle biases, is believed to be one of the primary factors responsible for the underrepresentation of women in leadership positions. The phenomenon has been described as the "glass ceiling" (Morrison et al., 1987) or "glass labyrinth" (Eagly & Carli, 2007) to characterize the invisible barriers women face as they attempt to advance through the ranks of organizations. Despite initiatives aimed at increasing women's representation at the top of organizational hierarchies, the subtle strength of gender stereotypes and expectations of women as leaders have led to an endurance of representation disparities (Heilman, 2001; Ely et al., 2011). As such, continued research is warranted into evaluative biases and how they might disadvantage aspiring women leaders.

Role Congruity Theory

One theoretical explanation for the existence of gender bias in leadership is through the lens of the Role Congruity Theory of prejudice toward female leaders (Eagly & Karau, 2002). Role Congruity Theory was born out of Social Role Theory (Eagly, 1987; Eagly & Wood, 2012), which argues that individuals are expected to engage in activities and exhibit behaviors that are

consistent with their culturally defined gender roles. Men's traditional gender role encourages them to display an *agentic* orientation (e.g., assertiveness, confidence, competitiveness) while women are expected to display a *communal* orientation (e.g., warm, compassionate, considerate; Bakan, 1966; Eagly, 1987). Despite advances in gender equality–evidenced by increasing ratings of women's competence in the workplace over time (e.g., Gallup, 2017)–gendered stereotypes of agency and communion have endured well into the 21st century (Eagly et al., 2020). These stereotypic perceptions extend to those in leadership, where empirical evidence indicates women in top leadership positions are viewed as more communal (e.g., more caring and relationship-oriented) and men as more agentic (e.g., more forceful and task-oriented; Rosette & Tost, 2010; Lyness & Heilman, 2006).

Role Congruity Theory then suggests that the perceived alignment between the gender roles to which women and men are ascribed and the role of leaders determines how they are viewed as leaders. Research has demonstrated that the role of a leader has traditionally been viewed as requiring agentic qualities such as competitiveness, self-confidence, aggressiveness, and ambition (e.g., Schein, 1973, 1975, 2001; Heilman et al., 1989; Massengill & di Marco, 1979; Lee & Hoon, 1993). Accordingly, men are viewed as well aligned to the role of a leader. However, women's communal gender role is dissimilar from the role leaders are traditionally expected to play. This mismatch, or incongruence, between the stereotypical role of women and the prototypical traits of leaders is hypothesized to explain bias against women in leadership positions (Eagly & Karau, 2002). Specifically, Role Congruity Theory posits that this incongruence results in two types of prejudice against women. The first can be attributed to a descriptive bias, or the perception of differences between the stereotypic qualities of women and the qualities that are required of leaders. The result of this descriptive bias is that women are

viewed as having less potential for leadership due to leadership abilities being stereotypically associated with men more than women. Another outcome of descriptive stereotypes is that people tend to believe men and women engage in different styles of leadership, whether this is true or not. A meta-analysis comparing the leadership styles of men and women indicated that while women are expected to excel in interpersonally oriented leadership and men in task-oriented leadership, they did not actually differ in leadership styles utilized across a sample of organizational studies (Eagly & Johnson, 1990). The second form of prejudice results from prescriptive beliefs about gender roles (i.e., how women and men ought to behave). Because of societal pressures to conform to gender roles, men and women have separate leadership styles in which they are expected to engage. In addition to how women ought to behave, prescriptive stereotypes also dictate how women ought not to behave (Heilman, 2001). Deviating from these injunctive norms for workplace behavior, including leadership style, tends to result in disapproval from others (Cialdini & Trost, 1998). While prescriptive bias could theoretically affect leaders of both genders, Eagly and Karau (2002) insinuate women are more disadvantaged since leadership roles require more agency than communion. They accordingly named their theory "Role Congruity Theory of prejudice *toward female leaders"* and concluded the following:

> Women leaders' choices are thus constrained by threats from two directions: Conforming to their gender role would produce a failure to meet the requirements of their leader role, and conforming to their leader role would produce a failure to meet the requirements of their gender role (2002; p. 576).

Prescriptive bias against women, as suggested by Role Congruity Theory, has been examined empirically by considering the relationship between leader behaviors and leadership

evaluations for leaders of different genders. That is, does displaying (or failing to display) certain leader behaviors lead to equal rewards (or punishments) in leader evaluations for men and women? Research has indicated the answer to this question is often "no," especially when women exhibit role incongruent agentic leader behaviors. For example, research has established that women who attempt to adopt a more masculine leadership style (i.e., role incongruent) are subsequently penalized in leadership evaluations compared to those of men despite exhibiting similar behaviors as men (e.g., Eagly et al., 1992; Johnson et al., 2008). In addition to receiving less favorable leader evaluations, studies suggest women displaying counter-role agentic behaviors can also result in others' perceptions of coldness, interpersonal derogation, and general dislike; these reactions against women have been described as a "backlash effect" (e.g., Heilman & Chen, 2005, Rudman, 1998, Flynn & Ames, 2006). Less research attention has been devoted to considering backlash men might face as leaders, although evidence from other areas of organizational behavior indicates men might also suffer consequences for violating prescriptive gender norms. These consequences have been observed when men display counter-role behavior such as asking for family leave (Allen & Russell, 1999; Wayne & Cordeiro, 2003; Rudman & Mescher, 2013) or failing to engage in masculine behaviors expected of them (Chen, 2008; Moss-Racusin et al., 2010). One study that did specifically look at leaders demonstrated that while women who succeeded in male-dominated roles were viewed as hostile and disliked, successful men in a stereotypically feminine role were viewed as wimpy and were not respected (Heilman & Wallen, 2010). However, this study manipulated gender-typed roles but did not consider whether men and women actually engaged in counter-role behavior. More attention needs to be dedicated to understanding whether and under what conditions men might face consequences for engaging in counter-role behavior as leaders.

Lack of Fit Model

Heilman's Lack of Fit Model (1983, 1995, 1997, 2001) offers a similar theoretical explanation for bias against female leaders. Heilman's seminal work on the Lack of Fit Model focused more broadly on the occurrence of sex bias in various organizational settings (1983), although this was followed with subsequent publications that specifically applied the model to women in leadership (Heilman et al., 1995; Heilman, 2001). The Lack of Fit model posits that descriptive and prescriptive gender stereotypes combine to create a perception of a "lack of fit" between attributes typically associated with women and those believed to be required to successfully perform certain jobs, such as leadership roles. Lack of fit is proposed to result in self-directed bias (i.e., self-limiting behavior) and other-directed bias (i.e., discrimination) that disadvantage women by preventing them from ascending up the corporate ladder. While the Lack of Fit model is similar in many ways to Role Congruity Theory, one key difference is that the Lack of Fit Model acknowledges that men should also be subject to consequences for violating prescriptive gender norms. However, it suggests violations of gender norms should result in different consequences for men and women; Heilman postulates that while women who violate role norms are perceived as lacking femininity and are accordingly viewed as cold and disliked, penalties men receive will be related to their perceived lack of masculinity (e.g., perceptions of passiveness, lack of respect; Heilman, 2012).

Leadership Prototypes

Role Congruity Theory and the Lack of Fit model indicate that men and women are held to different role expectations as leaders, and that these expectations constrain leaders in what types or styles of leadership they tend to exhibit. These constraints are suggested to disadvantage women in particular due to the fact that violations of role expectations often lead to penalties in

leadership evaluations. Our understanding of how followers' expectations help determine how leaders are evaluated comes from research on implicit leadership theories (ILTs) such as Leader Categorization Theory.

Leader Categorization Theory (Lord et al., 1984) is an implicit leadership theory that proposes that individuals possess mental schemas, or prototypes, for how they believe leaders should act. Humans rely on prototypes to help simplify cognitive processes as we perceive and draw patterns from the world around us. Leader Categorization Theory suggests that people leverage their past experiences with leaders to build cognitive knowledge structures about leadership and the role of leaders. As individuals continue adding experiences to these knowledge structures over time, the cognitive processes through which they perceive and make judgements about unfamiliar people become less efficient. To streamline the cognition process, individuals use category knowledge they possess of leaders to generate prototypes, or ideal visions of leader attributes, traits, and behaviors. They then use these prototypes as a perceptual reference point to which they compare unfamiliar individuals to determine if they sufficiently match their schema of a leader.

While going through this process leads to cognition efficiencies, it also can lead to bias against women because leader prototypes are often contaminated by things like gender role expectations or other perceptions of individuals in social groups. That is, many individuals likely hold different prototype schemas of men and women leaders due to descriptive stereotypes shaped by past experiences and prescriptive stereotype expectations about how men and women ought to behave as leaders. Research has shown that the process through which prototypes can affect the cognition process is complicated. In short, prototypes can affect what stimuli individuals attend to (e.g., individuals have a harder time noticing agentic behavior from women

compared to from men; Scott & Brown, 2006), how they encode information (e.g., perceiving the same behavior as different depending on the gender of the actor; Phillips & Lord, 1982), and how they retrieve schema-consistent information (or fail to retrieve schema-inconsistent information; Hogue & Lord, 2007). Indeed, evidence indicates that agentic leader prototypes (e.g., strength) tend to be endorsed for men and communal prototypes (e.g., sensitivity) tend to be more strongly associated with women (Johnson et al., 2008). Thus, there seem to be not only prototypes for leaders in general, but also prototypes that might differ for male and female leaders. This likely has consequences for perceptions of a female leader's competence, expectations for future behavior, and evaluations of her ability (Hogue & Lord, 2007).

The degree of congruence between followers' expectations for how leaders should act and the leader behaviors displayed by leaders seems to have important ramifications for how they are subsequently evaluated; research has demonstrated that followers strongly favor leaders who act in ways that match their implicit leader prototypes (e.g., Nye & Forsyth, 1991). Regardless of whether a leader's behavior might be considered most appropriate within a specific situation or context, it might be perceived negatively if it is inconsistent with a follower's implicit theory or expectations (Yukl, 2013). Thus, individuals' expectations for leader behaviors are an important consideration in understanding how leaders are evaluated.

Research has identified eight universally held leader prototypes (e.g., attractiveness, intelligence, strength; Offerman et al., 1994; Epitropaki & Martin, 2004). However, these prototypes reflect traits of leaders and are not generally reflective of leader behaviors. In other words, implicit leadership theories like Leader Categorization Theory suggest that individuals develop prototypes of how leaders should *be,* but not necessarily how they should *act.* This subtle difference is especially meaningful in the context of the present research which will seek

to examine how incongruence between followers' leadership expectations and what leaders actually do (i.e., leader behaviors) can impact leader evaluations. Thus, this research will refer to followers' leader behavior expectations rather than prototypes. Conceptually, these expectations operate similarly to follower prototypes but are distinct in that they refer to followers' expectations for leader behaviors rather than traits.

Stereotypes of Leaders

In addition to the role of past experiences in developing followers' expectations for leader behaviors, societal conceptualizations and stereotypes of leadership help inform followers' implicit theories of effective leadership. For a long time, conceptualizations of effective leadership tended to inordinately emphasize the importance of masculine or agentic leader qualities (e.g., competitiveness, assertiveness, forcefulness; Yukl, 2013). This conclusion was evident based on three related lines of research: the think manager–think male paradigm (Schein, 1973); the agency–communion paradigm (Powell & Butterfield, 1979); and the masculinity–femininity paradigm (Shinar, 1975). All three research paradigms came to this conclusion by comparing the similarity of cultural stereotypes of leaders to stereotypes of men and women, albeit by using slightly different methodology. Results from all three approaches were consistently clear: leadership was viewed primarily in masculine terms (Koenig et al., 2011).

As explained by Role Congruity Theory, agentic leader traits and behaviors associated with leaders align closely with the traditional gender role ascribed to men but represent a misalignment with communal qualities (e.g., compassion, helpfulness) to which women were traditionally expected to conform. Societal expectations for how women should behave, coupled with masculine conceptualizations of leadership requirements, resulted in a mismatch in how

capable women were viewed to be as leaders, as suggested by Role Congruity Theory (Eagly & Karau, 2002). As noted previously, while men could also be disadvantaged due to prescriptive bias in how communal behaviors exhibited by men would be received, Role Congruity Theory insinuates this bias against men is less relevant because leadership roles are considered to have primarily masculine requirements. Thus, the agentic role of leaders is a key assumption made by Role Congruity Theory's explanation of why women are consistently rated as being less effective leaders than men.

However, societal views of leadership can change over time (Bass & Bass, 2008). More often than in the past, effective leadership is more recently being viewed as requiring a balanced mix of both task-focused and relations-focused leadership behavior. This is in part evidenced by a meta-analysis of leadership stereotypes that suggests that the masculine construal of leadership has diminished over time (Koenig et al., 2011). While their findings do not rule out that changing gender stereotypes could be driving this change, this is unlikely as other evidence indicates that stereotypes about men and women are still strong today (Eagly et al., 2020). The most likely interpretation is that popular conceptions of leadership are instead changing to become more androgynous (Koenig et al., 2011). Possible reasons for changing conceptions are numerous. Some attribute it to actual changes in the nature of leadership during modern times. For example, social and technological changes have added layers of complexity to organizations, possibly rendering traditional command-and-control influence tactics less effective than democratic, participatory leadership styles (Gergen, 2005; McCauley, 2004; Lipman-Blumen, 2000). The slow, but steady, increase in representation of women in leadership positions could be another contributing factor to changing views of leadership. Evidence indicates that first-hand exposure to women leaders can lead to changes in perceptions of leader roles (Beaman et al., 2009). It is

therefore possible that mere increased exposure to competent women over the past several decades has led to more androgynous views of leadership.

To a certain extent, changing conceptualizations of effective leadership are also reflected in theoretical developments in the leadership literature. For example, recently proposed leadership theories more often emphasize the need for interpersonal-oriented leadership than they did in the past (Avolio et al., 2009; Eagly and Carli, 2007; van Dierendonck, 2011). One increasingly popular model of leadership is servant leadership, which calls for leaders to focus on nurturing followers' personal and professional growth (Greenleaf, 1977; Avolio et al., 2009; Bass & Bass, 2008). Notably, practicing servant leadership requires leaders to engage in high levels of stereotypically feminine leader behaviors (Barbuto & Gifford, 2010). Other recent leadership theories, such as spiritual leadership (Fry, 2003; 2005), authentic leadership (Avolio et al., 2004; Ilies, Morgeson, & Nahrgang, 2005; Shamir & Eilam, 2005), and humble leadership (Schein & Schein, 2018), similarly seem to advocate for leaders to act in communal ways and prioritize utilizing a relations-oriented leadership style.

Leader Behaviors

The proposed research will examine gender differences in leader behavior expectations and display along specific behaviors from Yukl's Hierarchical Taxonomy of Leadership Behaviors (Yukl, 2012). Yukl's Taxonomy represents an integration of ten prominent historical leader behavior taxonomies and includes fifteen leader behaviors across four meta-categories: task-oriented behaviors (i.e., concerned with the accomplishment of task objectives), relations-oriented behaviors (i.e., concerned with managing relationships with and between followers), change-oriented behaviors (i.e., concerned with encouraging and facilitating change), and external behaviors (i.e., boundary-spanning behaviors; Yukl, 2013).

Despite the prominence of Yukl's Taxonomy in the leadership literature, it does not appear to have been applied to the study of leadership and gender. This is likely the case because research on gender differences and bias in leadership more often measures leader traits (e.g., agency and communion) rather than behaviors. Research that does measure behavior almost exclusively models leadership behavior along the two broad factors of leadership extracted from the Ohio State Studies: consideration and initiating structure (Fleishman, 1953). Some have equated this behavioral paradigm to the agency/communion trait paradigm and suggest that men are more likely to exhibit higher levels of task-oriented (i.e., structuring) behaviors while women are more likely to excel in displaying relations-oriented (i.e., consideration) behaviors (Eagly & Karau, 2002). Past meta-analytic evidence did suggest women tend to exhibit slightly higher amounts of interpersonally oriented leader behaviors than men, but no gender differences were found in task-oriented behavior (Eagly & Johnson, 1990). This suggests that stereotypic agentic and communal trait differences might be more prominent than behavioral differences between men and women leaders.

Despite the lack of gender differences in task-oriented behavior and only small differences in relations-oriented behavior found in Eagly and Johnson's (1990) meta-analysis from thirty years ago, continued investigation into gender differences in leader behavior is warranted for two reasons. First, this meta-analysis only considered differences along the two broad meta-categories of leader behavior. Applying Yukl's taxonomy, which takes a more specific approach to modeling leader behavior, will allow for a more fine-grained investigation into gender and leader behavior. Additionally, it is possible that gender differences have changed in the decades that have passed since studies comprising Eagly and Johnson's meta-analysis were conducted. The changing nature of leadership mentioned previously, in combination with

the increased representation of women in leadership, has likely led to women feeling less pressure to conform to masculine leadership stereotypes (or the hiring of women who do not conform to these standards). In other words, leader behavior of women (and men) on average might be evolving, highlighting the need for continued investigation of leader behavior and gender.

The present research will take a narrow focus and consider specific leader behaviors rather than broad meta-categories of leader behavior or a long list of individual behaviors within meta-categories. This narrow scope is for two reasons. First, the focus on specific types of behaviors allows for experimental manipulation while avoiding confounds that may occur with manipulating multiple types of behaviors. The second reason is that this research aims to bridge a gap between how gender researchers and leadership researchers model and measure leadership. That is, gender researchers often measure leadership traits according to the agency/communion or masculinity/femininity paradigms (e.g., agentic traits like assertiveness or confidence; communal traits like compassion or warmth; Bakan, 1966; Eagly, 1987). While these approaches have strengths, they are not useful in measuring leadership at a granular level, nor are they able to measure how leaders act as has been taxonomized in the leadership literature as task and relations-oriented behavior. Many leader behavior taxonomies exist, but Yukl's has been lauded as it combines the factor analytic approach to taxonomy development and the literature integration approach (Yukl, 2012). Yukl's taxonomy contains a list of leader behaviors that fall under the meta-categories of task-oriented behaviors (e.g., Monitoring, Problem Solving), relations-oriented behaviors (e.g., Supporting, Recognizing), and change-oriented behaviors (e.g., External Monitoring, Envisioning Change; Yukl, 2012). The difficulty in bridging the gap between the trait approach and behavior approach lies in the fact that agentic and communal

behaviors are often viewed as similar to task-oriented and relations-oriented behaviors, respectively. However, not all task-oriented behaviors are agentic in nature, and not all relations-oriented are communal. For example, one could argue that Monitoring, a task-oriented leader behavior concerned with checking and regulating the performance of subordinates, is conceptually similar to elements of agency like being assertive and achievement-oriented. Other task-oriented behaviors, like Problem Solving (i.e., concerned with identifying problems and coming up with solutions; Yukl, 2012) seems to be less clearly tied to agentic qualities. Similarly, while a relations-oriented behavior like Supporting (i.e., showing consideration, acceptance, and concern for the needs and feelings of others; Yukl, 2012) is consistent with communal qualities such as being compassionate and considerate, a relations-oriented behavior like Recognizing (i.e., giving praise for effective performance; Yukl, 2012) might be viewed as having less overlap with communal traits. Other behaviors, especially those in the change-oriented meta-category, such as Encouraging Innovation (i.e., promoting innovation, creativity, and flexibility; Yukl, 2012), seem to fall in a neutral space in the agentic/communal behavior continuum.

Accordingly, the present research will focus on three leader behaviors from Yukl's Taxonomy: one relations-oriented behavior that is communal in nature (Supporting), one task-oriented behavior that is agentic in nature (Monitoring Operations and Performance, hereafter referred to simply as Monitoring), and one change-oriented behavior that is neutral in the agency/communion paradigm (Encouraging Innovation). Examples of Supporting behaviors include "showing concern for the needs and feelings of individual members, providing support and encouragement when there is a difficult or stressful task, and expressing confidence members can successfully complete it" (Yukl, p. 461, 2012). Evidence indicates Supporting

behaviors are strongly related to effectiveness perceptions and follower ratings of leader satisfaction (Bass, 1990; Yukl, 1999; Yukl et al., 2019). Supporting behaviors overlap with a key component of a transformational leader behavior called individualized consideration, which is defined as focusing on the development and mentoring of followers and attending to their individual needs (Bass, 1985; Bass & Avolio, 1990). A meta-analysis on gender differences in transformational leader behaviors found that women on average displayed more individualized consideration behaviors (d = .19) across twenty-eight studies (Eagly et al., 2003). One conclusion that can be drawn from this research is that Supporting behaviors might be viewed as more communal or feminine.

Monitoring behavior includes "checking on the progress and quality of work, examining relevant sources of information to determine how well important tasks are being performed, and evaluating the performance of members in a systematic way" (Yukl, p. 460, 2012). Evidence indicates Monitoring behaviors are also related to perceptions of leadership effectiveness and leader satisfaction (Komaki, 1986; Komaki et al., 1989; Kim & Yukl, 1995; Yukl et al., 2019). Monitoring behavior is conceptually very similar to Full Range Leadership Theory's "management by exception–active" behavior (i.e., attending to follower mistakes and progress in meeting objectives) that falls under the umbrella of transactional leader behaviors (Bass, 1985; Bass & Avolio, 1990). In a meta-analysis comparing men and women on transformational and transactional leader behaviors, men were found to display more management by exception–active behaviors than women (d = .12) across twelve studies (Eagly et al., 2003), indicating Monitoring might be perceived as a more agentic or masculine leader behavior.

Encouraging Innovation behaviors involve those that encourage followers to "look at problems from different perspectives, to think outside the box when solving problems, to

experiment with new ideas, and to find ideas and other fields that can be applied to their current problem or task" (Yukl, p. 462, 2012). Engaging in these behaviors helps create a climate of psychological safety and mutual trust. A host of survey studies, field and lab experiments, and case studies have demonstrated that Encouraging Innovation behaviors are linked to evaluations of leadership effectiveness (e.g., Bass & Yammarino, 1991; Elenkov et al., 2005; Howell & Avolio, 1993; Edmondson, 2003). Encouraging Innovation behavior is conceptually related to Full Range Leadership Theory's "intellectual stimulation" behavior (i.e., encouraging creativity, innovation, critical thinking, and problem-solving) that falls under the umbrella of transformational leader behaviors (Bass, 1985; Bass & Avolio, 1990). In a meta-analysis comparing men and women on transformational and transactional leader behaviors, women were found to display slightly more intellectual stimulation behaviors than men (d = .05) across thirty-five studies (Eagly et al., 2003). This difference was tied for the smallest gender difference in behaviors among all of the transformational/transactional behaviors studied, suggesting Encouraging Innovation might be perceived as less strongly masculine or feminine than other leader behaviors.

It was noted that Supporting, Monitoring, and Encouraging Innovation behaviors are all related to evaluations of leadership effectiveness, but how do they compare? One study that examined the validity of behaviors in an earlier version of Yukl's Taxonomy of Leadership Behaviors (this version did not include Encouraging Innovation) found that Monitoring behaviors had a slightly larger correlation with leadership effectiveness than did Supporting behaviors (r = .27 compared to r = .22; Kim & Yukl, 1995), although another found Supporting to have a higher correlation (r = .56 compared to r = .42; Yukl et al., 2019; this study found Encouraging Innovation to have a correlation of r = .54 with effectiveness). Two meta-analyses

of other leader behavior paradigms might also lend some insight into relative differences in effectiveness of Monitoring and Supporting behaviors. The first found that consideration behaviors (i.e., relations-oriented) were more positively related to leader effectiveness than were structuring behaviors (i.e., task-oriented; $\rho = .52$ versus $\rho = .39$) but did not measure behavior at a more specific level than these two meta-categories (Judge et al., 2004). The second compared the relative effectiveness of transformational leader behaviors and transactional leader behaviors (Judge & Piccolo, 2004); while the validity of the transactional behavior Management by Exception–Active (i.e., similar to Monitoring) was reported ($\rho = .15$), the meta-analysis did not report validities for individual behaviors, such as Individualized Consideration (i.e., similar to Supporting), that comprised the meta-category of transformational leadership.

Hypotheses

The present research will test a series of hypotheses to investigate a) whether there are differences in follower expectations for leader behaviors between men and women leaders, and b) how incongruence between follower expectations and leaders' display of communal and agentic leader behaviors might play a differential role in how men and women leaders are evaluated. Supporting and Monitoring behaviors will be used as examples of communal and agentic leader behavior, respectively. Encouraging Innovation will be included as an example of a change-related (i.e., non-gendered) leader behavior.

As noted previously, it has been proposed that the process through which leader prototypes are generated might disadvantage women due to the fact that prototypes might be contaminated by role expectations, which are incongruent with women's stereotypic gender role. Regardless of whether actual behavioral differences exist between men and women, meta-analytic evidence indicates the enduring existence of strong descriptive stereotypes that women

are more communal than men, and that men are more agentic than women (Eagly et al., 2020). Findings also indicate that prescriptive stereotypes of male and female leaders have remained stable over the past four decades (Zehnter et al., 2018). If gender role stereotypes remain salient, it follows then that followers should have higher expectations for role congruent behaviors of leaders. Indeed, one study found that followers expect women to engage in more servant leadership behaviors than men, which are commonly viewed as distinctly feminine (Hogue, 2016). This finding supports the idea that followers might hold differing levels of expectations for communal leader behaviors from men and women leaders, but it remains to be seen if the opposite is true of expectations for agentic behaviors. In line with Social Role Theory, evidence of enduring descriptive and prescriptive gender stereotypes, and research on leader prototypes, I hypothesize the following:

Hypothesis 1a: Followers will have greater expectations for Supporting behaviors from female than male leaders.

Hypothesis 1b: Followers will have greater expectations for Monitoring behaviors from male than female leaders.

Hypothesis 1c: Followers will expect equal levels of Encouraging Innovation behaviors from male and female leaders.

Of additional interest is whether certain follower characteristics predict expectations for gender-congruent behaviors from their leaders. One of the characteristics that is expected to moderate the relationship between leader gender and followers' leader behavior expectations is the gender role orientation of followers. The literature on gender and leadership suggests perceptions and evaluations of leaders are influenced by various follower intrapsychic processes; one of these important factors is the gender role beliefs and attitudes of followers (Ayman &

Korabik, 2010). If followers have greater expectations for role congruent behaviors from men and women leaders due to enduring descriptive and prescriptive gender stereotypes, it follows that this effect should be stronger among individuals who subscribe to more traditional views of gender roles.

> *Hypothesis 2:* Follower gender role orientation will moderate the relationship between leader gender and followers' leader behavior expectations such that followers with a traditional gender role orientation will expect higher levels of role congruent leader behaviors (i.e., Supporting behaviors from women and Monitoring behaviors from men) and lower levels of role incongruent leader behaviors than followers with an egalitarian gender role orientation (See Figures 1 and 2).

Several other exploratory moderators will be considered. One of the most commonly studied moderators of how individuals perceive or evaluate leaders is follower (or rater) gender. While research has demonstrated that men and women hold different leadership schemas (e.g., Schein, 2001) and men and women each think that masculine and feminine leadership, respectively, is more attractive (e.g., Stoker et al., 2012), it is unclear whether or how follower gender would interact with leader gender to predict leader behavior expectations. Thus, it is not hypothesized that follower gender will moderate the relationship between leader gender and leader behavior expectations. An additional exploratory moderator is follower age; however, it is difficult to foresee the effect of age on followers' leader behavior expectations. On one hand, older individuals might have more experience with female managers, causing them to have less gender role-consistent leader behavior expectations. Conversely, older individuals also tend to have a more traditional gender role orientation (Howell & Day, 2000), which might lead them to have greater gender role-consistent leader behavior expectations.

The next set of hypotheses relate to how leader evaluations are influenced by incongruence between follower expectations and leaders' behavior, how this relationship operates under different conditions of expectation fulfilment, and the extent to which these relationships might differ for men and women leaders. Some inferences about how expectation fulfillment might relate to evaluations can be drawn from research on implicit leadership theories and Leader Categorization Theory, as evidence indicates that a higher match between followers' expectations and leaders' behavior leads to more favorable leadership evaluations. For example, followers who endorse warm and friendly leader prototypes rate socioemotional-oriented leaders more favorably than task-oriented leaders, and followers who endorse dominant and controlling leader prototypes rate task-oriented leaders more favorably than socioemotional-oriented leaders (Nye & Forsyth, 1991). To this degree, we have some understanding of how leader behavior expectations lead to leadership evaluations.

However, less is known about how leadership evaluations are impacted by varying levels of expectation–behavior incongruence. For example, how are leaders evaluated when expectations for certain leader behaviors are unmet? What about when they are exceeded? The met expectations hypothesis (Porter & Steers, 1973) suggests that in situations in which individuals' expectations for various work-related experiences are unmet, they react negatively. Conversely, situations in which low expectations are exceeded can also sometimes lead to negative reactions, suggesting a curvilinear relationship might exist between expectation congruence and evaluations (Irving & Montes, 2009). Leader behavior theorists have similarly contended that higher levels of effective leader behaviors such as initiating structure and consideration do not always lead to higher evaluations. Rather, some optimal level might exist in the eyes of followers (Fleishman, 1995). Applying the met expectations hypothesis to this

context might suggest that an optimal level of various leader behaviors exists and is dictated by followers' leader behavior expectations.

If incongruence between followers' expectations and a leader's behavior results in negative leadership evaluations, and if followers have different expectations of men and women leaders as suggested in Hypothesis 1, men or women might experience lower evaluations for exhibiting the same leader behavior. Indeed, prior research has demonstrated how prescriptive gender stereotypes can lead to evaluative bias against women leaders (e.g., Rudman, 1998; Rudman & Glick, 2001). Notably, however, most of this research tends to conflate gender-typed occupations or tasks with actual behaviors; that is, they consider how women (men) are evaluated when they succeed in a male-typed (female-typed) job or task but do not measure the display of actual behaviors. This suggests a need for more research that measures, or controls and manipulates, specific leader behaviors.

Another shortcoming of research on counter-role leadership styles is that it has almost exclusively focused on evaluative bias against women, which is the primary focus of Role Congruity Theory. However, the Lack of Fit model asserts that men should also be subject to evaluative biases for displaying counter-role behavior. Indeed, there is a general tendency for deviations from injunctive norms to elicit disapproval (Cialdini & Trost, 1998), suggesting men should similarly be evaluated negatively for violating gender role norms as leaders. However, less is known about whether male leaders face similar penalties to the same degree that female leaders do when they display counter-role leader behavior. Results from one study indicate that men do receive negative reactions for succeeding in counter-role communal-typed tasks (Heilman & Wallen, 2010). Another found that male leaders who ask for help are viewed as less competent than those who do not ask for help; this penalty was unique to male leaders and was

not observed for female leaders (Rosette et al., 2015). However, further investigation is required to determine if men are subject to comparable consequences as women for violating injunctive norms as leaders.

To investigate how deviations from followers' leader behavior expectations might produce negative evaluations of men and women leaders, two conditions of follower expectation-leader behavior incongruence will be considered: 1) when follower expectations for a specific leader behavior are exceeded (i.e., expectations are lower than the leader's actual behavior), and 2) when follower expectations for a specific leader behavior are unmet (i.e., expectations are higher than the leader's actual behavior). The majority of research on gender role-incongruent leadership has considered the first condition. In other words, when female leaders display counter-role leadership behaviors or styles or succeed in counter-role tasks or industries, are they evaluated negatively compared to men? Findings have consistently shown that female leaders who display counter-role behaviors receive backlash and other negative reactions (e.g., Rudman, 1998; Rudman & Glick, 1999, 2001; Johnson et al., 2008; Wang et al., 2013). Less research has considered consequences for male leaders who display counter-role behavior, but some evidence has found a similar backlash effect for men (e.g., Heilman & Wallen, 2010). Accordingly, it is hypothesized that when leaders of both genders display leader behaviors that are inconsistent with gendered follower expectations for leader behaviors, they will receive more negative reactions compared to leaders of the other gender.

Hypothesis 3a: When follower expectations for Supporting are low but leaders exhibit high levels of Supporting behavior, men will be rated as less effective than women (see Figures 3 and 4).

Hypothesis 3b: When follower expectations for Monitoring are low but leaders exhibit high levels of Monitoring behavior, women will be rated as less effective than men (see Figures 3 and 4).

Hypothesis 3c: When follower expectations for Encouraging Innovation are low but leaders exhibit high levels of Encouraging Innovation behavior, men and women will not differ in effectiveness ratings.

Considerably less research has examined the consequences for leader evaluations and reactions when leaders fail to display gender role consistent behavior compared to when they display counter-role behavior. However, inferences can be drawn from related research on OCBs. One study showed that when women–who were expected to engage in more helping OCBs than men–failed to display these role consistent OCBs, they were evaluated more negatively than men who also did not display helping OCBs (Heilman & Chen, 2005). Role Congruity Theory and the Lack of Fit model would suggest that similarly to when leaders display counter-role behaviors, negative reactions will occur when leaders fail to engage in role-congruent behaviors. Accordingly, the following is hypothesized to occur when followers' leader behavior expectations are unmet:

Hypothesis 3d: When follower expectations for Supporting are high but leaders exhibit low levels of Supporting behavior, women will be rated as less effective than men (see Figures 4 and 7-8).

Hypothesis 3e: When follower expectations for Monitoring are high but leaders exhibit low levels of Monitoring behavior, men will be rated as less effective than women (see Figures 6-8).

Hypothesis 3f: When follower expectations for Encouraging Innovation are high but leaders exhibit low levels of Encouraging Innovation behavior, men and women will not differ in effectiveness ratings.

Method

To test these hypotheses, I conducted an experiment in which the gender of the leader, level of Supporting behavior, level of Monitoring behavior, and level of Encouraging Innovation behavior were manipulated across experimental conditions. Therefore, the study is a 2 x 2 x 2 x 2 between-subjects experimental design. I chose to utilize an experimental approach rather than a field design for three reasons. First, it provided greater control and avoided potential confounds with leader gender–such as industry, job type, or job level–that might be problematic in a field correlational study. Second, an experimental approach allowed for clearer insight into causal connections, the lack of which has plagued prior literature. Finally, this approach gave me an opportunity to learn and apply polynomial regression and response surface methodology techniques. While polynomial regression could be used in a field survey, the ability to experimentally manipulate leader behaviors likely created greater variability in the behavior perception variables than would otherwise be found in a field study.

Sample

Based on a priori power analysis using G*Power 3.1 (Faul et al., 2007), at least 474 participants were needed to achieve 80% power to detect medium sized effects. In order to maximize variability in several key variables, I sought to collect a sample of at least 550 data points. Participants were recruited from Qualtrics Panels, an online data collection platform. In order to qualify for the study, participants answered a series of screening questions (see Appendix C) to ensure they were at least 18 years old and have experience working in a full-time

job (i.e., 35+ hours per week in the last six months) in which they regularly interacted with an immediate supervisor or manager. I also sampled to ensure gender and age representativeness (i.e., 50% men and 50% women; 33% in 18-34, 35-54, and 55+ age groups). To maximize data quality, I included several open-ended questions to screen out bots and nonsensical responding. Those who qualified for the study, completed the procedure and survey, and passed the data quality checks were compensated directly from Qualtrics in the amount and format advertised (e.g., cash, gift cards, or other rewards).

The final sample consisted of 564 respondents. Just over half were men (50.4%) and the mean age was 44.7 years old (*SD* = 14.5; range: 18-81). In response to the question, "Out of your entire working career in which you have had a manager or supervisor, approximately what percent of the time has your manager or supervisor been a woman?" the mean percentage was 48.17 (SD = 25.26). About 3 percent of respondents reported never having worked for a female manager and another 3 percent reported never having worked for a male manager, but most (66.7%) reported having a female manager for 25 to 75 percent of their career.

Procedure

Participants who qualified based on the screening questions and agreed to the informed consent were randomly assigned to read a series of vignettes (see Appendix D) describing a new male or female supervisor who displays high or low levels of Supporting, Monitoring, and Encouraging Innovation behaviors. After reading the first vignette, which asked them to imagine they are part of a work group who will soon be assigned a new supervisor, described the situation, and briefly introduced the new supervisor, participants were presented with an attention check item and then answered a series of questions that measured their expectations for Supporting, Monitoring, and Encouraging Innovation behaviors from the supervisor. Next, a

series of three consecutive vignettes described the fictional supervisor's actions through their first few weeks on the job vis-à-vis the extent to which they have displayed Supporting, Monitoring, and Encouraging Innovation behaviors. Following each vignette, participants were required to correctly answer a simple comprehension check to ensure they properly read and comprehended the vignettes. After reading through the vignettes, participants were then asked to rate their perceptions of the supervisor's behavior, evaluate the effectiveness of the supervisor's leadership, and provide a qualitative description of the leader's actions. Finally, participants' gender role orientation was measured.

Manipulation

Four variables were manipulated within the supervisor vignettes (see Appendix D): the gender of the fictional supervisor (i.e., male or female) and the level (i.e., high or low) of Supporting, Monitoring, and Encouraging Innovation behaviors displayed by the supervisor. Thus, this experiment is a 2 x 2 x 2 x 2 design. Participants were randomly assigned to one of the resulting sixteen conditions.

The majority of the first vignette remained constant across all conditions and introduced the scenario. The vignette asked participants to imagine that they were in the scenario presented as to enhance participants' engagement with the story. The story depicted a situation in which the participant, who works for a professional services firm, is about to be assigned a new supervisor. This industry was selected because there is rough equivalence in gender representativeness (U.S., 2020) and most jobs in this sector have traditional organizational leadership hierarchies in which individual contributors would have a direct manager to whom they report. At the end of the vignette, participants learned the name of their new supervisor and were shown a headshot of them. The gender of the supervisor was manipulated in the written description of the supervisor's

name and pronouns (i.e., Ken/Kelly, he/her, etc.). Participants also received a brief summary of the supervisor's education and work experience; this summary remained constant across conditions. In addition, the supervisor vignettes contained one of two headshots of the supervisor (i.e., male or female; see Appendix D) depending on the condition.

After presenting this background information about the situation and supervisor, the experiment first assesses participants' expectations for their new supervisor's leadership behavior. The survey then asked participants to imagine several weeks had passed and that their new supervisor has now spent several weeks in their current role. Participants were next presented with a second vignette that described their supervisor's behavior through their first several weeks on the job. Supporting, Monitoring, and Encouraging Innovation behaviors were each manipulated into conditions of high and low levels through written descriptions of the supervisor's actions (see Appendix D).

Prior to conducting the main study, I conducted a small pilot study with a sample of undergraduate students (n = 50) at a large public university in the Midwestern United States. The primary purpose of the pilot study was to select headshot photos to use in the gender manipulation. A set of ten stock photos depicting professional headshots of businesspeople (five male and five female) were presented to participants, who were asked to rate their perceptions of each person's professionalism and attractiveness and provide estimates of their age. Two headshots that were selected for use in the main study–one of a man and one of a woman–received equivalent ratings in estimated age, perceived attractiveness, and perceived professionalism (see Table 1). The secondary purpose of the pilot study was to ensure sufficient variance would be observed in leader behavior expectations, leader behavior perceptions, and

leader effectiveness perceptions. Pilot results indicated that sufficient variance would be observed in each of these variables.

Measures

Leader Behavior Expectations

After reading the initial vignette that described the work group and briefly introduced the new leader, respondents' expectations for Supporting, Monitoring, and Encouraging Innovation leader behavior from the supervisor were measured using three subscales adapted from Yukl's Managerial Practices Survey (MPS; 2012). Each MPS subscale contains four items that ask followers to rate the extent to which their manager displays each behavior (see Appendix E). For the purposes of this study, the item stem was adapted to assess respondents' expectations that the leader will display these behaviors in the future (i.e., "I expect that [Ken/Kelly] will…"). A sample item from the Supporting subscale is "…show concern for the needs and feelings of individual members of the work unit." A sample item from the Monitoring subscale is "…check on the progress and quality of the work." A sample item from the Encouraging Innovation subscale is "…talk about the importance of innovation and flexibility for the success of the unit." Responses were rated on a five-point scale (1 = Not at all, 5 = To a very great extent). Scale reliabilities were adequate for all three subscales (Supporting $\alpha = .86$; Monitoring $\alpha = .81$; Encouraging Innovation $\alpha = .85$). CFA indicated expectations for each behavior were distinct rather than representing a single general leadership expectations factor[2].

[2] A three factor CFA model with Supporting, Monitoring, and Encouraging Innovation expectations demonstrated significantly better fit ($\chi^2 = 174.97$, SRMR = .04, RMSEA = .07, CFI = .96, TLI = .95) compared to a single factor model ($\chi^2 = 651.19$, SRMR = .07, RMSEA = .14, CFI = .83, TLI = .79).

Leader Behavior Perceptions

After reading the second vignette that describes how the supervisor has handled leading the team so far, respondents' perceptions of the supervisor's behavior were measured using the Monitoring, Supporting, and Encouraging Innovation subscales from the MPS (Yukl, 2012). Each subscale contains four items and were rated on a five-point scale (1 = Not at all, 5 = To a very great extent; see Appendix E). The stem for all items was "Based on the information provided, I think Ken/Kelly…." A sample item from the Supporting subscale is "…provides support and encouragement when there is a difficult or stressful task." A sample item from the Monitoring subscale is "…evaluates how well important tasks or projects are being performed." A sample item from the Encouraging Innovation subscale is "…encourages innovative thinking and creative solutions to problems." All three subscales were found to possess high levels of reliability (Supporting $\alpha = .94$; Monitoring $\alpha = .96$; Encouraging Innovation $\alpha = .94$).

Perceived Leader Effectiveness

Perceptions of the supervisor's leadership effectiveness were measured with a three-item scale previously used in laboratory experiments to assess the perceived effectiveness of leaders presented in vignettes (Johnson et al., 2008; see Appendix E). The three items are: "[Ken/Kelly] will be effective"; "[Ken/Kelly] will succeed in this role"; and "[Ken/Kelly] will improve performance at this company." Responses were rated on a 7-point Likert agreement scale (1 = strongly disagree, 7 = strongly agree). The scale possessed high reliability ($\alpha = .94$).

Gender Role Orientation

Respondent gender role orientation was assessed using the Gender Role Stereotypes Scale (Mills et al., 2012). The scale asks respondents to indicate the extent to which they believe each task should be done by the man, should be done by the woman, or the man and woman

share the responsibility equally, when there is a relationship between a man and a woman. The scale contains eight items–four items are traditionally masculine tasks (e.g., mow the lawn) and four items are traditionally feminine tasks (e.g., prepare meals). Responses were rated on a 1-5 scale (1 = should always be done by the man, 2 = should usually be done by the man, 3 = equal responsibility, 4 = should usually be done by the woman, 5 = should always be done by the woman). Responses to the feminine tasks were reverse coded such that higher scale values represent traditional gender role orientations and lower scale values represent egalitarian gender role orientations. Adequate scale reliability was observed ($\alpha = .76$), in line with reliabilities reported by the authors of the scale ($\alpha = .75$ and $\alpha = .78$; Mills et al., 2012).

Exploratory Measures

Leader likability was also measured with a three-item scale previously used in laboratory experiments to assess the likability of leaders presented in vignettes (Johnson et al., 2008; see Appendix C). Studies on gender and leadership often use two different measures of leadership evaluations as some research indicates communal behaviors are often more closely tied to outcomes such as liking or satisfaction, while agentic behaviors are often more closely tied to outcomes such as respect or competence (Wojciszke et al., 2009). This scale has three items: "[Ken/Kelly] will be liked by his/her employees"; "[Ken/Kelly] seems likeable"; and "[Ken/Kelly]'s employees will like working for him/her." Responses were rated on a 7-point Likert agreement scale (1 = strongly disagree, 7 = strongly agree). This scale was also found to possess high reliability ($\alpha = .96$).

Additionally, General Leadership Impression (GLI) was measured using the GLI Measure (Cronshaw & Lord, 1987). The GLI Measure contains five items that assess a respondent's perceptions of how "leader-like" they perceive someone to be. This scale was used

twice in the survey; first, respondents' expectations of the supervisor vis-à-vis the GLI items were measured after reading the introductory vignette (e.g., "To what degree do you expect Ken/Kelly to fit your image of a leader?"), and second, respondent's evaluations of the supervisor's leadership were assessed with the GLI measure after reading all of the vignettes (e.g., "To what degree did Ken/Kelly fit your image of a leader?"). Responses were rated on a 5-point Likert scale. High reliability was observed for both the GLI expectations measure ($\alpha = .84$) and the GLI evaluations measure[3] ($\alpha = .94$).

Demographics

The survey also asked respondents to report their gender, age, employment status, and experience working for female managers (see Appendix E).

Results

Descriptive statistics and bivariate correlations for all variables of interest are presented in Table 2. Of note, across both supervisor gender conditions, follower expectations for all three behaviors were high; expectations for Monitoring behaviors were the highest ($M = 4.13$, $SD = .67$) followed by expectations for Supporting behaviors ($M = 3.99$, $SD = .73$) and Encouraging Innovation behaviors ($M = 3.98$, $SD = .74$). Additionally, perceptions of all three behaviors were significantly positively related to evaluations of effectiveness, liking, and GLI.

Results suggest the experimental manipulations of Supporting, Monitoring, and Encouraging Innovation behaviors, respectively, were all successful. In separate regression models with predicting perceptions of each behavior, experimental condition (i.e., high versus low levels of behavior) for each respective behavior predicted perceptions of: Supporting

[3] CFA was conducted to determine if effectiveness, liking, and GLI evaluations were distinct factors. Compared to a single factor model ($\chi^2 = 1452.08$, SRMR = .06, RMSEA = .24, CFI = .82, TLI = .77), a three factor model with effectiveness, liking, and GLI evaluations demonstrated significantly improved fit ($\chi^2 = 254.72$, SRMR = .03, RMSEA = .10, CFI = .97, TLI = .96).

behavior, $F(1,562) = 1214.66$, $MSE = .62$, $p < .001$, such that participants in the "high" Supporting condition ($M = 4.15$, $SD = .80$) perceived higher levels of Supporting behavior than did participants in the "low" condition ($M = 1.84$, $SD = .77$); Monitoring behavior $F(1,562) = 969.68$, $MSE = .74$, $p < .001$, such that participants in the "high" Monitoring condition ($M = 4.20$, $SD = .69$) reported higher levels of Monitoring behavior than did participants in the "low" condition ($M = 1.95$, $SD = 1.00$); and Encouraging Innovation behavior $F(1,562) = 997.82$, $MSE = .74$, $p < .001$, such that participants in the "high" Encouraging Innovation condition ($M = 4.07$, $SD = .76$) perceived higher levels of Encouraging Innovation behavior than did participants in the "low" condition ($M = 1.77$, $SD = .96$). A MANOVA model suggests Supporting condition was also a significant predictor of Monitoring perceptions, $F(1,556) = 16.88$, $p < .001$, partial $\eta^2 = .03$, and of Encouraging Innovation perceptions, $F(1,556) = 27.07$, $p < .001$, partial $\eta^2 = .05$. However, these effect sizes are much smaller than the effect of Supporting condition on Supporting perceptions (partial $\eta^2 = .71$). Similarly, Monitoring condition was also a significant predictor of Supporting perceptions, $F(1,556) = 52.34$, $p < .001$, partial $\eta^2 = .09$, and of Encouraging Innovation perceptions, $F(1,556) = 25.96$, $p < .001$, partial $\eta^2 = .05$. These effect sizes are again much smaller than the effect of Monitoring condition on Monitoring perceptions (partial $\eta^2 = .65$). Finally, Encouraging Innovation condition was also a significant predictor of Monitoring perceptions, $F(1,556) = 5.36$, $p = .02$, partial $\eta^2 = .01$, and of Supporting perceptions, $F(1,556) = 20.27$, $p < .001$, partial $\eta^2 = .04$. Once again, however, these effect sizes are much smaller than the effect of Encouraging Innovation condition on Encouraging Innovation perceptions (partial $\eta^2 = .66$).

Agreement descriptive statistics were computed to assess the rate of discrepancies between followers' behavior expectations and perceptions ratings (Table 3). According to

Shanock et al. (2010), before conducting polynomial regression congruence analyses it is important to first determine if, how many, and in what direction discrepancies exist between predictor variables. If few respondents report discrepant values, there is limited practical value in conducting polynomial regression congruence analyses. Because leader behavior levels were manipulated in the present study, sufficient variance was expected in the behavior perception variables. However, prior to data collection it was unknown how much variance would be observed in the behavior expectations variables. Following the procedure outlined by Shanock et al. (2010) and used by others, predictor scores were first standardized. Participants who then reported expectation or perception levels that were one-half of one standard deviation above or below the other predictor were considered to have discrepant values. Values were considered to be in agreement if expectation and perception levels were within one-half of one standard deviation of each other. As can be seen in Table 3, well over half of the sample had discrepant expectations and perceptions levels for all three behaviors. These statistics support the conclusion that investigating the effect of followers' behavior expectations-perceptions incongruence on supervisor evaluations makes sense from a practical perspective.

Hypothesis Tests

Hypotheses 1a and 1b predicted that the level of Supporting behavior and Monitoring behavior expected by followers would be greater for female and male leaders, respectively, and Hypothesis 1c predicted there would be no difference in the level of Encouraging Innovation behaviors expected by follower gender. To assess H1a, a simple regression model was run with follower expectations for Supporting behavior as the dependent variable and supervisor gender (dummy-coded, male = 1, female = 2) as the predictor to determine if follower expectations differed as a function of supervisor gender. This model was significant, $F(1,562) = 4.04$, $MSE =$

.53, β = .08, p = .045, R^2 = .01, suggesting that follower expectations for Supporting behavior from the female supervisor (M = 4.05, SD = .70) were greater than expectations for Supporting behavior from the male supervisor (M = 3.93, SD = .76) and supporting H1a. Because Supporting expectations were significantly positively associated with experience with a female supervisor, this analysis was repeated while controlling for this factor. After controlling for experience with a female supervisor, the gender difference found in the original model became marginally significant, β = .08, t = 1.82, p = .07. To assess H2b and H2c, similar regression models were run to the first but with follower expectations for Monitoring behavior and expectations for Encouraging Innovation behavior, respectively, as the outcome variables. The model predicting follower expectations for Monitoring behavior was not significant, F(1,562) = .95, MSE = .45, β = .04, p = .33, R^2 = .002, suggesting expectations for Monitoring behavior did not differ significantly from the male supervisor (M = 4.10, SD = .66) compared to the female supervisor (M = 4.16, SD = .68). As such, H1b was not supported. Lastly, the model predicting follower expectations for Encouraging Innovation behavior was also not significant, F(1,562) = .92, MSE = .55, β = .04, p = .34, R^2 = .002. This suggests expectations for Encouraging Innovation behavior did not differ significantly between the male supervisor (M = 3.95, SD = .73) and the female supervisor (M = 4.01, SD = .75), supporting H1c. Because Encouraging Innovation expectations were also significantly positively associated with experience with a female supervisor, this analysis was also repeated while controlling for this factor. After controlling for experience with a female supervisor, respondents still did not have different levels of Encouraging Innovation expectations as a function of supervisor gender, β = .03, t = .77, p = .44. Thus, H1a and H1c were supported but H1b was not.

H2 predicted that the relationship between supervisor gender and follower expectations would be moderated by follower gender role orientation such that followers with a traditional gender role orientation (compared to an egalitarian orientation) will expect higher levels of gender role congruent leader behaviors (i.e., Supporting behaviors from women and Monitoring behaviors from men) and lower levels of gender role incongruent leader behaviors from supervisors. To test this, two separate moderated regression models were run (i.e., one each with Supporting expectations and Monitoring expectations as the outcome variables). In each model, supervisor gender and follower gender role orientation were entered as predictors in the first step, and their interaction was entered in the second step. Gender role orientation was grand mean centered and supervisor gender was effects coded (male = 1, female = -1) prior to being entered into each model. The overall model predicting expectations for Supporting behavior (Table 4) was significant, $F(3,560) = 4.83$, $MSE = .52$, $p = .002$, but the interaction between gender role orientation and supervisor gender was not significant ($b = .02$, $t = .28$, $p = .78$). Similarly, the overall model predicting expectations for Monitoring behavior (Table 5) was also significant, $F(3,560) = 4.58$, $MSE = .44$, $p = .004$, but the interaction between gender role orientation and supervisor gender was again not a significant predictor ($b = -.01$, $t = -.16$, $p = .87$). Thus, H2 was not supported.

Hypotheses 3a, 3b, 3d, and 3e predicted that various conditions of incongruence between followers' expectations for Supporting and Monitoring behavior and the leader's behavior (i.e., unmet or exceeded expectations) will have different effects on the evaluations of male and female leaders. Specifically, Hypotheses 3a and 3b predicted that when expectations for counter role behaviors (i.e., Supporting behaviors for men and Monitoring behaviors for women) are exceeded, effectiveness evaluations will be lower compared to their counterparts for whom these

behaviors are role consistent. Conversely, Hypotheses 3d and 3e predicted that when expectations for leader behaviors are unmet, leaders for whom these behaviors are role consistent will have lower evaluations compared to their counterparts for whom these behaviors are counter role. Hypotheses 3c and 3f predicted that incongruence between followers' expectations for Encouraging Innovation behaviors will not have different effects on the evaluations of male and female leaders.

To test these hypotheses, a series of moderated polynomial regression models were run. Utilizing polynomial regression and response surface methodology to test congruence hypotheses instead of linear regression avoids problems associated with difference scores; this approach was first recommended by Edwards (1994; 2001) and has since been used in other areas of organizational research when testing the effect of fit or congruence between predictors on an outcome of interest (e.g., Wiegand, Drasgow, & Rounds, 2020; Humberg, Nestler, & Back, 2019). Prior to analysis, all continuous predictors were centered around the midpoint of their respective scales (i.e., 3 was subtracted from each score as behavior expectations and perceptions variables were all rated on a 5-point Likert scale). Centering is helpful for interpretation purposes and reduces the potential for multicollinearity (Aiken & West, 1991); grand mean centering is recommended for linear regression, but for polynomial regression Edwards (1994) recommends centering predictors around the midpoint of each scale.

Three polynomial regression models were run to test the effects of Supporting, Monitoring, and Encouraging Innovation follower expectation–behavior perception congruence, respectively, on evaluations of effectiveness. Each model was first built by adding the following predictor terms: follower expectations for Supporting, Monitoring, or Encouraging Innovation behavior (X), perceptions of that behavior (Y), the behavior expectations-perceptions interaction

(XY), and the squared terms for behavior expectations (X^2) and behavior perceptions (Y^2). Next, supervisor gender (dummy coded) was added to the model as a moderator variable (W). Following guidance from Aiken and West (1991) and Edwards (2002), five additional terms were then added to the model in an additional step to test for moderation: interactions between supervisor gender and 1) follower expectations (WX), 2) behavior perceptions (WY), 3) the follower expectations squared term (WX^2), 4) the behavior perceptions squared term (WY^2), and 5) the behavior expectation-perceptions interaction term (WXY). This set of five interaction terms collectively represents the moderating effect of supervisor gender. According to Edwards, if the model containing the five moderator interaction terms explains a statistically and practically significant amount of variance in the outcome over and above the model without the interaction terms (ΔR^2), then it is appropriate to conclude the existence of a moderator variable and conduct follow up analyses (Edwards, 2002; Edwards & Parry, 1993).

To test Hypotheses 3a and 3d (i.e., the effect of Supporting behavior expectations-perceptions congruence on effectiveness), a polynomial regression model was first run with the five terms described above (i.e., X, Y, XY, X^2, and Y^2). This model was significant, $F(5,559) = 57.26$, $MSE = 1.87$, $p < .001$. Because the overall model was significant, surface tests were then conducted, and interpretation was aided by viewing the response surface graph for this model (Figure 5). The four surface test values are denoted as a_1 (the slope of the line of perfect agreement as related to effectiveness), a_2 (curvature along the line of perfect agreement as related to effectiveness), a_3 (the slope of the line of incongruence as related to effectiveness), and a_4 (curvature of the line of incongruence as related to effectiveness). As can be seen in Table 6, a_1 and a_3 were significant but a_2 and a_4 were not. This means that the slopes along the line of perfect agreement (X = Y) and the line of incongruence (X = -Y) were both different from zero but that

neither slope had significant curvature to be considered nonlinear. As can be seen from Figure 5, effectiveness ratings were highest when supervisors exhibited high levels of Supporting behavior regardless of expectations levels. When supervisors exhibited low levels of Supporting behavior, effectiveness evaluations decreased as expectations increased. Across both genders of supervisors, exceeded expectations for Supporting behavior (i.e., low expectations and high levels of behavior) received far higher evaluations than did unmet expectations for Supporting behavior (i.e., high expectations and low levels of behavior).

To test for the moderating influence of supervisor gender and evaluate Hypotheses 3a and 3d, supervisor gender was first added to the model. Next, the five supervisor gender interaction variables (i.e., WX, WY, WXY, WX^2, and WY^2) were added as an additional step to the previous model. The resulting moderated polynomial regression model did not explain significant variance over and above the previous model, $\Delta R^2 = .008$, $p = .24$, indicating that congruence between Supporting behavior expectations and perceptions did not have a different effect on effectiveness evaluations for male and female supervisors. Thus, Hypotheses 3a and 3d were not supported.

To test Hypotheses 3b and 3e (i.e., the effect of Monitoring behavior expectations-perceptions congruence on effectiveness), a polynomial regression model was first run without the supervisor gender moderator terms. This model was significant, $F(5,558) = 81.39$, $MSE = 1.64$, $p < .001$. As can be seen in Table 7, a_3 was significant but a_1, a_2 and a_4 were not. This means that the slope along the line of perfect agreement (X = Y) was not different from zero (i.e., Monitoring behavior expectations and perceptions did not have an additive effect) but the line of incongruence (X = -Y) was different from zero. Neither slope had significant curvature to be considered nonlinear. As can be seen from Figure 6, effectiveness ratings were highest when

followers had low expectations for Monitoring behavior and supervisors exhibited high levels of Monitoring behavior. When Monitoring behavior perceptions were high, evaluations decreased slightly as expectation levels increased. When supervisors exhibited low levels of Monitoring behavior, effectiveness evaluations also decreased as expectations increased. Across both genders of supervisors, exceeded expectations for Monitoring behavior (i.e., low expectations and high levels of behavior) received far higher evaluations than did unmet expectations for Monitoring behavior (i.e., high expectations and low levels of behavior).

To test for the moderating influence of supervisor gender and evaluate Hypotheses 3b and 3e, supervisor gender was first added to the model. Next, the five supervisor gender interaction variables were added as a second step to the first model. The moderated polynomial regression model did not explain significant variance over and above the first model, $\Delta R^2 = .002$, $p = .79$, indicating that congruence between Monitoring behavior expectations and perceptions did not have a different effect on effectiveness evaluations for male and female supervisors. As a result, Hypotheses 3b and 3e were not supported.

To test Hypotheses 3c and 3f (i.e., the effect of Encouraging Innovation behavior expectations-perceptions congruence on effectiveness), a polynomial regression model was first run without the supervisor gender moderator terms. This model was significant, $F(5,558) = 50.71$, $MSE = 1.95$, $p < .001$. As can be seen in Table 8, a_1 and a_3 were significant but a_2 and a_4 were not. This means that the slopes along the line of perfect agreement (X = Y) and the line of incongruence (X = -Y) were both different from zero but that neither slope had significant curvature to be considered nonlinear. As can be seen from Figure 7, effectiveness ratings were highest when supervisors exhibited high levels of Encouraging Innovation behavior, for the most part regardless of expectations levels. When supervisors exhibited low levels of Encouraging

Innovation behavior, effectiveness evaluations decreased as expectations increased. Across both genders of supervisors, exceeded expectations for Encouraging Innovation behavior (i.e., low expectations and high levels of behavior) received far higher evaluations than did unmet expectations for Encouraging Innovation behavior (i.e., high expectations and low levels of behavior).

To test for the moderating influence of supervisor gender and evaluate Hypotheses 3c and 3f, supervisor gender was first added to the model. Next, the five supervisor gender interaction variables were added as a second step to the prior model. The amount of variance explained by this moderated polynomial regression model over and above the previous model was significant, $\Delta R^2 = .014$, $p = .04$, indicating that congruence between Encouraging Innovation behavior expectations and perceptions did have some sort of different effect on effectiveness evaluations for male and female supervisors. According to Edwards, if the addition of the five interaction terms results in an increased R^2 that is both statistically and practically significant, then it is appropriate to conclude the existence of a moderation effect and conduct appropriate follow up analyses[4] (Edwards, 2002; Edwards & Parry, 1993). To follow up this significant interaction, separate simple surface graphs were plotted for the male and female supervisor conditions (Figures 8 and 9). As can be seen from the graphs, effectiveness ratings were highest for both male and female supervisors when Encouraging Innovation behavior was high. However, for male supervisors, effectiveness was similarly high regardless of expectations level while for female supervisors, effectiveness was highest when expectations were low or high, but lower when expectations were near the midpoint of the scale. Additionally, in the male supervisor

[4] While the set of interaction terms explained only an additional 1.4% of variance, an amount that is of questionable practical significance, I still conducted follow up analyses if for nothing else than developmental experience.

simple slopes, when expectations were low, effectiveness increased drastically as Encouraging Innovation behavior increased. In the female supervisor condition, however, when expectations were low, effectiveness ratings remained mostly stable even as Encouraging Innovation behavior increased. When followers' low expectations were exceeded by high levels of Encouraging Innovation behavior, effectiveness did not appear to differ between the male and female supervisor conditions. Likewise, when followers' high expectations are unmet (i.e., low levels of behavior), effectiveness did not appear to differ between the male and female supervisor conditions. Thus, while the significant interactions do provide evidence moderation and differences can be seen when comparing the simple surfaces, none of the differences related to the hypothesized relationships between behavior expectations and perceptions incongruence on effectiveness.

Exploratory Analyses

Several additional exploratory analyses were also run. First, MANOVA was conducted to determine if any behavior perceptions or leadership evaluation outcomes differed between the male and female supervisor condition. Results indicated no differences by supervisor gender in perceptions of Monitoring or Encouraging Innovation behavior or in evaluations of effectiveness or GLI. However, participants perceived more Supporting behavior from the female ($M = 3.14$, $SD = 1.35$) than the male condition ($M = 2.87$, $SD = 1.43$), $F(1,562) = 5.32$, $MSE = 1.94$, $p = .02$. This difference was true in the low Supporting conditions but not in the high Supporting conditions. Participants also rated the female supervisor condition as more likeable ($M = 4.59$, $SD = 1.78$) than the male condition ($M = 4.25$, $SD = 1.84$), $F(1,562) = 5.22$, $MSE = 3.26$, $p = .02$.

GLI expectations were considered as an alternate outcome variable to determine if followers' GLI expectations differed between male and female supervisors. Notably, GLI

expectations were moderately strongly related to expectations for Supporting, Monitoring, and Encouraging Innovation behavior (r = .61, .57, and .62, respectively). To examine gender differences in GLI expectations, a simple regression model was run with follower expectations for the GLI scale as the dependent variable and supervisor gender (dummy-coded, male = 1, female = 2) as the predictor to determine if follower expectations differed as a function of supervisor gender. This model was significant, $F(1,562) = 7.19$, $MSE = .40$, $p < .01$, $b = .14$, $R^2 = .01$ suggesting that follower expectations for GLI were greater from the female supervisor ($M = 4.01$, $SD = .64$) than from the male supervisor ($M = 3.87$, $SD = .62$).

It was noted earlier that rater gender would be examined as a moderator of the relationship between leader gender and leader behavior expectations. To do this, moderated regression models were run to consider the interaction between leader gender and rater gender in predicting follower expectations for Supporting behavior, Monitoring behavior, and Encouraging Innovation behavior, as well as GLI expectations. Both supervisor and rater gender were dummy coded (male = 1, female = 2) prior to analyses. In the model predicting expectations for Supporting behavior, neither rater gender ($b = -.09$, $t = -1.00$, $p = .32$) nor the rater gender-supervisor gender interaction ($b = -.01$, $t = -.08$, $p = .93$) were significant. Likewise, in the model predicting expectations for Monitoring behavior, neither rater gender ($b = -.10$, $t = -1.23$, $p = .22$) nor the rater gender-supervisor gender interaction ($b = -.03$, $t = -.31$, $p = .76$) were significant. In the model predicting expectations for Encouraging Innovation behavior, once again rater gender ($b = -.03$, $t = -.33$, $p = .74$) nor the rater gender-supervisor gender interaction ($b = .07$, $t = .61$, $p = .54$) were significant. Finally, in the model predicting GLI expectations, neither rater gender ($b = .10$, $t = 1.27$, $p = .20$) nor the rater gender-supervisor gender interaction ($b = .06$, $t = .58$, $p = .56$)

were significant. Thus, there is no evidence of rater gender having a moderating effect in the relationship between leader gender and leader behavior expectations.

Rater age was also proposed as a potential moderator of the relationship between leader gender and leader behavior expectations. To examine this, an additional set of moderated regression models were run to consider the interaction between leader gender and rater gender in predicting follower expectations for Supporting behavior, Monitoring behavior, and Encouraging Innovation behavior, and GLI expectations. Rater age was grand mean centered and supervisor gender was dummy coded (male = 1, female = 2) prior to analyses. In the model predicting expectations for Supporting behavior, neither rater age (b = -.01, t = -1.90, p = .06) nor the rater age-supervisor gender interaction (b = .00, t = -.06, p = .96) were significant. In the model predicting expectations for Monitoring behavior, neither rater age (b = -.001, t = -.26, p = .79) nor the rater age-supervisor gender interaction (b = .001, t = -.27, p = .78) were significant. In the model predicting expectations for Encouraging Innovation behavior, rater age was significant (b = -.01, t = -2.23, p = .03) such that expectations increased as age decreased. However, the rater age-supervisor gender interaction was not significant (b = -.001, t = .27, p = .78) in predicting expectations for Encouraging Innovation behavior. Finally, in the model predicting GLI expectations, rater age was significant (b = -.01, t = -2.72, p < .01) such GLI expectations increased as age decreased. However, the rater age-supervisor gender interaction was not significant (b = .00, t = -.08, p = .94) in predicting GLI expectations. Thus, there was also no evidence of rater age having a moderating effect in the relationship between leader gender and leader behavior expectations.

It was also previously mentioned that gender and leadership research often considers ratings of likeability in addition to evaluations of effectiveness as an alternative outcome

variable. As such, I repeated all polynomial regression tests to see if Supporting, Monitoring, and Encouraging Innovation behavior expectations-perceptions incongruence, respectively, have different effects on ratings of liking for men and women supervisors.

To test the effect of Supporting behavior expectations-perceptions congruence on liking, a polynomial regression model was first run without the supervisor gender moderator terms. This model was significant, $F(5,559) = 242.42$, $MSE = 1.05$, $p < .001$. As in the model predicting effectiveness, a_1 and a_3 were significant but a_2 and a_4 were not. To test for the moderating influence of supervisor gender, supervisor gender was added to the model. Next, the five supervisor gender interaction variables were added as a second step to the previous model. This moderated polynomial regression model did not explain significant variance over and above the previous model, $\Delta R^2 = .004$, $p = .18$, indicating that congruence between Supporting behavior expectations and perceptions did not have a different effect on liking evaluations for male and female supervisors.

To test the effect of Monitoring behavior expectations-perceptions congruence on liking, a polynomial regression model was first run without the supervisor gender moderator terms. This model was significant, $F(5,559) = 53.37$, $MSE = 2.83$, $p < .001$. As in the model predicting effectiveness, a_3 was significant but a_1, a_2, and a_4 were not. To test for the moderating influence of supervisor gender, supervisor gender was then added to the model. Next, the five supervisor gender interaction variables were added as a second step to the previous model. The moderated polynomial regression model did not explain significant variance over and above the previous model, $\Delta R^2 = .007$, $p = .44$, indicating that congruence between Monitoring behavior expectations and perceptions did not have a different effect on liking ratings for male and female supervisors.

To test the effect of Encouraging Innovation behavior expectations-perceptions congruence on liking ratings, a polynomial regression model was first run without the supervisor gender moderator terms. This model was significant, $F(5,559) = 23.90$, $MSE = 2.73$, $p < .001$. As in the model predicting effectiveness, a_1 and a_3 were significant but a_2 and a_4 were not. To test for the moderating influence of supervisor gender, supervisor gender was then added to the model. Next, the supervisor gender interaction variables were added as a second step to the previous model. This moderated polynomial regression model also did not explain significant variance over and above the previous model, $\Delta R^2 = .009$, $p = .27$, indicating that congruence between Encouraging Innovation behavior expectations and perceptions did not have a different effect on liking ratings for male and female supervisors.

Discussion

This study contributed to the gender and leadership research literature by explicitly measuring follower expectations for specific leader behaviors and considering how several conditions of follower expectation-behavior incongruence affected evaluations of both female and male supervisors. In line with past findings (Hogue, 2016) and research on descriptive and prescriptive gender stereotypes, respondents in the female supervisor condition reported higher expectations for Supporting, a communal behavior, than respondents in the male supervisor condition. However, respondents in the male supervisor condition did not expect higher levels of agentic Monitoring behavior compared to respondents in the female supervisor condition. This is a notable finding and might indicate a double standard in which women leaders are expected to exhibit high levels of both role-congruent and incongruent behavior, while men are only expected to exhibit high levels of role-congruent behavior. It is also possible that Monitoring behaviors are considered to be less strongly agentic than Supporting behaviors are communal.

In polynomial regression analyses aggregated across gender conditions, results across Supporting, Monitoring, and Encouraging Innovation behaviors indicated that unmet expectations (i.e., high expectations but low levels of behavior) resulted in lower evaluations than when expectations and perceptions were both low. Exceeded expectations were rewarded with higher evaluations compared to when expectations were higher for Monitoring behavior, but expectations did not seem to impact evaluations when Supporting and Encouraging Innovation perceptions were high. Overall, these findings did not provide much support for the met expectations hypothesis as the Y = -X line in each graph did not display an inverse curvilinear relationship; instead, exceeded expectations were generally rewarded compared to "met" expectations (i.e., when expectations and perceptions were congruent).

Results suggest that Supporting and Monitoring behavior expectation-perception incongruence, respectively, did not impact effectiveness or liking evaluations of male and female supervisors differently (i.e., supervisor gender was not a significant moderator in polynomial regression analyses for these two behaviors). However, it is possible that small differences existed between the male and female conditions in specific areas of the response surface graphs but could not be detected by full surface moderation testing. In other words, it is possible that some moderation effects might exist but that polynomial regression moderation testing was not precise enough to detect small differences in specific quadrants of the surface graph. This limitation is inherent in full surface moderation testing and could not be overcome. It can nevertheless be concluded that medium-sized or larger moderation effects for Supporting or Monitoring behavior did not exist in this sample.

Analyses did indicate a statistically significant moderation effect of supervisor gender in the effect of Encouraging Innovation expectations-perceptions incongruence on effectiveness, but

the practical significance of this finding is questionable (i.e., the moderation terms only explained an additional 1.4% of variance in effectiveness ratings). Furthermore, by looking at the simple surface graphs, it seems that the source of this significant moderation effect was that low expectations combined with low Encouraging Innovation perceptions resulted in far lower evaluations for men than for women. While interesting, this finding was unrelated to the hypothesized source of moderation (i.e., incongruent expectations and perceptions).

Finally, exploratory analyses found that participants in the female supervisor condition perceived higher levels of Supporting behavior than did participants in the male supervisor condition. Not surprisingly due to the substantial relationship between Supporting behavior perceptions and liking evaluations, the female supervisor was also rated as more likeable than the male supervisor. The finding that participants perceive higher levels of a communal behavior from the female than male supervisor is notable. A study by Scott and Brown (2006) concluded that individuals a) have a more difficult time encoding agentic behavior from female than male leaders (but not communal behavior from male leaders), and b) less easily encode agentic than communal behavior from women leaders. Findings from the present study seemingly indicate that individuals more easily perceive communal behavior from the female compared to the male supervisor.

Limitations and Future Directions

The use of an experimental design allowed for the manipulation of supervisors' display of two individual leader behaviors while controlling for potential confounds such as situational aspects (e.g., industry) and supervisor characteristics (e.g., experience). This approach also ensured sufficient variance in leader behavior perceptions, a necessary step to observing meaningful incongruence between followers' expectations and leaders' behavior. The

manipulation was generally successful although there did seem to be a small impact of positive perceptions of behaviors relating to positive perceptions of other behaviors as indicated in MANOVA results.

The study was limited by several factors inherent in lab studies, such as less realism and lack of participant interpersonal interaction with supervisors (i.e., participants' only source of information about the supervisor was based on reading a few descriptive sentences). Depicting the supervisor in the vignettes as new to the job and organization could also have affected participants' expectations and evaluations; perhaps they were more lenient than they would be with a supervisor who was not new. Photos were included to enhance the manipulation of supervisor gender; despite being piloted for equivalence in age, attractiveness, and professionalism, it is still a possibility that they introduced confounds. In addition, Encouraging Innovation expectations were measured and behaviors manipulated to compare results for this less inherently gendered behavior to results for the agentic and communal behavior. However, hypotheses related to this behavior predicted null effects; truly confirming null effects would require more stringent standards than those employed in this study.

Future research should continue to examine the role of follower expectations in leading to how women leaders are evaluated in real and contrived settings. For example, if a relatively normal distribution of expectations for some behavior could be sampled, assessing individuals' expectations, perceptions, and evaluations of their actual managers in a field study could be insightful. However, it might be challenging to capture meaningful variance in perceptions of real managers' behavior in a field sample. Further, this study asked participants to rate their expectations of a new supervisor, whereas in field studies it would likely be extremely difficult to gather data from a sample of individuals who are about to enter into a new relationship with a

supervisor. In field samples where respondents are not entering into a new relationship with a supervisor, expectations could be influenced by respondents' past experiences with the supervisor.

Related to this idea, conducting longitudinal research could yield interesting insights into how follower expectations change over time. For example, some emphasize the role of individuating information on social evaluations and suggest stereotypes effects are severely attenuated by individuating information (Landy, 2008). This perspective might suggest that time spent working with a specific supervisor, or experience working with supervisors of a specific gender more generally, might decrease stereotypes. On the other hand, others argue that first impression biases (e.g., self-fulfilling prophecy or behavioral confirmation effects) remain salient and that individuating information does not lead to significant stereotype reductions (Wessel & Ryan, 2008). To this end, leader and gender stereotypes could be explored further by measuring follower expectations via longitudinal research designs.

This study made a contribution by manipulating and measuring agentic and communal behaviors (rather than traits) using two behaviors from Yukl's leader behavior taxonomy. Future gender and leadership research should continue to follow the behavioral approach as this is far less frequently considered than the trait approach in gender and leadership research. Future studies could either include other behaviors from Yukl's taxonomy or behaviors from other leadership taxonomies or models. Researchers should also consider including more specific agentic or communal behaviors employed by leaders that might not be explicitly taxonomized as leader behaviors (e.g., helping behaviors, assertiveness, etc.). In a similar vein, it would be interesting to explore other gendered behaviors that are not directly related to effective

leadership (e.g., friendliness) to explore the extent to which displaying gendered non-leadership behaviors impacts leadership evaluations.

This study could also be replicated but with behaviors or traits that are considered to be ineffective or "dark" (e.g., narcissism, arrogance, "micromanaging" behaviors, interpersonal insensitivity, selfishness, etc.) to see if gender differences emerge in followers' tolerance of ineffective leadership qualities that may or may not be gender stereotyped. In other words, instead of measuring expectations for and evaluating the presence or absence of good leadership, do followers expect different levels of poor leadership attributes from men and women, and is the presence or absence of these attributes evaluated differently when displayed by male versus female leaders?

An additional limitation of this study was that internal consistency reliabilities for the adapted expectations measures were adequate but less than ideal, which has the potential to attenuate effects. Therefore, it could be useful to dedicate attention to developing and validating measures of follower expectations.

Concluding Thoughts

The endurance of descriptive and prescriptive gender stereotypes is troubling and warrants continued investigation of evaluative biases against women leaders. Considering how follower expectations color and combine with behavior perceptions to produce leader evaluations is an understudied avenue of research that could help further explain evaluative biases against women leaders and potentially lead to a richer understanding of the glass labyrinth phenomenon.

APPENDICES

APPENDIX A: Tables

Table 1: Male and Female Headshot Pilot Results

	Male photo		Female photo		Mean difference	
	M	*SD*	*M*	*SD*	*t*	*p*
Estimated Age	36.32	5.87	37.30	5.49	-1.24	.22
Attractiveness	5.82	1.91	6.18	1.77	-1.39	.17
Professionalism	6.94	1.90	7.34	1.76	-1.18	.24

Note. n = 50. Age indicated in years. Attractiveness and professionalism were rated on a 1-10 scale (oriented such that higher numbers indicated greater values).

Table 2: Descriptive Statistics and Bivariate Correlations

Variable	*M(SD)*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Supervisor Gender	1.50(.50)	-																
2. Supporting Expectations	3.99(.73)	**.08**	(.86)															
3. Monitoring Expectations	4.13(.67)	.04	**.51**	(.81)														
4. Encouraging Innovation Expectations	3.98(.74)	.04	**.67**	**.61**	(.85)													
5. GLI Expectations	3.94(.64)	**.11**	**.61**	**.57**	**.62**	(.84)												
6. Supporting Condition	.50(.50)	-.05	.07	.07	.04	.03	-											
7. Monitoring Condition	.50(.50)	.03	-.05	.02	.00	.01	.01	-										
8. Encouraging Innovation Condition	.50(.50)	-.01	.03	-.05	.00	.00	.03	-.03	-									
9. Supporting Perceptions	3.00(1.40)	**.10**	**.14**	**.16**	**.15**	**.15**	**.83**	**.17**	**.12**	(.94)								
10. Monitoring Perceptions	3.09(1.41)	.02	.05	**.15**	**.12**	**.13**	**.11**	**.80**	.04	**.30**	(.96)							
11. Encouraging Innovation Perceptions	3.00(1.44)	.05	**.13**	.08	**.11**	**.14**	**.15**	**.10**	**.80**	**.29**	**.24**	(.94)						
12. Effectiveness	4.41(1.68)	.06	.04	.07	.07	**.10**	**.40**	**.50**	**.33**	**.58**	**.64**	**.55**	(.94)					
13. Liking	4.42(1.81)	**.10**	**.10**	**.11**	**.11**	**.12**	**.70**	**.22**	**.23**	**.82**	**.37**	**.41**	**.75**	(.96)				
14. GLI Evaluations	3.00(1.12)	.04	.07	**.11**	**.09**	**.13**	**.46**	**.51**	**.25**	**.66**	**.69**	**.48**	**.86**	**.76**	(.94)			
15. Gender Role Orientation	2.50(.41)	-.04	**-.14**	**-.15**	**-.13**	**-.13**	-.05	-.03	-.04	**-.14**	-.05	**-.12**	**-.11**	**-.11**	**-.11**	(.76)		
16. Respondent Gender	1.49(.51)	.04	-.05	-.06	-.05	.05	-.03	-.01	.00	-.02	-.04	-.05	.03	-.03	-.03	**.17**	-	
17. Respondent Age	44.68(14.53)	-.04	**-.12**	-.03	**-.13**	**-.17**	-.07	-.05	.01	-.07	-.07	-.06	-.06	-.03	**-.12**	.02	**.10**	-
18. Female manager experience	48.15(25.28)	**.09**	**.10**	.01	**.09**	**.10**	.01	-.07	.00	.03	-.02	.02	-.03	.00	-.01	**.09**	**.25**	**-.21**

Note. N = 564. Bolded values are significant at $p < .05$. Cronbach's alpha reported on the diagonal. All measures were rated on 5-point scales except for effectiveness and liking (7-point scales). All scale measures are oriented such that a higher mean indicates greater levels except for gender role orientation (higher values indicate more egalitarian attitudes). Gender variables were dummy-coded, 1 = male, 2 = female. Condition variables were dummy-coded, 0 = low, 1 = high. Age was indicated in years. Female manager experience indicated as percent of time in respondents' career working for a female manager or supervisor.

Table 3: Frequencies of Followers' Leader Behavior Expectation Levels Over, Under, and In Agreement with Perception Levels

Behavior	Agreement Groups	N	N Percent
Supporting			
	Expectations higher than perceptions	193	34.2
	In agreement	169	30.0
	Expectations lower than perceptions	202	35.8
Monitoring			
	Expectations higher than perceptions	177	31.4
	In agreement	193	34.2
	Expectations lower than perceptions	194	34.4
Encouraging Innovation			
	Expectations higher than perceptions	199	35.3
	In agreement	162	28.7
	Expectations lower than perceptions	203	36.0

Note. N = 564. Agreement is defined as predictor levels being within 0.5 standard deviations of each other.

Table 4: H2 Regression Models Predicting Supporting Expectations as a Function of Supervisor Gender, Gender Role Orientation, and their Interactions

Variable	Supporting Expectations							
	Step 1				*Step 2*			
	b	*β*	*t*	*p*	*b*	*β*	*t*	*p*
Intercept	3.99		130.96	<.001	3.99		130.74	<.001
Supervisor gender	-.06	-.08	-1.90	.06	-.06	-.08	-1.89	.06
Gender role orientation	-.24	-.13	-3.22	.001	-.24	-.13	-3.19	.002
Supervisor gender * gender role orientation interaction					.02	.01	.28	.78

Note. N = 564. The overall model was significant $F(3,560) = 4.83$, $MSE = .52$, $p = .002$, $R^2 = 03$. Gender role orientation was grand mean centered and supervisor gender was effects coded (male = 1, female = -1). Higher gender role orientation values indicate traditional orientation and lower values indicate egalitarian orientation.

Table 5: H2 Regression Models Predicting Monitoring Expectations as a Function of Supervisor Gender, Gender Role Orientation, and their Interactions

Variable	Monitoring Expectations							
	Step 1				*Step 2*			
	b	*β*	*t*	*p*	*b*	*β*	*t*	*p*
Intercept	4.13		147.75	<.001	4.13		147.51	<.001
Supervisor gender	-.02	-.04	-.84	.40	-.02	-.04	-.84	.40
Gender role orientation	-.24	-.15	-3.57	<.001	-.24	-.15	-3.57	<.001
Supervisor gender * gender role orientation interaction					-.01	-.01	-.16	.87

Note. N = 564. The overall model was significant $F(3,560) = 4.58$, $MSE = .44$, $p = .004$, $R^2 = 02$. Gender role orientation was grand mean centered and supervisor gender was effects coded (male = 1, female = -1). Higher gender role orientation values indicate traditional orientation and lower values indicate egalitarian orientation.

Table 6: Supporting Behavior Expectations-Perceptions Discrepancy Predicting Effectiveness

Variable	*b* (se)	*t*
Constant	4.63 (.12)	38.06*
Supporting expectations	-.08 (.17)	-.44
Supporting perceptions	.64 (.08)	8.37*
Supporting expectations squared	.02 (.09)	.16
Supporting expectations x Supporting perceptions	.06 (.06)	.99
Supporting perceptions squared	-.09 (.04)	-2.08*
R^2	.34	57.26* (*F*)
Surface tests		
Linear congruence (a_1)	.56*	3.26*
Quadratic congruence (a_2)	-.02	-.18
Linear incongruence (a_3)	-.71	-3.48*
Quadratic incongruence (a_4)	-.13	-1.05

Note. $N = 564$. *$p < .05$

$a_1 = (b_1 + b_2)$, where b_1 is unstandardized coefficient for Supporting expectations and b_2 is unstandardized coefficient for Supporting perceptions. $a_2 = (b_3 + b_4 + b_5)$, where b_3 is unstandardized coefficient for Supporting expectations squared, b_4 is unstandardized coefficient for the cross-product of Supporting expectations and perceptions, and b_5 is unstandardized coefficient for Supporting perceptions squared. $a_3 = (b_1 – b_2)$.

$a_4 = (b_3 – b_4 + b_5)$.

b unstandardized regression coefficient, *se* standard error. Significance depends in part on standard errors, thus *a* values of equivalent magnitude may not both be significant.

Table 7: Monitoring Behavior Expectations-Perceptions Discrepancy Predicting Effectiveness

Variable	*b* (se)	*t*
Constant	4.58 (.13)	35.79*
Monitoring expectations	-.37 (.20)	-1.84
Monitoring perceptions	.65 (.08)	7.79*
Monitoring expectations squared	.19 (.10)	1.95
Monitoring expectations x Monitoring perceptions	.08 (.06)	1.38
Monitoring perceptions squared	-.08 (.04)	-2.00*
R^2	.42	81.39* (*F*)
Surface tests		
Linear congruence (a_1)	.28	1.37
Quadratic congruence (a_2)	.19	1.75
Linear incongruence (a_3)	-1.02	-4.52*
Quadratic incongruence (a_4)	.03	.22

Note. $N = 564$. *$p < .05$

$a_1 = (b_1 + b_2)$, where b_1 is unstandardized coefficient for Monitoring expectations and b_2 is unstandardized coefficient for Monitoring perceptions. $a_2 = (b_3 + b_4 + b_5)$, where b_3 is unstandardized coefficient for Monitoring expectations squared, b_4 is unstandardized coefficient for the cross-product of Monitoring expectations and perceptions, and b_5 is unstandardized coefficient for Monitoring perceptions squared. $a_3 = (b_1 – b_2)$.

$a_4 = (b_3 – b_4 + b_5)$.

b unstandardized regression coefficient, *se* standard error. Significance depends in part on standard errors, thus *a* values of equivalent magnitude may not both be significant.

Table 8: Encouraging Innovation Behavior Expectations-Perceptions Discrepancy Predicting Effectiveness

Variable	*b* (se)	*t*
Constant	4.50 (.13)	36.00*
Encouraging Innovation expectations	-.12 (.18)	-.65
Encouraging Innovation perceptions	.55 (.07)	7.79*
Encouraging Innovation expectations squared	.08 (.10)	.88
Encouraging Innovation expectations x Monitoring perceptions	.08 (.06)	1.46
Encouraging Innovation perceptions squared	-.06 (.04)	-1.38
R^2	.31	81.39* (*F*)
Surface tests		
Linear congruence (a_1)	.44	2.34*
Quadratic congruence (a_2)	.11	1.01
Linear incongruence (a_3)	-.67	-3.39*
Quadratic incongruence (a_4)	-.05	-.42

Note. $N = 564$. *$p < .05$

$a_1 = (b_1 + b_2)$, where b_1 is unstandardized coefficient for Encouraging Innovation expectations and b_2 is unstandardized coefficient for Encouraging Innovation perceptions. $a_2 = (b_3 + b_4 + b_5)$, where b_3 is unstandardized coefficient for Encouraging Innovation expectations squared, b_4 is unstandardized coefficient for the cross-product of Encouraging Innovation expectations and perceptions, and b_5 is unstandardized coefficient for Encouraging Innovation perceptions squared. $a_3 = (b_1 – b_2)$.

$a_4 = (b_3 – b_4 + b_5)$.

b unstandardized regression coefficient, *se* standard error. Significance depends in part on standard errors, thus *a* values of equivalent magnitude may not both be significant.

APPENDIX B: Figures

Figure 1: Hypothesized Interaction: The Effect of Gender Role Orientation on Supporting Expectations

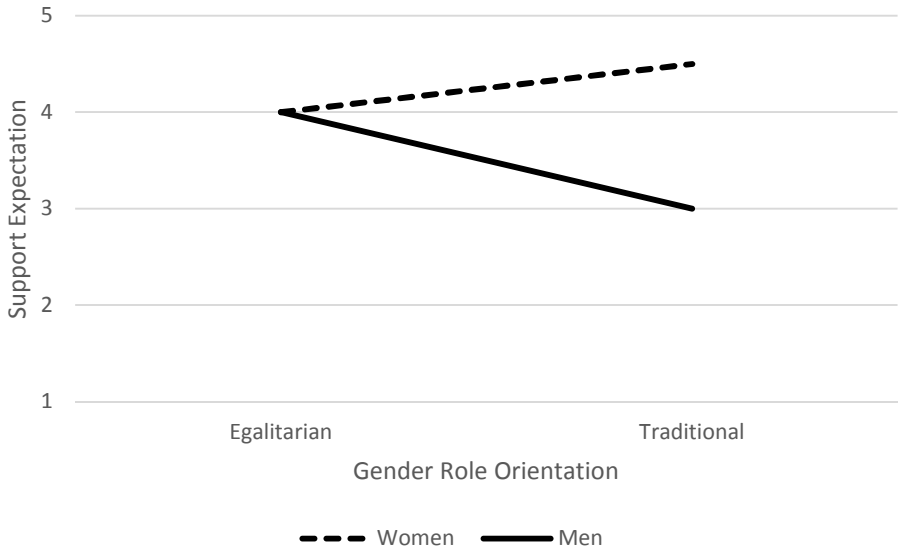


Figure 2: Hypothesized Interaction: The Effect of Gender Role Orientation on Monitoring Expectations

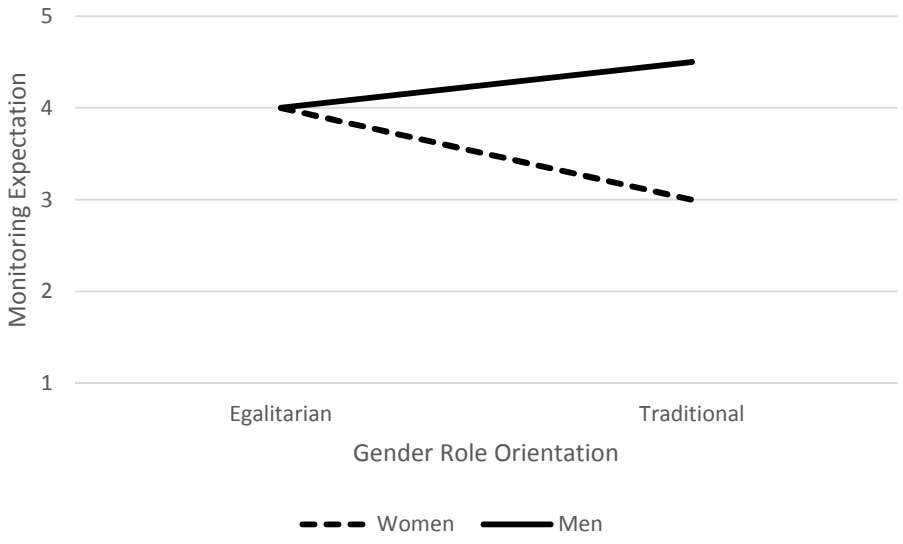


Figure 3: Hypothesized Interaction Response Surface for Expectations and Supporting (Monitoring) Behavior on Evaluations of Female (Male) Managers

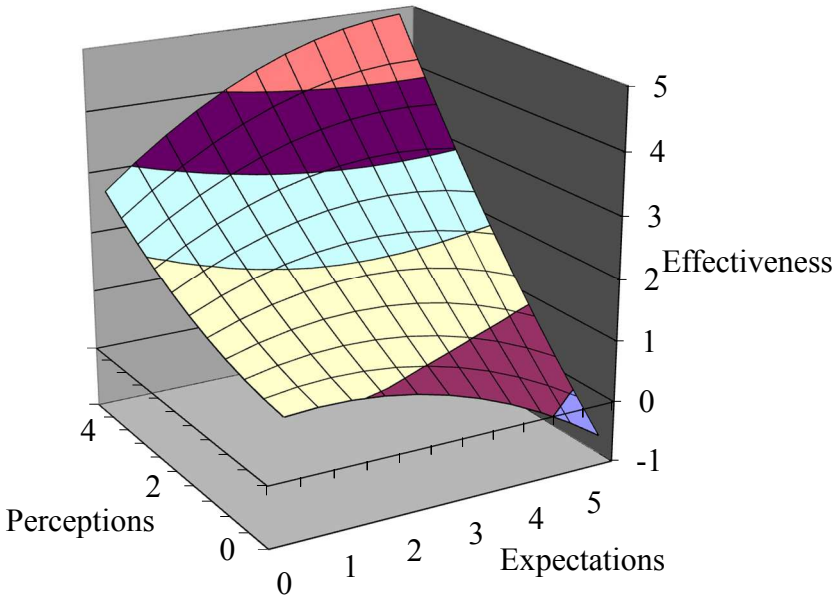


Figure 4: Hypothesized Interaction Response Surface for Expectations and Supporting (Monitoring) Behavior on Evaluations of Male (Female) Managers

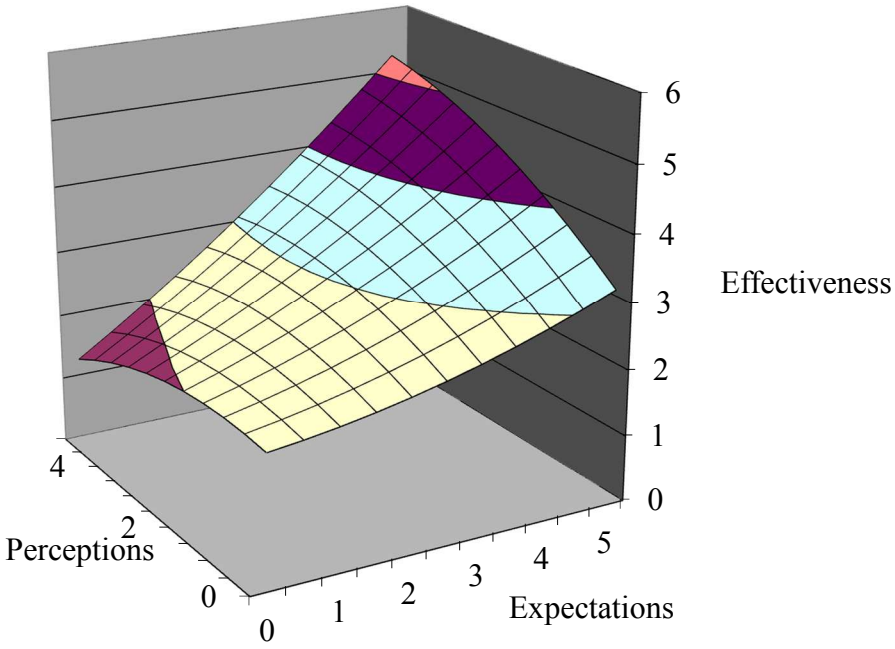


Figure 5: Response Surface for Supporting Behavior Expectations and Perceptions on Effectiveness (Male and Female Supervisor Conditions)

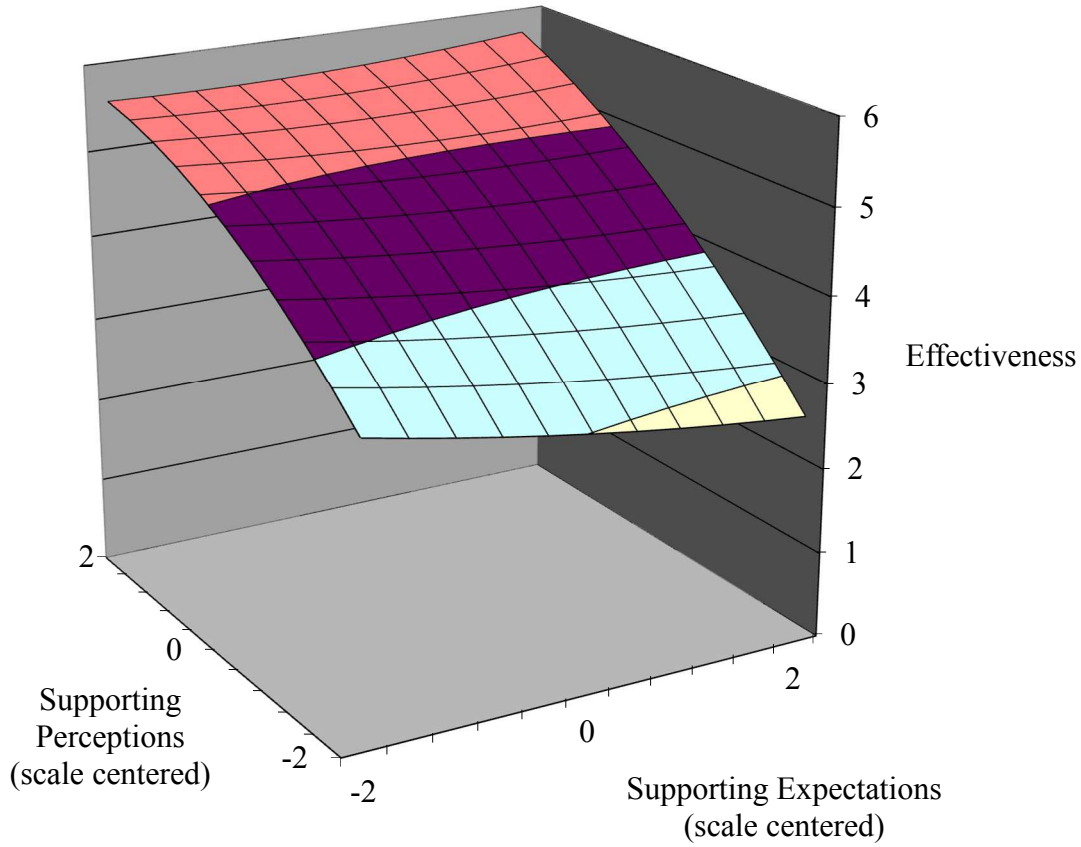


Figure 6: Response Surface for Monitoring Behavior Expectations and Perceptions on Effectiveness (Male and Female Supervisor Conditions)

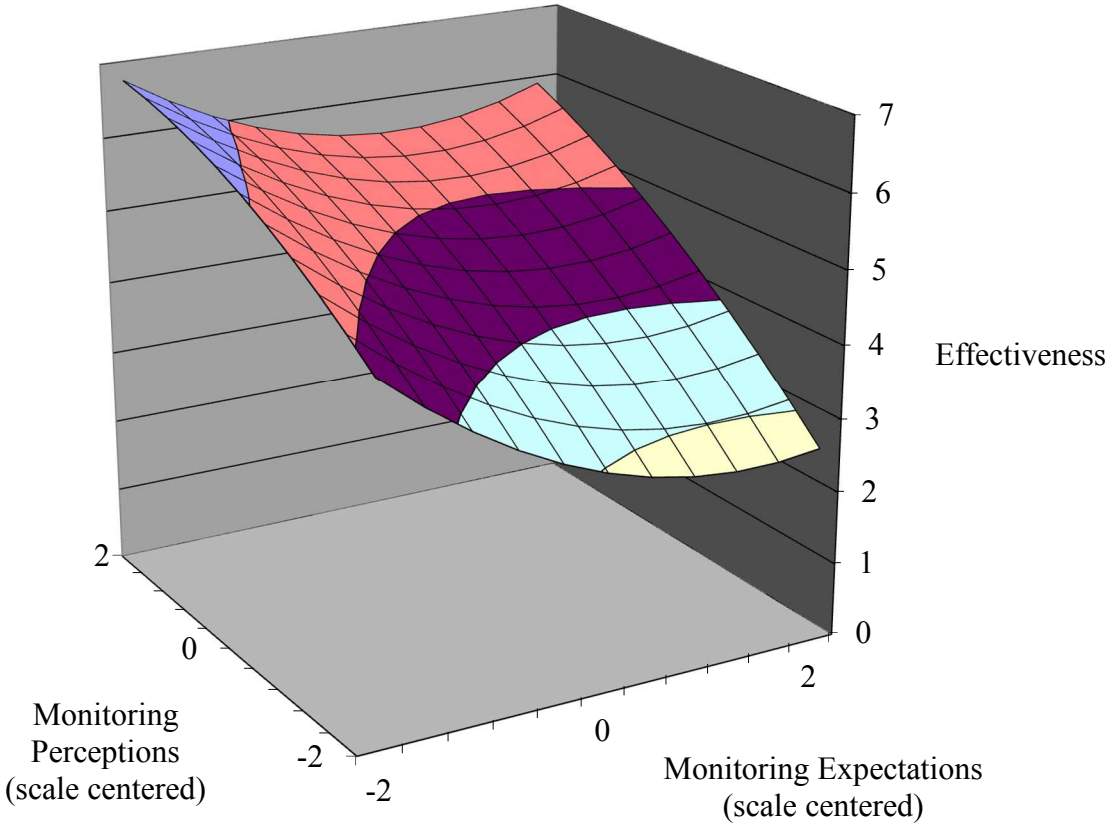


Figure 7: Response Surface for Encouraging Innovation Behavior Expectations and Perceptions on Effectiveness (Male and Female Supervisor Conditions)

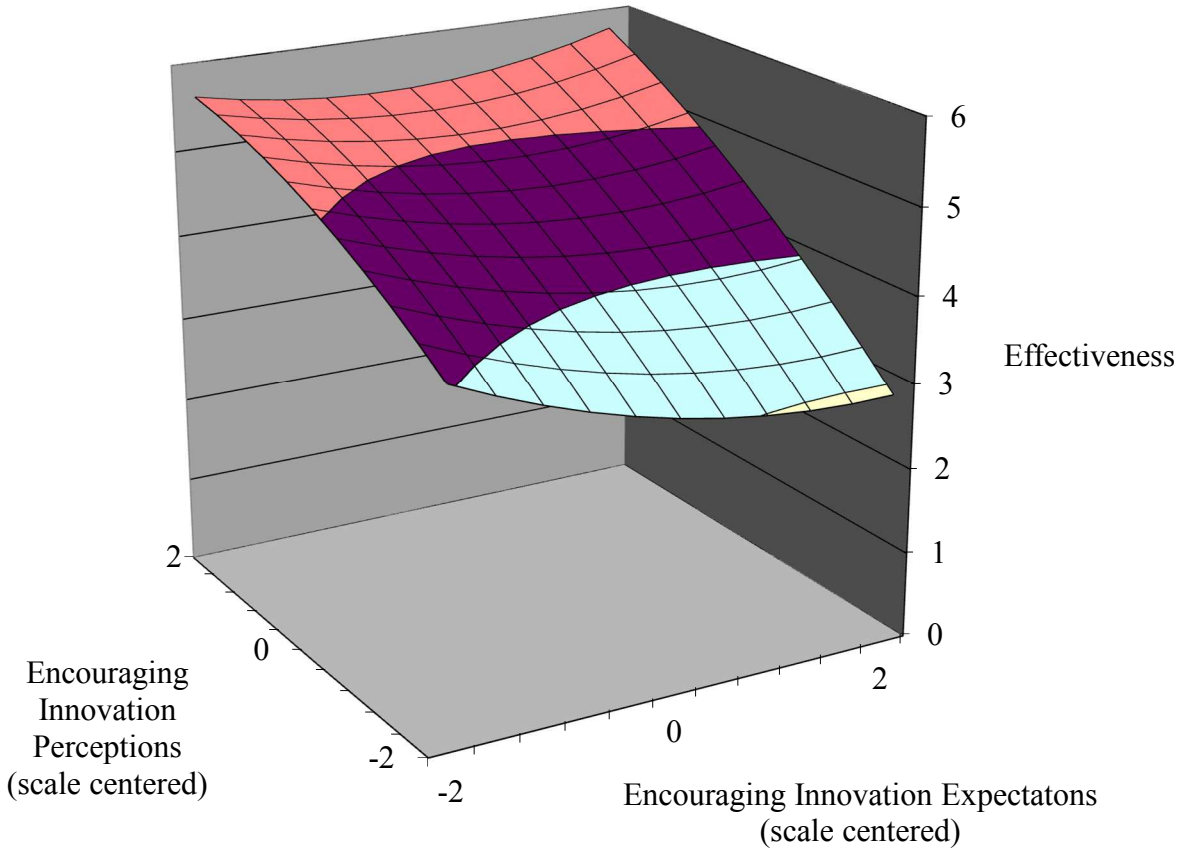


Figure 8: Simple Surface for Encouraging Innovation Behavior Expectations and Perceptions on Effectiveness (Male Supervisor Condition Only)

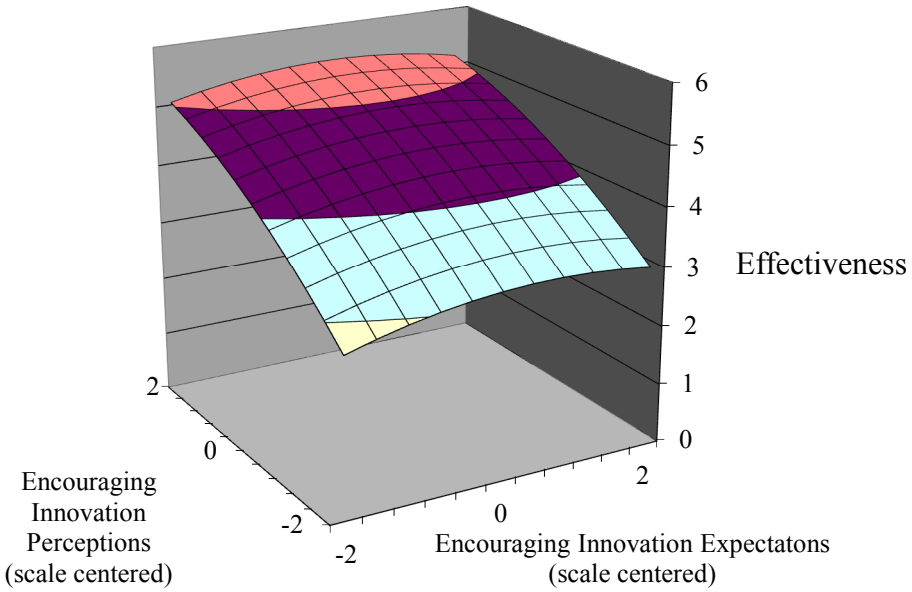
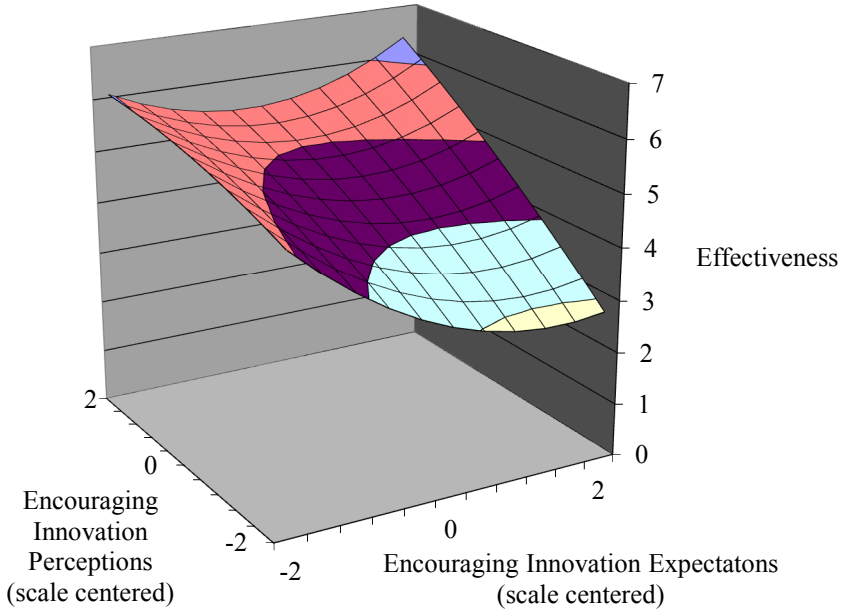


Figure 9: Simple Surface for Encouraging Innovation Behavior Expectations and Perceptions on Effectiveness (Female Supervisor Condition Only)



APPENDIX C: Survey Screening Questions

1) What is your gender?
2) What is your age?
3) Of the past six months, how many months have you worked in a full-time job? (0-6)
    a. If less than 2, term
4) Do you (or did you) have a direct supervisor or manager to whom you report?
    a. If no, term
5) How often do (or did you) interact with your direct supervisor or manager? (0-5 days per week)
    a. If less than 2 days per week, term

APPENDIX D: Supervisor Vignettes

Vignette 1a (presented to all participants)
The following vignette will describe a workplace situation. Imagine that you are in this situation and consider how you would react. Afterwards, you will be asked a series of questions.

You are an employee at ABC corporation, a professional services firm. You have worked for ABC for the past three years. During this time, you have had the same supervisor to whom you have reported for all three years. However, your supervisor was recently promoted and left for their new job a few weeks ago. Luckily, ABC moved quickly and has already hired your new supervisor, who will start next week.

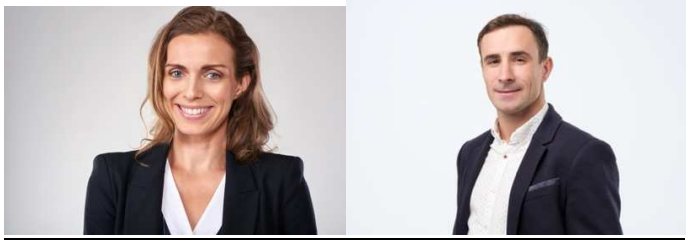
Attention Check 1:
Which best describes the current status of your supervisor?

a) My supervisor for the past three years is still currently my supervisor.
b) I do not currently have a supervisor, and ABC has not yet hired a replacement.
c) I do not currently have a supervisor. ABC has already hired a replacement who will start next week.

Vignette 1b (participants were randomly assigned to male/female supervisor conditions where they saw one of the two photos below with the supervisor's name underneath)

You recently found out that your new supervisor will be *Ken/Kelly* Green. You have never met *Ken/Kelly* and do not know much about *him/her* until you receive an email from a company leader introducing *Ken/Kelly* and summarizing *his/her* resume. Based on this summary, you learn that *Ken/Kelly* has a bachelor's degree in Business Management, and *he/she* had previously worked in a supervisory role at another company for the past four years. The email says that *Ken/Kelly* is excited to begin *his/her* new role as your supervisor next week.

You are excited to meet *Ken/Kelly* next week, but also a bit anxious to find out about *his/her* leadership style and what it will be like to have *him/her* as your supervisor.



Kelly Green Ken Green

Attention Check 2:
Which of the following two statements are true?

a) My new supervisor is a man.
b) My new supervisor is a woman.
c) My new supervisor has prior experience as a supervisor.
d) My new supervisor does not have prior experience as a supervisor.

Vignette 2 (participants were randomly assigned to high or low Supporting, Monitoring, and Encouraging Innovation conditions)

Fast forward one month. This next vignette will describe your interactions with your new supervisor after their first few weeks on the job. Imagine that you are in this situation and consider how you would react. Afterwards, you will be asked a series of questions about the supervisor's leadership.

High Supporting

In the three weeks since Ken/Kelly began working as your supervisor, you have made several observations about his/her tendencies as a leader. For example, Ken/Kelly regularly shows acceptance and positive regard towards you by doing things like trying to spend some time getting to know you. Ken/Kelly also provides sympathy and support to you when you are anxious or upset about work stressors and difficulties. In addition, Ken/Kelly seems to always be willing to listen to you and help you if you are having a hard time, even with topics relating to your personal life. Another theme you've noticed is that Ken/Kelly spends time trying to boost your self-confidence and provide encouragement when you're working through difficult tasks or experiencing setbacks.

Low Supporting

In the three weeks since Ken/Kelly began working as your supervisor, you have made several observations about his/her tendencies as a leader. For example, Ken/Kelly jumped right into a working relationship with you and your teammates without spending time getting to know you. A few times when you have been anxious or upset about work stressors and difficulties, Ken/Kelly didn't seem to take much notice or give you any special attention. In addition, if you are having a hard time with anything, especially related to topics in your personal life, you don't think you would be very comfortable discussing it with Ken/Kelly as he/she seems to prefer keeping work and personal lives separate. Another theme you've noticed is that Ken/Kelly isn't one to go out of his/her way to give you encouragement or a confidence boost when you're working through difficult tasks or experiencing setbacks. He/she simply expects the work to be done.

High Monitoring

You have also noticed that Ken/Kelly emphasizes achievement and performance. He/she is very attuned to the daily operations and performance of his/her team. To do this, he/she frequently walks around to observe you and his/her other subordinates and to ask questions about the work you are doing. Ken/Kelly has already held several progress review meetings with you to discuss projects, review your work, and give feedback on your performance. Ken/Kelly then applies information gathered from observations and progress meetings to decisively take action; this might be finding ways to address deficits or poor performance, or to praise high performance.

Low Monitoring

You have also noticed that Ken/Kelly hasn't talked much about topics relating to achievement or performance. He/she doesn't seem to be very attuned to the daily operations and performance of his/her team. You don't see him/her around much, and he/she doesn't ask you

many questions about the work you are doing. Ken/Kelly hasn't held any progress review meetings with you to discuss projects, review your work, or give feedback on your performance. Because of this, Ken/Kelly doesn't know much about what you are doing and hasn't taken any actions to address deficits or poor performance, or to praise high performance.

High Encouraging Innovation

Another thing you have noticed so far is that Ken/Kelly often encourages you and your team to be creative and innovative. He/she says that when approaching problems, the team should look at the situation from different perspectives and think outside the box when developing solutions. Ken/Kelly also encourages the team to experiment with new ideas or to find ideas in other fields that could be applied to current problems or tasks. Ken/Kelly says that he/she wants the team to feel comfortable suggesting new ideas or different ways of doing things. He/she seems very open to change and actively encourages you and the team to be as well.

Low Encouraging Innovation

Another thing you have noticed so far is that Ken/Kelly has mentioned multiple times that he/she is someone who thinks continuity and stability are important. He/she says that when approaching problems and developing solutions, the team should start by using proven methods. When thinking about how to solve current problems and tasks, Ken/Kelly encourages the team to apply what they did in relevant past situations. Ken/Kelly also says new ideas are not worth exploring in situations where current processes work just fine. He/she seems to really prefer that the team keep doing things in the same way as we have been.

Attention Check 3:

Which of the following statements is true about your supervisor's leadership?

When I'm working through difficult tasks or experiencing setbacks, my supervisor…

a) Spends time trying to boost my self-confidence and provide encouragement.
b) Isn't one to go out of their way to give encouragement or a confidence boost.

Attention Check 4:

Which of the following statements is true about your supervisor's leadership?

a) My supervisor is very attuned to the daily operations and performance of the team.
b) My supervisor doesn't seem to be very attuned to the daily operations and performance of the team.

Attention Check 5:

Which of the following statements is true about your supervisor's leadership?

a) My supervisor encourages the team to be creative and innovative.
b) My supervisor encourages the team to stick to tried and true methods.

APPENDIX E: Study Measures

Leader Behavior Expectations (adapted from Yukl, 2012)
Stem: "I expect that Ken/Kelly will…"
Supporting
–Show concern for the needs and feelings of individual members of the work unit.
–Provide support and encouragement when there is a difficult or stressful task.
–Express confidence that members of the unit can perform a difficult task.
–Show sympathy and understanding when a member is worried or upset.
Monitoring
–Check on the progress and quality of the work.
–Evaluate how well important tasks or projects are being performed.
–Request progress reports for an important task or assignment.
–Evaluate the job performance of unit members in a systematic way.
Encouraging Innovation
–Encourage innovative thinking and creative solutions to problems.
–Talk about the importance of innovation and flexibility for the success of the unit.
–Encourage members to look for better ways to accomplish work unit objectives.
–Ask questions that encourage members to think about old problems in new ways.
Scale: 1 = Not at all, 2 = To a limited extent, 3 = To a moderate extent, 4 = To a considerable extent, 5 = To a very great extent

General Leadership Impression - Expectations (adapted from Cronshaw & Lord, 1987)
–How much leadership do you expect Ken/Kelly will exhibit?
–How willing would you be to choose Ken/Kelly as a leader?
–How typical of a leader do you expect Ken/Kelly to be?
–To what extent do you expect Ken/Kelly to engage in leader behavior?
–To what degree do you expect Ken/Kelly to fit your image of a leader?
Scale: 1 = Not at all, 2 = To a limited extent, 3 = To a moderate extent, 4 = To a considerable extent, 5 = To a very great extent

How would you describe the leader's leadership? Please write at least two sentences.

Leader Behavior Perceptions (Managerial Practices Survey; Yukl, 2012)
Stem: "Based on the vignettes, I think Ken/Kelly…"
Supporting
–Shows concern for the needs and feelings of individual members of the work unit.
–Provides support and encouragement when there is a difficult or stressful task.
–Expresses confidence that members of the unit can perform a difficult task.
–Shows sympathy and understanding when a member is worried or upset.
Monitoring
–Checks on the progress and quality of the work.
–Evaluates how well important tasks or projects are being performed.
–Requests progress reports for an important task or assignment.
–Evaluates the job performance of unit members in a systematic way.
Encouraging Innovation

–Encourages innovative thinking and creative solutions to problems.
–Talks about the importance of innovation and flexibility for the success of the unit.
–Encourages members to look for better ways to accomplish work unit objectives.
–Asks questions that encourage members to think about old problems in new ways.
Scale: 1 = Not at all, 2 = To a limited extent, 3 = To a moderate extent, 4 = To a considerable extent, 5 = To a very great extent

Perceived Leader Effectiveness (Johnson et al., 2008)
–Ken/Kelly will be effective.
–Ken/Kelly will succeed in this role.
–Ken/Kelly will improve performance at this company.
Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Agree, 7 = Strongly Agree

Leader Likability (Johnson et al., 2008)
–Ken/Kelly will be liked by his/her employees.
–Ken/Kelly seems likeable.
–Ken's/Kelly's employees will like working for him/her.
Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Agree, 7 = Strongly Agree

General Leadership Impression Measure (Cronshaw & Lord, 1987)
–How much leadership did Ken/Kelly exhibit?
–How willing would you be to choose Ken/Kelly as a leader?
–How typical was Ken/Kelly of a leader?
–To what extent did Ken/Kelly engage in leader behavior?
–To what degree did Ken/Kelly fit your image of a leader?
Scale: 1 = Not at all, 2 = To a limited extent, 3 = To a moderate extent, 4 = To a considerable extent, 5 = To a very great extent

Gender Role Stereotypes Scale (Mills et al., 2012)
Please indicate the extent to which you believe each task should be done by the man, should be done by the woman, or the man and woman share the responsibility equally, when there is a relationship between a man and a woman.

1. Mow the lawn
2. Drive the car when both the man and the woman are traveling
3. Prepare meals
4. Propose marriage
5. Perform basic maintenance of vehicles, such as changing the oil
6. Handle financial matters, such as paying bills
7. Perform household cleaning
8. Wash, fold, and put away laundry
9. Purchase groceries
10. Earn most of the money to support the family
11. Wrap gifts (e.g., birthday or holiday presents)
12. Decorate the house

13. Shovel snow to clear driveways and sidewalks
14. Stay home with a child who is sick

Scale: 1 = should always be done by the man, 2 = should usually be done by the man, 3 = equal responsibility, 4 = should usually be done by the woman, 5 = should always be done by the woman

Demographics

–Gender, age, and employment status measured in screening questions (see Appendix C).

–In your working career, have you worked for a female manager/supervisor before? (Yes/no)

–If yes: Out of your entire working career, approximately what percent of the time have you worked for a female manager/supervisor? (0-100 slider)

REFERENCES

REFERENCES

Allen, T. D., & Russell, J. E. (1999). Parental Leave of Absence: Some Not So Family□Friendly Implications. *Journal of applied social psychology*, *29*(1), 166-191.

Avolio, B. J., Gardner, W. L., Walumbwa, F. O., Luthans, F., & May, D. R. (2004). Unlocking the mask: A look at the process by which authentic leaders impact follower attitudes and behaviors. *Leadership Quarterly*, *15,* 801-823.

Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual review of psychology*, *60*, 421-449.

Ayman, R., & Korabik, K. (2010). Leadership: Why gender and culture matter. *American psychologist*, *65*(3), 157.

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Rand McNally.

Barbuto, J. E., & Gifford, G. T. (2010). Examining gender differences of servant leadership: An analysis of the agentic and communal properties of the Servant Leadership Questionnaire. *Journal of Leadership Education*, *9*(2), 4-21.

Bass, B. M. (1985). Leadership and performance beyond expectations.

Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational dynamics*, *18*(3), 19-31.

Bass, B. M. & Avolio, B. J. (1990). *Transformational leadership development: Manual for the Multifactor Leadership Questionnaire.* Palo Alto, CA: Consulting Psychologist Press.

Bass, B. M., & Bass, R. (2008). *The bass handbook of leadership: theory, research, and managerial applications* (4th ed.). New York: Free Press.

Bass, B. M., & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for under- standing successful performance. *Applied Psychology: An International Review, 40*(4), 437–454.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias?. *The Quarterly journal of economics*, *124*(4), 1497-1540.

Catalyst. (2020). *Pyramid: Women in S&P 500 Companies.* Retrieved from https://www.catalyst.org/research/women-in-sp-500-companies/.

Cialdini, R. B., & Trost, M. R. (1998). *Social influence: Social norms, conformity and compliance.* In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (p. 151–192). McGraw-Hill.

Chen, J. J. (2008). *Ordinary vs. extraordinary: Differential reactions to men's and women's prosocial behavior in the workplace* (Doctoral dissertation, New York University).

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation.* Erlbaum, Hillsdale, NJ.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Harvard Business Press.

Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological bulletin*, *129*(4), 569.

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*(1), 3-22.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American psychologist*, *75*(3), 301.

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 458-476). SAGE Publications Ltd, https://www.doi.org/10.4135/9781446249222.n49

Edmondson, A. (2003). Speaking up in the operating room. *Journal of Management Studies*, 40, 1419–1452.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational behavior and human decision processes*, *58*(1), 51-100.

Edwards, J. R. (2001). Ten difference score myths. *Organizational research methods*, *4*(3), 265-287.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36,* 1577-1613.

Elenkov, D. S., Judge, W., & Wright, P. (2005). Strategic leadership and executive innovation influence: An international multi-cluster comparative study. *Strategic Management Journal,* 26, 665–682.

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.

Epitropaki, O., & Martin, R. (2004). Implicit leadership theories in applied settings: factor structure, generalizability, and stability over time. *Journal of applied psychology*, *89*(2), 293.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of applied psychology*, *37*(1), 1.

Fleishman, E. A. (1995). Consideration and structure: Another look at their role in leadership research. In F. Dansereau & F. J. Yammarino (Eds.), *Leadership: The multiple-level approaches* (pp. 51–60). Stamford, CT: JAI Press.

Flynn, F. J., & Ames, D. R. (2006). What's good for the goose may not be as good for the gander: The benefits of self-monitoring for men and women in task groups and dyadic conflicts. *Journal of Applied Psychology*, *91*(2), 272.

Fry, L. W. (2003). Toward a theory of spiritual leadership. *The leadership quarterly*, *14*(6), 693-727.

Fry, L. W. (2005). Editorial: Introduction to The Leadership Quarterly special issue: Toward a paradigm of spiritual leadership. *The Leadership Quarterly, 16*(5), 619-622. https://doi.org/10.1016/j.leaqua.2005.07.001

Gallup. (2017). Americans No Longer Prefer Male Boss to Female Boss. Retrieved March 11, 2019, from https://news.gallup.com/poll/222425/americans-no-longer-prefer-male-boss-female-boss.aspx

Gergen, D. (2005). Women leading in the twenty-first century. In L. Coughlin, E. Wingard, & K. Hollihan (Eds.), *Enlightened power: How women are transforming the practice of leadership* (pp. xv-xxix). San Francisco, CA: Jossey-Bass.

Greenleaf, R.K. (1977). *Servant Leadership*. Paulist Press, New York, NY.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. Staw & L. Cummings (Eds.), Research in organizational behavior (Vol. 5, pp. 269–298). Greenwich, CT: JAI Press.

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 and Personality, 10*, 3–26.

Heilman, M. E. (1997). Sex discrimination and the affirmative action remedy: The role of sex stereotypes. *Journal of Business Ethics, 16*, 877–889.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*, 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. (1995). Sex stereotypes: Do they influence perceptions of managers?. *Journal of Social behavior and Personality*, *10*(4), 237.

Heilman, M. E., & Chen, J. J. (2005). Same behavior, different consequences: reactions to men's and women's altruistic citizenship behavior. *Journal of Applied Psychology*, *90*(3), 431.

Heilman, M. E., & Wallen, A. S. (2010). Wimpy and undeserving of respect: Penalties for men's gender-inconsistent success. *Journal of Experimental Social Psychology, 46*(4), 664–667. https://doi.org/10.1016/j.jesp.2010.01.008

Hogue, M. (2016). Gender bias in communal leadership: Examining servant leadership. *Journal of Managerial Psychology*.

Hogue, M., & Lord, R. G. (2007). A multilevel, complexity theory approach to understanding gender bias in leadership. *The Leadership Quarterly*, *18*(4), 370-390.

Howell, J. M., & Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated business unit performance. *Journal of Applied Psychology*, 78, 891–902.

Howell, S. E., & Day, C. L. (2000). Complexities of the gender gap. *Journal of Politics*, *62*(3), 858-874.

Humberg, S., Nestler, S., & Back, M. D. (2019). Response surface analysis in personality and social psychology: Checklist and clarifications for the case of congruence hypotheses. *Social Psychological and Personality Science*, *10*(3), 409-419.

Ilies, R., Morgeson, F. P., & Nahrgang, J. D. (2005). Authentic leadership and eudaemonic well-being: Understanding leader–follower outcomes. *The leadership quarterly*, *16*(3), 373-394.

Irving, P. G., & Montes, S. D. (2009). Met expectations: The effects of expected and delivered inducements on employee satisfaction. *Journal of Occupational and Organizational Psychology*, *82*(2), 431-451.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational behavior and human decision processes*, *106*(1), 39-60.

Judge, T. A., & Livingston, B. A. (2008). Is the gap more than gender? A longitudinal analysis of gender, gender role orientation, and earnings. *Journal of applied psychology*, *93*(5), 994.

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *Journal of applied psychology*, *89*(5), 755.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of applied psychology*, *89*(1), 36.

Kim, H., & Yukl, G. (1995). Relationships of managerial effectiveness and advancement to self-reported and subordinate-reported leadership behaviors from the multiple-linkage mode. *The leadership quarterly*, *6*(3), 361-377.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological bulletin*, *137*(4), 616.

Komaki, J. L. (1986). Toward effective supervision: An operant analysis and comparison of managers at work. *Journal of applied psychology*, *71*(2), 270.

Komaki, J. L., Desselles, M. L., & Bowman, E. D. (1989). Definitely not a breeze: Extending an operant model of effective supervision to teams. *Journal of Applied Psychology*, *74*(3), 522.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 379–392.

Larsen, K. S., & Long, E. (1988). Attitudes toward sex-roles: Traditional or egalitarian? *Sex Roles, 19,* 1–12.

Lee, J., & Hoon, T. H. (1993). Business Students' Perceptions of Women in Management-the Case in Singapore. *Management Education and Development*, *24*(4), 415-429.

Lipman-Blumen, J. (2000). *Connective leadership: Managing in a changing world*. Oxford University Press.

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational behavior and human performance*, *34*(3), 343-378.

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. https://doi.org/10.1037/0021-9010.91.4.777

Massengill, D., & Di Marco, N. (1979). Sex-role stereotypes and requisite management characteristics: A current replication. *Sex Roles*, *5*(5), 561-570.

McCarthy, J. (2020, October 29). Less Than Half in U.S. Would Vote for a Socialist for President. Retrieved from https://news.gallup.com/poll/254120/less-half-vote-socialist-president.aspx

McCauley, C. D. (2004). *Successful and unsuccessful leadership.* In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.), *The nature of leadership* (p. 199–221). Sage Publications, Inc.

McKinsey. (2020, October 08). *Women in the Workplace 2020.* Retrieved from https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace.

Morrison, A. M., White, R. P., White, R. P., & Van Velsor, E. (1987). *Breaking The Glass Ceiling: Can Women Reach The Top Of America's Largest corporations?*. Pearson Education.

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.

Nye, J. L., & Forsyth, D. R. (1991). The effects of prototype-based biases on leadership appraisals: A test of leadership categorization theory. *Small Group Research*, *22*(3), 360-379.

Offermann, L. R., Kennedy Jr, J. K., & Wirtz, P. W. (1994). Implicit leadership theories: Content, structure, and generalizability. *The leadership quarterly*, *5*(1), 43-58.

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of applied psychology*, *99*(6), 1129.

Phillips, J. S., & Lord, R. G. (1982). Schematic information processing and perceptions of leadership in problem solving groups. *Journal of Applied Psychology, 67*, 486-492.

Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin, 80*, 151-176.

Powell, G. N., & Butterfield, D. A. (1979). The "good manager": Masculine or androgynous?. *Academy of Management Journal*, *22*(2), 395-403.

Rosette, A. S., Mueller, J. S., & Lebel, R. D. (2015). Are male leaders penalized for seeking help? The influence of gender and asking behaviors on competence perceptions. *The Leadership Quarterly*, *26*(5), 749-762.

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology, 95*(2), 221–235. https://doi.org/10.1037/a0018204

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counter stereotypical impression management. *Journal of personality and social psychology*, *74*(3), 629.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*(5), 1004.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of social issues*, *57*(4), 743-762.

Rudman, L. A., & Mescher, K. (2013). Penalizing men who request a family leave: Is flexibility stigma a femininity stigma?. *Journal of Social Issues*, *69*(2), 322-340.

Shamir, B., & Eilam, G. (2005). "What's your story?" A life-stories approach to authentic leadership development. *The leadership quarterly*, *16*(3), 395-417.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology, 57*, 95–100. doi: 10.1037/h0037128

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology, 60*, 340–344. doi: 10.1037/h0076637

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues, 57(4),* 675-688.

Schein, E. H., & Schein, P. A. (2018). *Humble leadership: The power of relationships, openness, and trust*. Berrett-Koehler Publishers.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*(4), 543-554.

Shinar, E. H. (1975). Sexual stereotypes of occupations. *Journal of vocational behavior*, *7*(1), 99-111.

Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational behavior and human decision processes*, *101*(2), 230-242.

Spence, J. T., & Helmreich, R. (1972). The Attitudes Toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *JSAS Catalog of Selected Documents in Psychology, 2,* 66–67.

Stoker, J. I., Van der Velde, M., & Lammers, J. (2012). Factors Relating to Managerial Stereotypes: The Role of Gender of the Employee and the Manager and Management Gender Ratio. *Journal of Business and Psychology, 27*(1), 31–42. JSTOR.

U.S. Bureau of Labor Statistics. (2020). *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity.* Retrieved from https://www.bls.gov/cps/cpsaat11.htm.

Van Dierendonck, D. (2011). Servant leadership: A review and synthesis. *Journal of management*, *37*(4), 1228-1261.

Wang, A. C., Chiang, J. T. J., Tsai, C. Y., Lin, T. T., & Cheng, B. S. (2013). Gender makes the difference: The moderating role of leader gender on the relationship between leadership styles and subordinate performance. *Organizational Behavior and Human Decision Processes*, *122*(2), 101-113.

Wayne, J. H., & Cordeiro, B. L. (2003). Who is a good organizational citizen? Social perception of male and female employees who use family leave. *Sex roles*, *49*(5), 233-246.

Wessel, J. L., & Ryan, A. M. (2008). Past the first encounter: The role of stereotypes. *Industrial and Organizational Psychology*, *1*(4), 409-411.

Wiegand, J., Drasgow, F., & Rounds, J. (2020). Misfit Matters: A Re-Examination of Interest Fit and Job Satisfaction. *Journal of Vocational Behavior*, 103524.

Wojciszke, Abele, & Baryla. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology, 39*, 973–990. doi: 10.1002/ejsp

Yukl, G. A. (1999). An evaluative essay on current conceptions of effective leadership. *European journal of work and organizational psychology*, *8*(1), 33-48.

Yukl, G. A. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives*, *26*(4), 66-85.

Yukl, G. A. (2013). *Leadership in Organizations* (8th). Pearson.

Yukl, G. A., Mahsud, R., Prussia, G., & Hassan, S. (2019). Effectiveness of broad and specific leadership behaviors. *Personnel Review, 48*(3), 774-783.

Zehnter, M. K., Olsen, J., & Kirchler, E. (2018). Obituaries of female and male leaders from 1974 to 2016 suggest change in descriptive but stability of prescriptive gender stereotypes. *Frontiers in Psychology*, *9*, 2286.