Not All Leaders Are Perceived Equal: The Interaction between Leader Gender, Perceiver Gender, and Emotion Suppression on Leader Ratings

Elsheba K. Abraham

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Psychology

Neil M. A. Hauenstein, Co-Chair

Roseanne J. Foti, Co-Chair

Daniel J. Beal

Ivan Hernandez

May 12, 2021
Blacksburg, VA

Keywords: leader judgments, leadership perceptions, gender bias, emotion suppression, ego depletion, repeated measures

Not All Leaders Are Perceived Equal: The Interaction between Leader Gender, Perceiver Gender, and Emotion Suppression on Leader Ratings

Elsheba K. Abraham

ABSTRACT

Females continue to be underrepresented in leadership despite research demonstrating that leadership effectiveness does not vary by leader gender (Paustian-Underdahl et al., 2014). The current study examines the gender bias in leadership through the lens of leadership perceptions and evaluations; in particular, how perceivers' ratings of a leader would change as a function of the leader's gender. Leadership judgments are based on the leader prototype activated in the perceiver and how consistent/inconsistent the leader is perceived to be with the activated prototype (Lord et al., 2001). Due to the mismatch between the communal-oriented female gender stereotype and agentic-oriented expectations of a successful leader (Eagly & Karau, 2002), it was expected that the female leader would be rated more negatively than the male leader. Furthermore, the perceiver's gender and prior engagement in emotion suppression are investigated as two additional factors that could bias information processing when evaluating leaders. Male perceivers, who tend to hold a stronger masculine understanding of leadership (Koenig et al., 2011), were expected to evaluate the female leader more harshly than the male leader. Additionally, those depleted of their finite self-regulatory resources due to prior emotion suppression (i.e. being in a state of ego depletion; Baumeister et al., 1998) were predicted to rely more heavily on their stereotypes when making subsequent judgments; hence, ego-depleted individuals would demonstrate more bias in their ratings of the female leader relative to the male leader.

In the current study, participants were randomly assigned to an emotion suppression or no suppression condition as they watched funny clips from the comedy series “The Office". Then, they watched four business videos featuring a leader and three business managers. Participants were also randomly assigned to one of the two versions of the business videos portraying either a male or female leader. Leadership perception and leader effectiveness ratings were collected after each of the four business videos, and leader competence and leader warmth ratings were measured once after all four videos. Additionally, behavior recognition accuracy of agentic and communal leadership behaviors that were displayed in the four business videos was assessed.

Contrary to expectations, the study findings demonstrate a dominant female leader effect; the female leader was evaluated more favorably than the male leader on all four leader judgments. This was observed both within the repeated measures and overall leadership ratings. An ego depletion effect was also observed; ego-depleted individuals showed lower accuracy in behavior recognition ratings and more leniency in leader warmth ratings. Furthermore, ego-depleted individuals showed less discernment by giving higher leader effectiveness ratings over time compared to non-ego-depleted individuals. Perceiver gender did not meaningfully affect leadership judgments. The unexpected pattern of bias in favor of the female leader instead of against her suggests that the nature of gender and leader stereotypes may be changing; the incongruence between the female stereotype and leader expectations may be decreasing, leading to more favorable evaluations of the female leader by both male and female perceivers. Moreover, the ability to provide fair and accurate judgments of leader effectiveness is reduced when depleted. Limitations and future research directions are discussed.

Not All Leaders Are Perceived Equal: The Interaction between Leader Gender, Perceiver Gender, and Emotion Suppression on Leader Ratings

Elsheba K. Abraham

GENERAL AUDIENCE ABSTRACT

The gender gap persists in leadership; although leader effectiveness has not been found to vary by the leader's gender, female leaders tend to be perceived and evaluated more negatively than male leaders. One reason for this is the mismatch between societal expectations for how women are ideally expected to behave and the expectations associated with a successful leader. In this study, gender bias in leader judgments and behavior recognition accuracy is examined by a leader's gender. Additionally, the perceiver's gender and prior engagement in emotion suppression are studied as two additional factors that can influence bias in leader ratings. Study findings demonstrate an unexpected but dominant female leader effect, where the female leader was perceived as more leader-like and rated more effective, more competent, and warmer than the male leader by both male and female perceivers. The amount of self-regulatory resources available also affected subsequent processing capabilities; those who suppressed their emotions and were depleted of their self-regulatory resources were less accurate in their behavior recognition ratings and were more lenient in their leader warmth ratings. Future research should explore if and how the nature of gender and leader stereotypes are changing, as evaluations of female leaders may not be as negatively-biased as it was previously.

Acknowledgements

This dissertation is the culmination of my 5-year journey at Virginia Tech. I have grown exponentially both professionally and personally, and would not have made it without the support of many people. First and foremost, I'd like to thank my dissertation committee, Dr. Neil Hauenstein, Dr. Roseanne Foti, Dr. Daniel Beal, Dr. Ivan Hernandez, and Dr. Charles Calderwood for your advice and expertise throughout the whole dissertation and prelim process. I have learned so much from all of you. In particular, I would like to thank my advisor Neil for his consistent guidance over the course of my PhD. Neil, I couldn't have asked for a better mentor – I am the researcher I am today because of you.

Next, I would like to thank my parents, KJ Abraham and Mary John, for their unwavering support. Pa and Ma, thank you for always being there for me no matter the time difference, for all the phone calls, and for your prayers. To my siblings, Benjamin Jon and Hanna Abraham, thank you for being my cheerleaders and for always keeping me in the loop with what was going on back home – it made it easier to do this PhD despite being so far away. I love you all dearly.

To my friends who are my chosen family, thank you for your friendship, support, and laughs throughout the years. To my Blacksburg friends: Annie Patrick, Sierrah Travis, Kory Wait, and Vivian Zagarese among many others, I treasure all the meals, movie nights, and hikes we've had – thank you for making my graduate school experience so memorable. To my SCIP lab friends from Day 1: Jessica Gladfelter and Manasia Sturdivant, I loved doing research with both of you and would not trade the hours we spent working/bonding in the Williams computer labs for anything. To my VT IO friends in cohorts before me: Maureen "Mo" McCusker, Yashna Shah, and Bryan Acton: thank you for so generously sharing your time, experience, and advice - you made my transition into graduate school so much easier. To my best friends from all other

phases of my life: Stella Yeoum, Xier Lu, Amanda Anuja, Sonia Grace, and Suki Prem Kumar – thank you for your constant support and consistent communication; you are my girlfriends until the end. Finally, I would like to thank God for my success. This journey has been a challenging one, but I drew upon Him for strength, peace, and confidence throughout this process.

I am proud of all I've accomplished so far with the support of so many, and am looking forward to what is to come.

Table of Contents

List of Figures

List of Tables

Not All Leaders Are Perceived Equal: The Interaction between Leader Gender, Perceiver Gender, and Emotion Suppression on Leader Ratings

Although the representation of females in leadership positions has been increasing (e.g., an 11.3% increase in female parliamentarians from 1995; UN Women, 2019), the gender gap in leadership continues to persist. For example, only 6.2% of CEOs in S&P 500 companies are women (Catalyst, 2020). Given overall leader effectiveness does not vary across gender (Eagly et al., 1995; Paustian-Underdahl et al., 2014), the underrepresentation of female leaders cannot be attributed to a systematic lack of competence on their part. Instead, this gender gap reflects a complex and unique set of challenges that women face, both while vying for leadership opportunities and after successfully rising to leadership positions. Aside from structural barriers (e.g., work hours that conflict with familial obligations, reduced access to networking opportunities; Carli & Eagly, 2018), there are also psychological factors that contribute to biased perceptions of female leadership. Research on differences in leadership perceptions across gender spans several decades, with a majority of studies focused on understanding cognitive schemas that guide perceptions and relatedly, leadership judgments (e.g., Eden & Leviathan, 1975, Lord et al., 1984).

Considering leadership is a socio-perceptual and relational process that unfolds between leaders and their followers (McCusker et al., 2018), leadership does not occur in isolation; hence, contextual factors that influence a follower's perception of a leader must be taken into account. The current study examines the effects of emotion regulation on the subsequent reliance on cognitive schemas when judging leaders. Not only is emotion regulation a mentally taxing task that is often employed in laboratory settings (e.g., Gross & Levenson, 1993; Richards & Gross, 2000), but it is also common in the workplace as individuals try to display certain emotions and

hide others in their daily social interactions (Cote, 2005). Emotion suppression is a specific emotion regulation strategy that can be employed to mask the behavioral expression of an emotion experienced (Gross & Levenson, 1993). Individuals may suppress their emotions when they find themselves in unexpected situations (e.g., being reprimanded by a leader in front of other coworkers) but still need to abide by unofficial workplace emotional display rules to maintain a face of professionalism. Effortful emotion suppression leads to a state of "ego depletion" in which self-regulation ability is impaired (Baumeister et al., 1998). Emotion suppression requires self-control, and the capacity to carry out self-control effectively is finite (Baumeister et al., 1998), As a result, individuals in a state of ego depletion are likely to rely on heuristics when making sense of social interactions.

The primary goal of this study is to examine the extent to which ego depletion through emotion suppression leads to gender bias in terms of four different types of judgments: the extent to which an individual is seen as a leader (i.e., *leadership perceptions*), leader effectiveness, leader competence and warmth. In addition, the current study will examine differential recognition accuracy of male/female leader behaviors as a function of emotion suppression. The fundamental argument is that when observing a leader in action, ego-depleted individuals, relative to non-ego-depleted individuals, will engage in greater heuristic processing due to increased reliance on both leader and gender schemas; this results in biased leadership judgments and biased recognition accuracy. Furthermore, the extent to which the *perceiver's* gender moderates how gender bias manifests across all five dependent variables is also examined, as males hold stronger biases against female leaders than females (Carli & Eagly, 2018).

Finally, a majority of gender bias leadership research is carried out using self-report leader evaluations administered at one time-point. The current research employs a repeated

measures approach for both leadership perceptions and leader effectiveness in the attempt to demonstrate that when perceivers are ego-depleted, gender bias manifests early in the judgment process, and these gender-biased perceptions are resistant to change despite increasing displays of leader behaviors by a female leader.

Bias against Female Leaders

Females often deal with the negative stereotypes and discrimination that come with being a woman in a position of authority (Carli & Eagly, 2018; Heilman & Caleo, 2018). Although there is a plethora of research on this topic, some of the prominent theories of focus in this study include leader categorization and connectionist theory (Lord et al., 1984; 2001), implicit leadership theory (ILT; Eden & Leviathan, 1975; Offermann et al., 1994) and role congruity theory (Eagly & Karau, 2002). These theories highlight how information processing of leader behaviors differs by gender, thereby influencing ratings of female leaders compared to male leaders.

Gender stereotypes evolved from societal norms within our culture, where females typically adopt the role of the caregiver while males are providers to the family. Women are expected to be more communal (i.e. warm, supportive, helpful) while men are associated with more agentic qualities (i.e. dominant, competent, assertive) (Carli & Eagly, 2018). According to the leader categorization theory (Lord et al., 1984), individuals have cognitive classifications of prototypical traits that are associated with leadership. Leaders who display these traits are more easily perceived as leaders compared to leaders who do not. Taking into consideration gender stereotypes and leader prototypical behavior, the agentic traits of men are more similar to beliefs of how leaders should behave (Schein, 1973). This makes it easier to see men as leaders, as their leadership behaviors are more consistent with leadership attributes typically present in a

perceiver's leadership prototype. In contrast, female leadership behaviors are inconsistent with the typical leadership prototype, resulting in prejudice against female leaders who are considered out of place in a "masculine" role (Kark & Eagly, 2010).

Due to the incongruity between expectations of female behavior with the expected abilities associated with a successful leader, there is a negative bias against female leaders (Eagly & Karau, 2002). Additionally, the status incongruity hypothesis proposes that female leaders are evaluated negatively and suffer backlash because they violate the traditional gender hierarchy (Rudman et al., 2012). Women are perceived to hold a lower social status and are expected to adopt more supportive roles in the group, while men are not questioned when they assume positions of power as it is afforded to them by their gender privilege (Ridgeway, 2001). Furthermore, female leaders face a double-bind; if they behave communally, they are not perceived as competent whereas if they behave agentically, they are disliked for lacking warmth and are evaluated more harshly (Carli & Eagly, 2018; Eagly & Carli, 2007). Not surprisingly, female leader behaviors are evaluated more negatively than males when they are more autocratic or directive (Eagly et al., 1992).

Perceiver Gender

Gender stereotypes stem from our collective societal expectation of how men and women should behave (Eagly & Karau, 2002). Additionally, research has shown that agentic traits associated with an ideal leader have remained stable over the past several decades (Offermann & Coats, 2018). However, even if everyone may hold the same knowledge of stereotypes, there is a distinction between knowing stereotypes and using them to guide judgment (Devine & Elliot, 1995). Thus, evaluations of female leaders can change according to *who* the perceiver is. The perceiver's gender may influence how leadership information is processed, which affects

subsequent leader judgments (Hogue & Lord, 2007; Johnson et al., 2008; Lord et al., 2001). Moreover, men tend to endorse masculine characteristics as part of leadership more strongly compared to females (Koenig et al., 2011; Offermann & Coats, 2018). Taken together, this suggests leader judgments vary by perceiver gender, and that male perceivers are more likely to rate female leaders as less effective than female perceivers.

Emotion Regulation at the Workplace

Due to social norms regarding the appropriate display of emotions, emotion regulation occurs in all types of social interactions including those occurring at work. While there are various prominent theories regarding emotions at the workplace (i.e. emotional labor; Hochschild, 1983; Grandey & Melloy, 2017; emotional intelligence; Mayer & Salovey, 1997), the framework used for the current study is Gross' (1998a;1998b) process model of emotion regulation (ER). The ER model provides a broader understanding of how individuals practice ER in a variety of social situations as they achieve a range of personal and functional goals (Perrone & Vickers, 2004). Emotion suppression, which is one of the five ER strategies included in the model, is the focus of this study. It is a response-focused ER strategy whereby an individual suppresses any facial or behavioral expression of an emotion *after* it is experienced.

Emotion Suppression and Ego Depletion

Emotion suppression requires self-control (or alternatively, self-regulation), which is defined as the capacity for altering one's responses and suppressing impulses to achieve standards, ideals, values, morals, and social expectations that are necessary to pursue long-term goals (Baumeister et al., 2007). Thus, emotion suppression is sustained by self-regulatory resources and enables an individual to achieve a regulatory goal in social interactions.

Self-control Strength Model. An individual's capacity for self-control is typically studied using the self-control strength model (Baumeister et al., 1998; Muraven et al., 1998). There are several core features that characterize this model: first, self-control is sustained by a limited resource similar to energy. Next, these regulatory resources are limited and are depleted by exertions of self-control. Taxing regulatory resources in one act of self-control depletes the available resources to sustain subsequent acts of self-control; this leads to a state of *ego depletion* whereby an individual has a decreased willingness or ability to self-regulate (Baumeister et al., 1998). Although the term "depletion" is used, it refers to any decrease in resources instead of complete exhaustion of resources (Baumeister and Vohs, 2016b). Finally, akin to a muscle that gets strained after exercise, the limited energy resource remains temporarily depleted after the act of self-control and needs a recovery period to get back to full capacity (Baumeister & Vohs, 2016b; Muraven et al., 1998).

Effects of Ego Depletion on Heuristic Processing. Although there have been recent concerns about the robustness of the ego depletion effect (e.g., Carter et al., 2015; Friese et. al., 2019; Hagger et al., 2016), studying ego depletion through emotion-based tasks has been argued to be a promising method for exploring this phenomenon (Baumeister & Vohs, 2016a). Many studies have examined ego depletion through emotion suppression, as affect regulation was included as one of the four main spheres of regulation in Baumeister and colleagues' original conceptualization of self-control (Baumeister et al., 1998; Muraven et. al., 1998). Emotion suppression is often used as a method to manipulate resource availability (e.g., Hofmann et al., 2012; Muraven, 2008; Vohs et al., 2005), leading to a state of ego depletion in which individuals exercise reduced top-down executive control (Baumeister & Vohs, 2016b).

In the context of stereotype-processing, ego-depleted individuals are more likely to rely on their stereotypes to make judgments instead of expending resources to engage in effortful thought processes. Stereotypes serve as cognitive heuristics as they are a less resource-intensive method for processing social contexts as a whole, instead of focusing on individuating pieces of information (McGarty et al., 2002). This aligns with past research that has demonstrated that ego-depleted individuals are more likely to express more prejudice against homosexuals (Gailliot et al., 2007), Blacks (Richeson & Shelton, 2003), and skinheads (Gordijn et al., 2004). If an individual is ego-depleted through emotion suppression, diagnostic behaviors are less accurately processed; in situations where individuating information is lacking or information is ambiguous, social judgments' of perceivers are guided by their gender stereotypes (Dunning & Sherman, 1997; Kunda & Thagard, 1996; Hamilton et al., 1990; Von Hippel et al., 1995). Relative to non-ego-depleted perceivers, perceivers in an ego-depleted state due to emotion suppression are more likely to be guided by their initial categorizations of the observed leader. Female leadership behaviors will be perceived as more inconsistent with the perceiver's leadership schema compared to male leadership behaviors; consequently, ratings of the female leader will be more negatively biased than ratings of the male leader.

Current Study

There is increasing recognition that studying emotion regulation and leadership in conjunction with one another can provide a greater understanding of both of these topics (Humphrey et al., 2016). However, a majority of studies focus on the influence that emotion regulation by leaders and their subsequent emotional displays can have on followers' perceptions of the leader's effectiveness (e.g., Dasborough & Ashkanasy, 2002; Eberly & Fong, 2013; Jordan & Lindenbaum, 2015; Madera & Smith, 2009; Schaubroeck & Shao, 2012; Trichas et al., 2017).

In contrast, this study focuses on how emotion regulation by the *follower* (or perceiver) affects subsequent judgments (and recognition accuracy) of a leader, and if the gender bias in these information processing outcomes varies by perceiver gender. From an ego depletion perspective, individuals who suppress their emotions become ego-depleted and are more likely to rely on leader and gender stereotypes when observing leaders. As a result, ego-depleted individuals are less likely to perceive and judge leaders based on leadership behaviors that actually occur; instead, they rely more on leader/gender schemas because they have reduced processing resources available. Due to the mismatch between the female and leader stereotypes, female leaders would be more likely to be evaluated with more bias, and this pattern of bias may be more pronounced in male than female perceivers.

Repeated Measures Design

Cross-sectional self-report assessments are commonly utilized in research due to their availability and convenience of administration. However, to provide a more robust test of the processes that underlie gender bias in leadership perceptions and ratings of effectiveness, a repeated measures approach is used in this study. Collection of repeated measures affords the study of change occurring across measurement occasions and how the constructs relate to one another within an individual (Gabriel et al., 2017). It also allows for a more immediate assessment of a construct after an event has occurred (Beal & Gabriel, 2018), hence giving a deeper insight into the influence of events within a particular measurement interval on subsequent ratings.

Literature Review

Past research has demonstrated that overall, leader effectiveness does not vary by leader gender (Eagly et al., 1995; Paustian-Underdahl et al., 2014). Nonetheless, systematic differences are apparent when comparing the judgments of female leaders compared to male leaders. Meta-analyses have shown female leaders are rated less favorably when in stereotypically masculine leadership roles and work in male-dominated organizations (Eagly et al., 1992; Paustian-Underdahl et al., 2014). Leadership ratings also vary based on how leadership as a construct is assessed; males are preferred over females when leadership is measured by perceptions of competence and ability (Eagly et al., 1992, 1995). Taken together, this suggests that the lower ratings that female leaders receive are not simply due to them being objectively less effective; instead, the assessment of a leader's effectiveness is influenced by many factors.

Leader Categorization Theory

Within leader effectiveness research, a subordinates' perceived effectiveness or approval of a leader is often used as the dependent variable/criterion measure (Kaiser et al., 2008). Individuals' leadership ratings are based on information that they process from their surroundings and their interactions with the leader (Lord et al., 1984; 2001). While subjective leader effectiveness ratings directly assess a leader's ability to guide followers toward goal achievement (Eagly et al., 1995), these ratings are susceptible to personal biases that affect how individuals perceive and subsequently assess leaders.

The conceptual foundation for research on leadership perceptions is leadership categorization theory (Lord et al., 1984; Lord & Maher, 1993), an extension of Rosch's (1978) categorization theory. Leadership categorization theorists posit that individuals have cognitive representations or mental schemas of leader categories to enable efficient processing of

information in the external environment. The categories consist of a variety of leader attributes, and each category is represented by prototypes. As prototypes reflect the most representative items of a category (Rosch, 1978), leadership prototypes provide a standard for how leaders are expected to behave (Lord et al., 1984). Leadership perception is understood as the result of a two-stage matching process: first, when individuals view (or are primed about) a stimulus target, the most relevant leadership prototype is activated; then, the observed leader behaviors are compared with the activated prototype pattern (Lord et al., 2001).

Intersectionality of Gender and Leadership

Although there is a great deal of empirical support for leadership categorization theory, it overlooks the contextual influences that affect the leadership prototype activation process. Leadership is typically defined without gender expectations (e.g., a leader-follower interaction process where the leader guides followers to achieve a shared accomplishment; Yammarino, 2013), but the reality is that effective leadership beliefs carry a masculine construal (Heilman et al., 1989; Powell & Butterfield, 1989; Schein, 1973). To handle this complexity, Lord et al. (2001) proposed a connectionist-based model to expound on the flexible and dynamic nature of the leadership categorization process. In the connectionist model, leadership attributes within an individual's leadership prototype are activated differently according to each situation. The resulting prototype network depends both on the activation of bottom-up behavioral inputs observed from the leader and the constraints placed by higher-level contextual factors that can be both external (e.g., leader gender, organization type) and internal (e.g., affect, self-schema) to the perceiver.

Female versus Male Leadership Prototype

ILTs represent the understanding of the collection of attributes that an ideal leader is believed to possess. Offermann, Kennedy, and Wirtz (1994) identified eight factors represented in a typical leader prototype including, sensitivity, dedication, tyranny, charisma, attractiveness, masculinity, intelligence, and strength. They found that the factor structure replicated across gender for both students and working adults. Recent research has generally replicated these findings; Epitropaki and Martin (2004) found six factors, where charisma and strength were collapsed into the dynamism factor, but no evidence of an attractiveness factor. In 2018, Offermann and Coats replicated the Offermann et al. (1994) study and found evidence for the original eight factors, plus a ninth factor identified as creativity.

Across both Offermann et. al's (1994; 2018) studies and the Epitropaki and Martin (2004) study, masculinity was identified as a common trait in individuals' ILTs. In addition to masculinity, both strength and tyranny have been associated with male leadership whereas only sensitivity has been connected to female leadership (Johnson et al., 2008). Moreover, Offermann and Coats (2018) found that both male and female participants used male-oriented but not female-oriented terms when asked to describe the characteristics of their ideal leader. It is clear that leader prototypes contain greater male than female attributes. A meta-analysis conducted by Koenig et al. (2011) to review the masculinity/femininity of leader stereotypes underscores this fact; across three different leadership paradigms, leader stereotypes are considered more masculine overall and leaders are believed to show more agentic than communal behaviors.

According to Eagly & Karau's (2002) role congruity theory, it is difficult to perceive females as leaders as there is a mismatch between the communal nature of the female stereotype and the agentic expectations for a successful leader. Gender stereotypes are beliefs about the

traits and attributes thought to characterize men and women (Scott & Brown, 2006), and they are shared by both genders (Ellemers, 2018). Stereotype content can be divided between two core dimensions: *warmth* that reflects a desire to know others' intents, and *competence* that represents the ability to pursue a task (Fiske et al., 2002). Closely associated with gender stereotypes is the concept of gender roles, which consists of both descriptive (what they *actually* do) and prescriptive (what they *should ideally* do) expectations of how males/females behave (Eagly & Karau, 2002). Men are thought to reflect competence as they are agentic and task-oriented, displaying attributes such as independence and assertiveness. In contrast, women are perceived to represent warmth and are expected to be communal, interpersonally-oriented, helpful, and nurturing (Broverman et al., 1972; Eagly & Karau, 2002). In comparison to women, it is easier to process men as leaders because male and leader role expectations are more aligned (Eagly & Karau, 2002; Eagly, 2007; Heilman, 2001, Heilman et al., 1989).

The strong similarity in how males and leaders are viewed has been established by multiple research findings. For example, in Schein's (1973) seminal study, male middle managers were asked to rate successful managers, men in general, or women in general on a list of 92 attributes. There was a greater overlap between the characteristics of successful managers and males compared to females. Both successful managers and men were perceived to be emotionally stable, direct, and possessing leadership ability, amongst many other attributes. This "think manager, think male" mindset was replicated years later; compared to women in general, men in general were rated more similar to a successful manager (Heilman et al., 1989), and male, instead of female, executives were rated more similar to executives in general (Sczesny et al., 2004). Additionally, good managers were described primarily using masculine characteristics

(Powell et al., 2002); this preference was observed in another study, regardless of the employee's gender or the gender of his/her manager (Stoker et al., 2012).

In addition to the incongruity between the female and leader stereotype, female leaders are often placed in a double bind as they deal with the opposing agentic versus communal demands of the leader and gender role respectively. Females are not rewarded when they are helpful and carry out other communal behaviors as expected but are penalized when they do not engage in these behaviors (Scott & Brown, 2006; Heilman & Chen, 2005; Phelan & Rudman, 2010). Female leaders are also restricted by the communal expectations attached to their leadership; a meta-analysis by Smith et al. (2013) found that while women yield influence only when carrying out communal behaviors, men are influential regardless if they are behaving communally or agentically. Moreover, female leaders are expected to balance their communal tendencies with agentic behaviors; Johnson et al. (2008) found that participants perceived a male leader as effective if he showed strength but an effective female leader had to show strength *and* sensitivity. Thus, not only are female leaders challenged with a smaller conceptual overlap in expectations tied to their gender and leader role, but they also are faced with more constrained prescriptions of how they should behave as leaders.

Perceiver Gender and Leadership Perceptions

Males and females share similar leader attributes in their ILTs (Offermann & Coats, 2018) and share knowledge of the descriptive stereotypes of gender roles (Ellemers, 2018; Haines et al., 2016). While female and male raters tend to agree on the perceived characteristics possessed by a male leader (e.g. Deal & Stevenson, 1998; Johnson et al., 2008; Powell & Butterfield, 1989; Sczesny et al., 2004) and provide similar ratings of the male leader (e.g. Johnson et al., 2008; Sczesny, 2003; Stoker et al., 2012), they vary when it relates to female

leaders. Although there is an overall negative bias against female leaders, men tend to be harsher evaluators than women (Carli & Eagly, 2018).

A perceiver's gender may moderate leader judgments across male and female leaders due to two distinct but related reasons. First, a perceiver's gender influences the accessibility of attributes within his/her leadership prototype (Deal & Stevenson, 1998; Lord et al., 2001; Nye & Forsyth, 1991). An individual's self-schema is influenced by past experiences and affects how leadership is perceived (Hogue & Lord, 2007). Accordingly, communal leader behaviors such as sensitivity commonly associated with women may be more frequently activated in a female perceiver's leadership prototype while agentic traits such as dominance may be more central in a male perceiver's leadership prototype (Lord et al., 2001). Empirical evidence demonstrates a pattern of preference for communal versus agentic behaviors by gender; when asked to estimate how likely male/female middle managers carried out certain leader behaviors, female managers rated female leaders higher on communal behaviors such as supporting and rewarding while male managers rated male leaders higher on agentic behaviors such as problem-solving and intellectual stimulation (Martell & DeSmet, 2001). Similarly, Powell et al. (2008) found that female raters viewed female leaders who carried out transformational leadership more favorably, a leadership style characterized by various communal behaviors such as mentoring and motivating followers. Johnson et al. (2008) found that feminine, but not masculine, individuals valued sensitivity when evaluating the effectiveness of female leaders. Hence, female perceivers tend to heavily weigh communal/warmth behaviors when evaluating female leaders whereas male perceivers tend to heavily weigh agentic/competent behaviors when evaluating male leaders.

In addition, men hold a stronger masculine understanding of leadership than women (Koenig et al., 2011). This carries implications on how males perceive female leaders, as how much an individual associates communality with women (and agency with men) can negatively affect ratings of agentic females (Rudman & Glick, 2001). Men are more likely than women to feel threatened by female leaders and to display higher levels of prejudice toward women in authority (Rudman & Kilianski, 2000; Rudman et al., 2012). Furthermore, prescriptive stereotyping has been found to lead men (but not women) to show bias against females in a masculine role (Gill, 2004). In comparison to female raters, male raters are more likely to describe female leaders negatively (Deal & Stevenson, 1998) and perceive a greater stereotype overlap between males and a successful good manager (Ryan et al., 2011; Stoker et al., 2012). The "think manager-think male" mindset persists in men, who are less likely to realize the challenges that gender stereotyping brings to female leaders compared to women (Schein, 2001).

As a result, male raters manifest greater bias in ratings of female leaders than female raters (Eagly et al., 1992; Ayman et al., 2009; Kwon & Milgrom, 2010). Comparing across two meta-analyses of leader effectiveness and gender, while Eagly et. al. (1995) initially failed to find a pro-female leader bias in female raters, Paustian-Underdahl et. al. (2014) found that female leaders were evaluated more favorably than male leaders when there was a higher percentage of females in the rater group; however, the female leader preference disappeared as the group became gender-balanced. More recently, Gloor et al. (2020) demonstrated that female leaders are viewed as less prototypical leaders than male leaders when there was a higher percentage of male raters in the group. While there have been many studies that have found a perceiver gender effect on negative evaluations of female leaders, there are several others that have not found gender differences (e.g., Brescoll, 2011, Okimoto & Brescoll, 2010; Rudman et

al., 2012) or the opposite effect, such that women evaluate other female leaders more negatively (e.g., Parks-Stamm et al., 2008; Warning & Buchanan, 2009). Although these findings appear contradictory, a closer look reveals that perceiver gender differences on assessments of female leader effectiveness depend on the motivation for devaluing a female leader (Vial et al., 2016). Women are more likely to show bias against female leaders when they also hold some authority, and self-perceptions of their competence are threatened by the female leader (Park-Stamm et al., 2008).

Emotion Regulation and Ego Depletion

In comparison to daily interactions, perceivers in an experimental context are faced with fewer distractions that detract from their ability to process information in an unbiased manner. Hence, in order to produce the gender biases discussed in previous sections, it is necessary to manipulate the perceiver's ability to engage in effortful processing thereby increasing their reliance on schematic/stereotypic processing. Across the various possible methods for taxing mental processing in laboratory studies, ego depletion induced through an emotion suppression manipulation is used for two reasons: first, emotion suppression is a resource-demanding process (Baumeister et al., 1998; Gross, 1998a). Past research also has shown that emotion suppression increases reliance on stereotypes when making subsequent judgments (Burns et al., 2008; Muraven, 2008). For example, Abraham and Hauenstein (2020) found that participants actively suppressing emotions showed evidence of greater bigotry against Blacks in subsequent judgments compared to participants not instructed to suppress their emotions.

Second, as emotion regulation is common in the work context, it is a likely contributing factor to gender bias in leadership judgments. Emotion regulation is inherent in interpersonal interactions and is important in almost all jobs (Beal & Trougakos, 2013). Across work

interactions, individuals are often motivated to regulate their emotions for a variety of reasons, including abiding by organizational display rules (Diefendorff & Richard, 2003), aligning personal perceptions of emotional expression with the perceived display rule (Diefendorff & Grosserand, 2003) or fulfilling a higher-order personal goal such as being happy (Niven, Totterdell, et al., 2013). It is common for individuals to engage in emotion regulation as they interact with their coworkers and supervisors (e.g. Kafetsios et al., 2012; 2014; Gabriel et al., 2020; Niven, Sprigg, et al., 2013; Shumski Thomas et al., 2018; Totterdell et. al., 2012). Given its prevalence in the workplace, research on emotion regulation and leadership has received increased attention recently (Humphrey et al., 2016) and observational studies have shown that greater emotion regulation negatively affects both perceptions of meeting effectiveness (Shumski Thomas et al., 2018) and leadership perceptions (Glaso & Einarsen, 2008; Xu et al., 2014).

Emotion Suppression

One of the primary frameworks used to study emotion regulation at the workplace includes Gross' (1998a; 1998b) process model of emotion regulation (ER). ER is the process "by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions" (Gross, 1998a, p. 275). ER happens in a broad range of settings both alone or in the presence of others (Troth et al., 2018). Furthermore, regulation is not necessarily tied to a particular professional display rule (e.g. customer service representatives who need to smile and adopt a friendly demeanor) but instead applies more generally to both formal and informal workplace interactions. The ER process model consists of 5 emotion regulation strategies divided into antecedent-focused regulation (i.e. situation selection, situation modification, attention deployment, cognitive reappraisal) carried out *before* the emotion is experienced, and response-focused regulation (e.g., emotion suppression) of the emotion *after* it

has been experienced (Gross 1998a;1998b). Emotion suppression is the focus of this study and is defined as the conscious inhibition of an emotional response that results in a discrepancy between the emotion being experienced and what is expressed (Gross & Levenson, 1993). Often in the workplace, individuals cannot predict their interactions and the corresponding emotions generated from that situation (Beal & Trougakos, 2013). Hence, emotion suppression is a relevant ER strategy to study within the workplace context, as it may be adopted at the moment in order to control the expression of emotions being experienced.

Ego Depletion. In the broader social psychology literature, emotion suppression is often studied within the context of ego depletion research in that suppression requires self-control (Baumeister et al., 1998; Muraven et al., 1998). Since self-control is sustained by limited resources, suppressing emotions in one task leads to decreased resources for self-control in a subsequent task. This is the underlying logic for ego depletion research. Labeled as both a *dual-task paradigm* (Hagger et al., 2010) and a *sequential-task paradigm* (Friese et al., 2019), participants in the experimental condition have to carry out two self-control tasks at time 1 (T1) and time 2 (T2); in contrast, participants in the control condition also perform two consecutive tasks but only perform the self-control task at T2. Since all actions of self-control are sustained by a common but limited resource (Muraven et al., 1998), the ego depletion effect is demonstrated if performance on the self-control task at T2 is significantly worse in the experimental group than the control group.

In ego depletion studies that utilize emotion suppression, participants are typically instructed to suppress their emotions while watching a film that creates a strong emotional reactance before completing the T2 task. For example, Muraven et al. (1998) found that compared to the control group, participants who suppressed their emotional reactions while

watching a 10-minute funny video by comedian Robin Williams performed significantly worse on an anagram task. In another study, amongst participants who were high in their motivation to control their prejudice, those who suppressed their emotions while watching an 8-minute comedy routine by Ellen DeGeneres rated Blacks more negatively than those who did not suppress their laughs (Muraven, 2008).

Summary of Conceptual Variables

Leadership categorization theory is the foundation of understanding leadership perceptions, but to study gender bias in leadership requires the integration of gender stereotypes given the assumed agentic-oriented leader characteristics of males are so different from the communal-oriented leader characteristics of females. Gender of the perceiver complicates the study of gender bias in leadership perceptions because the evidence suggests that the ILTs of females are similar, yet different from males. Finally, social-cognitive laboratory studies require a mechanism to facilitate heuristic processing. Ego depletion through emotional suppression was chosen as that mechanism because emotion suppression taxes executive functioning thereby facilitating heuristic processing. In addition, emotion suppression was chosen because emotion regulation is required during most social interactions in the workplace.

Processing Outcomes

The current study focuses on two sets of processing outcomes, gender bias in leadership judgments and gender bias in recognition accuracy. However, prior to reviewing the logic underlying the hypotheses for these two outcome variables, it is necessary to explicate "dual-processing" models of person perception.

Dual Processing Models

In dual-processing models of person perception (e.g., Brewer, 1988; Fiske & Neuberg, 1990), a perceiver categorizes a target into the relevant schema almost instantaneously without much effortful processing; this then guides subsequent interpretation about the target (Fiske & Neuberg, 1990; Heilman, 1995). Related to that, encoding is the process by which observed stimuli are translated into mental representations or information that can be accessed at a later time (von Hippel et al., 1995). Cognitive schemas are critical to the encoding process, as they enable individuals to match the incoming information with existing knowledge within these schemas and encode it into their long-term memory (Lord & Maher, 1993). Greater effort expended during the encoding process leads to greater accessibility of that encoded information from memory (Hastie & Kumar, 1979). In order to understand the social cognitive implications once a stimulus target is categorized as a leader, it is important to parse the differences that leadership prototype consistent and inconsistent information can have on perceiver ratings of the leader.

Processing of schema-consistent behavior. Encoding of behavioral actions congruent with an initial categorization requires fewer cognitive resources, as perceivers rely on the leader prototype to make general sense of a situation without processing specific details (Hastie & Kumar, 1979; Lord et al., 1984; Sherman et al., 2000). This leads to reinforcement of the initial leader categorization without creating strong memory traces for the encoding/storage of actual behaviors (Hastie & Kumar, 1979; Sherman et al., 2000). Furthermore, the lack of strong memory traces for schema-consistent behaviors results in false-positive memories of leader behaviors that the target leader did not perform (Hansbrough et al., 2020; Lord et al. 1984; Shondrick & Lord, 2010). Ultimately, perceivers' leadership judgments are skewed toward

increasing their confidence that the stimulus target is a leader (Hamilton et al., 1990; Heilman, 1995; Lord et al., 1984).

Processing of schema-inconsistent behavior. In contrast, schema-inconsistent information requires more effortful processing because the salience of the atypical behavior forces greater attentional effort to make sense of this information (Hastie & Kumar, 1979; Sherman et al., 2000). The effortful processing of schema-inconsistent behaviors creates stronger memory traces (Hastie & Kumar, 1979); as such, all other things being equal, perceivers are likely to have greater behavioral recognition accuracy for schema-inconsistent behaviors than schema-consistent behaviors (Foti & Hauenstein, 1993; Hastie & Kumar, 1979; Van Overwalle & Labiouse, 2004; Sherman et al., 2005). Although schema-inconsistent information is salient and more memorable, perceivers initially seek to discount the atypical behavior or make an external attribution about the cause of the behavior instead of changing the initial categorization (Crocker et al., 1983; Vonk, 1994). Moreover, perceivers are unlikely to alter their initial categorization of the target person, especially if the schema-inconsistent behaviors are infrequent. Thus, although schema-inconsistent behaviors require greater elaborative processing, the stronger memory traces for schema-inconsistent behaviors do not necessarily change initial perceptions of the target (Hamilton et al., 1990; Sherman et al., 2005). Individuals are more resistant to accepting schema-inconsistent information (Ditto & Lopez, 1992; Sherman et al., 2005), so even if inconsistent information is recognized, ratings still tend to bias towards schema-confirming responses (Foti & Hauenstein, 1993; Sherman et al., 2005).

Gender Bias in Leadership Judgments

Gender, along with other characteristics such as race and age, is used as a primary means of categorization (Ellemer, 2018; Brewer, 1988). As such, recognizing the target as male or

female then immediately cues gender stereotypes in the perceiver (Fiske & Neuberg, 1990; Scott & Brown, 2006). The primacy of gender stereotypes causes perceivers to activate networks associated with the leader's gender, thereby activating agentic traits (e.g., dominance) for male leaders versus communal traits (e.g., flexibility) for female leaders (See Figure 1, Hogue & Lord, 2007; Figure 5, Lord et al., 2001). Thus, a leader's gender can bias the perceiver's encoding and subsequent interpretation of leadership behaviors (Hogue & Lord, 2007).

Leader Gender Effects

Prior research has provided robust empirical support regarding the negative bias of leadership effectiveness ratings against female leaders (Eagly et al., 1992; Heilman et al., 1995). Building upon that, by tying in the discussion of how information processing of leadership behaviors is biased according to what is schema-consistent/inconsistent with the leader's gender, additional insight can be gained into the underlying processes that drive the differential ratings of male versus female leaders.

Perceivers generally default to using categories or schemas to guide judgment as this saves on encoding and memory demands; however, it occurs at the expense of evaluating each individual trait (Shondrick & Lord, 2010). In the context of leadership ratings, individuals rely on their prototypes that are activated at the preconscious level to guide their leadership perceptions (Lord et al., 2001). If the male/female leader behaves in a schema-consistent manner, the leader is likely to be judged within the boundaries of the leader and gender categorization that he/she has been placed in. As such, the processing of agentic behaviors by a male leader and the communal behaviors by a female leader involves low amounts of cognitive effort. The gender of the female leader is likely to play a boundary-constricting role in the categorization of her leadership; that is, she would be judged as a "good *female* leader" whereby the definition of

effective leadership is constrained by her gender. In contrast, as the male stereotype has greater overlap with the leader prototype, the male gender of the leader may not elicit salient boundary conditions for the definitions of leadership as compared to the female leader. As a result, he would be judged simply as a "good leader." Even though both the female and male leader can be recognized as good leaders, the female will still be rated as less effective than the male leader as her female gender limits the definition of effective leadership that she is associated with.

In terms of schema-inconsistent behavior, researchers have found that seeing a female leader activates the female leader prototype, resulting in the weaker encoding of agentic leader behaviors associated with a female compared to a male leader (Scott & Brown, 2006; Braun et al., 2018). Thus, agentic behaviors carried out by female leaders will be salient as perceivers engage in elaborative processing of that inconsistent information. The increased focus on her agentic behaviors elicits negative reactions by the perceivers (Rudman et al., 2012) and thereby can lead to perceptions of her being a "poor female leader."

The bottom line is that when gender intersects with leadership, behaviors that are schema-consistent/inconsistent vary as a function of leader gender. Communal behaviors are schema-consistent for females and agentic behaviors are schema-inconsistent for females. For males, agentic behaviors are schema-consistent, but it is less clear that communal behaviors are schema-inconsistent as evidence indicates the male leadership schema is more encompassing of all leader behaviors than the female leadership schema (Scott & Brown, 2006). Thus, even though both types of leadership behaviors are recognized for male leaders, an encoding bias exists when female leaders carry out leader behaviors that are inconsistent with their gender role expectations. If a female leader's leadership behaviors are not encoded, the perceiver will not

have information about her work potential in his/her leadership prototype; consequently, an accurate assessment of her competence as a leader is less likely (Lord & Maher, 1993).

In summary, there is greater congruence between the general expectations of behavior between males and leaders as compared to females and leaders; this influences the nodes that are activated in perceivers' leadership prototypes. As perceivers are guided by their activated gender-leader network more heavily than the actual behaviors manifested, female leaders are likely to receive more biased leadership ratings compared to male leaders.

Hypothesis 1: Gender bias will manifest in all leadership judgments such that a female leader will be:

1a. Rated as less *effective* overall than a male leader across the four assessments of leader effectiveness

1b. Perceived as less *leader-like* overall than a male leader across the four assessments of leadership perceptions

1c: Rated as less *competent* than a male leader in the competence ratings measured at the end of the experiment

1d: Rated as having greater *warmth* than a male leader in the warmth ratings measured at the end of the experiment

Leader Gender by Perceiver Gender Interaction

While a leader gender main effect is anticipated such that a female leader would be rated more negatively than a male leader, the pattern of effects is expected to vary as a function of the perceiver's gender. As mentioned previously, the accessibility of leadership attributes within a perceiver's schema differs according to the perceiver's gender and males tend to hold a stronger masculine understanding of leadership (Lord et al., 2001; Koenig et al., 2011). Hence, compared

to female perceivers, male perceivers likely activate different leadership prototypes and are more sensitized to incongruities between their male-oriented leader expectations and prescriptive beliefs regarding female behavior. This results in negatively biased ratings of female leaders from males compared to female perceivers. Although there is evidence that in some situations female perceivers show greater bias against female leaders than male leaders (Parks-Stamm et al., 2008; Warning & Buchanan, 2009), this "reverse gender bias" effect is unlikely if the female leader is being evaluated by a subordinate or a third-party observer.

Hypothesis 2: Female perceivers, relative to male perceivers, will rate the female leader as:

2a. More *effective* overall than the male leader across the four assessments of leader effectiveness

2b. Being more *leader-like* overall than the male leader across the four assessments of leadership perceptions

Leader Gender by Emotion Suppression Interaction

In order to form accurate judgments about a female leader, it is important to evaluate the leader based on as much information that is available about the leader's behavior. However, a perceiver's ability to incorporate new or schema-inconsistent information into their existing schema to inform their judgment is tied to their cognitive capacity available (Foti & Hauenstein, 1993; Van Overwalle & Labiouse, 2004). When the perceiver has a leadership prototype that has a narrow definition of what is associated with effective female leadership, schema-inconsistent information (i.e. agentic leadership behaviors from the female leader) will not be integrated into the existing schema until a threshold is crossed. From the information-processing perspective, having more information allows individuals more opportunities to readjust their initial impression of the leader. At this point, the perceiver observes female leadership behaviors at

such a frequency that they cannot solely be attributed to external factors anymore (Vonk, 1994). It is only once this behavioral information is included as part of the perceiver's impression of the female leader that the female leader will not be penalized for her schema-inconsistent (i.e. agentic) behaviors.

It is established that perceivers have existing biases against female leaders due to the mismatch between the female gender stereotypes on their leader prototypes. In order to overcome the initial categorization of the female leader, an individual needs sufficient cognitive resources to process the diagnostic information from the female leader. Building upon that, emotion suppression is examined as an important contextual factor that can amplify how gender biases manifest in perceptions and ratings of the female leader. Instead of effortfully processing each behavior carried out by a leader, an ego-depleted individual is more likely to rely on the initial generalized leadership prototype to guide ratings of both female and male leaders. As a result, gender bias in ratings of the female leader is expected to manifest more strongly from those who are ego-depleted compared to those who are not.

Hypothesis 3: Gender bias in leadership judgments will be stronger for those in the emotion suppression condition, relative to no emotion suppression condition, such that a female leader will be:

3a. Rated as less *effective* overall than a male leader across the four assessments of leader effectiveness

3b. Perceived as less *leader-like* overall than a male leader across the four assessments of leadership perceptions

Leader Gender by Emotion Suppression by Perceiver Gender Interaction

In line with the expectation that the gender bias in leader ratings would be stronger for those who suppressed their emotions, the pattern of rating bias by a perceiver's gender should also manifest more clearly amongst those who are ego-depleted relative to non-ego-depleted individuals. As male (female) perceivers are likely to have agentic (communal) attributes more frequently activated in their leadership prototype network, male and female perceivers who are ego-depleted are anticipated to be more strongly guided by their activated leadership prototype network when rating a male/female leader compared to those who are not ego-depleted.

Hypothesis 4: There will be an emotion suppression by leader gender by perceiver gender three-way interaction such that within the emotion suppression condition:

4a: Male participants will rate the male leader as more *competent* than the female leader, relative to female participants

4b: Female participants will rate the female leader as having greater *warmth* than the male leader, relative to male participants

Gender Bias in Leadership Recognition Accuracy

Research on memory and behavior ratings is typically guided by signal detection theory, whereby memory accuracy is evaluated using the metrics of hit rate (i.e. correctly identifying a behavior shown by the target being rated) and false alarm rate (i.e. mistakenly identifying a behavior that was not exhibited by a target) (Lord, 1985). Both measures reflect memory sensitivity or recognition accuracy, understood as an individual's ability to distinguish between behaviors that did happen and those that did not (Hansbrough et al., 2015).

Emotion Suppression Effect

Processing of individuating behaviors is effortful (McGarty et al., 2002); thus, ego-depleted individuals are unable to expend the cognitive resources necessary to process all details of a situation. This likely results in the decreased memory for situational details compared to those who are not ego-depleted.

Hypothesis 5: In the emotion suppression condition, overall recognition behavior accuracy (measured at the end of the experiment) will be lower for both leader agentic behavior and leader communal behavior, relative to the no emotion suppression condition.

Leader Gender by Emotion Suppression Interaction

In addition to the memory biases that manifest more strongly in ego-depleted individuals, memory for leadership behaviors also varies according to what is considered schema-consistent or inconsistent information, and this differs as a function of leader gender. Regardless if a male or female manifests the same leader behaviors, perceivers seek to confirm activated nodes associated with the gender of the leader. Furthermore, an activated node often indirectly activates associated nodes within the prototype network (Lord et al., 2001; Hogue & Lord, 2007). This leads the perceiver to attribute stereotypical characteristics to the leader for which diagnostic behaviors have not occurred. For example, observing a male leader activates nodes for masculine traits in the leader prototype. If the male leader is decisive, the perceiver may make the additional inference that the male leader is intelligent. Similarly, observing a female leader activates gender-related schemas first, and that subsequently influences how her leadership behaviors are observed and encoded.

As agentic behaviors by male leaders and communal behaviors by female leaders are considered schema-consistent, it is expected that there would be low recognition accuracy for

those behaviors as they do not require elaborative processing. In other words, processing behaviors that are consistent with the perceiver's initial categorization of the male/female leader likely lead to high false alarm rates.

Hypothesis 6: Within the emotion suppression condition, there will be an effect of leader gender on leader behavior recognition such that

6a: There will be a greater *false alarm rate for agentic behaviors* for the male leader compared to the female leader

6b: There will be a greater *false alarm rate for communal behaviors* for the female leader compared to the male leader

Research Questions

The following research questions further explore how leader behavior recognition accuracy differs depending on schema-inconsistent/consistent behaviors, and if leader behavior recognition varies by perceiver gender. They provide deeper insight into how the memory of leadership behaviors differ by leader and perceiver gender, thus allowing a more comprehensive understanding of how leadership judgments and recognition bias vary for a female compared to a male leader.

Agentic behaviors by female leaders are schema-inconsistent; as this information is more effortful to process and results in stronger memory traces (Hastie & Kumar, 1979), it is possible that memory recognition accuracy would be higher for agentic behaviors enacted by female leaders.

Research Question 1: Will there be a higher leader behavior recognition accuracy of *agentic behaviors* for female leaders compared to *communal behaviors*?

Furthermore, although past research has not found a gender difference in the encoding of agentic and communal behaviors as a function of the perceiver's gender (Scott & Brown, 2006), connectionist theory research suggests that perceivers' leadership prototype networks are activated differently according to the perceiver's gender (Lord et al., 2001). Since false alarm rates are influenced by the perceiver's initial categorization of the target such that they identify behaviors that are part of the categorization but did not actually occur (Lord, 1985), it is possible that male/female perceivers would have differential false alarm rates for agentic and communal leader behaviors.

Research Question 2: Will male participants have a greater *false alarm rate for agentic behaviors* for the male leader compared to the female leader, relative to female participants?

Research Question 3: Will female participants have a greater *false alarm rate for communal behaviors* for the female leader compared to the male leader, relative to male participants?

Study Overview

In the current study, the influence of emotion suppression on leader ratings across leader gender was examined, with perceiver gender included as a possible moderator. The study was conducted as a virtual experiment session through Zoom and Qualtrics. Participants first watched a 9-minute video compilation of funny clips from the comedy series "The Office". Based on the dual-task paradigm, this was the T1 task where participants were instructed to either watch the videos naturally or to suppress their emotions and maintain a neutral facial expression as they watched the videos. For the T2 task, participants watched a 16-minute video series of four videos depicting conversations between four business managers. There were two versions of these business videos, with the gender of the leader being the only difference across the two versions. The male leader or female leader was identified for the participant according to their condition,

and participants were informed that they would provide ratings about the leader based on the videos. After each of the four videos, they completed two measures: a 4-item rating of leadership effectiveness (Hais et al., 1997), and the 5-item Global Leadership Impression scale (GLI; Cronshaw & Lord, 1987; Lord et al., 1984) that measures the extent to which the target depicted in the video is perceived to be a leader. Finally, after watching the fourth video and providing their leader effectiveness and leadership perception ratings, participants completed measures of recognition memory for leader behaviors across all four videos and reported their perceptions of both the leader's competence and the leader's warmth.

Rationale for the Repeated Measures Design

As highlighted by connectionist theory, prototype activation patterns that guide leadership perceptions are dynamic and change according to contextual constraints and behavioral inputs of the observed target (Lord et al., 2001; Shondrick & Lord, 2010). Schema accessibility unfolds as connections among stereotype nodes related to gender and leadership are activated and inhibited (Lord et al., 2001; Shondrick & Lord, 2010). However, once an impression is formed about a leader, there is also a certain amount of stability tied to the leadership perception process; perceivers' interpretations of leader behaviors tend to align with their initial categorization of the leader, as the weights associated with each of the nodes within the activated leadership prototype pattern change slowly (Lord et al., 2001; Shondrick & Lord, 2010).

Even when participants are not cognitively taxed, the trajectory of leadership perceptions varies between male and female leaders (Brown et al., 1998; Foti et al., 2008). For example, Foti et al. (2008) used two sets of leadership videos, one where the emerging leader was male while the other was female, and asked participants to provide continuous leadership perception ratings

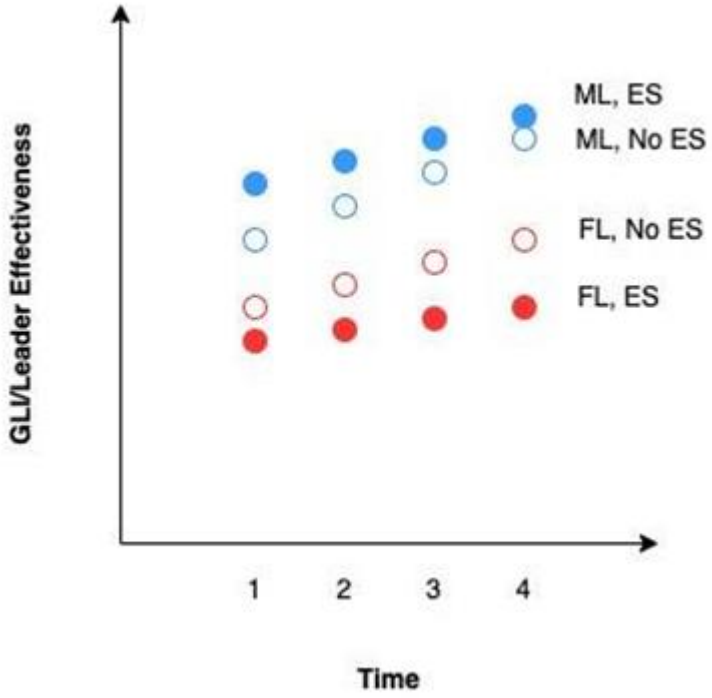
as they watched all nine videos. While they found that the strength of leadership perceptions for both male and female leaders increased as more leadership behaviors were displayed by each leader in their respective leader emergence condition, the patterns of change for leader perceptions over time varied by the leader's gender. In particular, while there was a gradual change in leadership perceptions for emerging male leaders, participants' leadership perceptions of emerging female leaders showed greater resistance to change.

In this study, a within-person, repeated measures approach was employed to investigate the influence of the ego depletion effect on the trajectories of leadership ratings over time. As included above, hypotheses 1-3 pertain to predictions about overall leader effectiveness ratings and leadership perceptions that are collapsed over all four timepoints. In addition to that, it was hypothesized that there would be within-person changes in leader effectiveness ratings and leadership perceptions across the four time points. The pattern of gender bias in the leadership effectiveness ratings and the leadership perception ratings that manifests as the result of ego depletion was predicted to vary as a function of leader gender (see Figure 1). More specifically, as the male leader exhibits more leader behaviors over time, the pattern of leader effectiveness and leadership perception ratings would converge between participants in the emotion suppression and non-suppression conditions. In contrast, as a female leader exhibits more leader behaviors over time, the pattern of leader effectiveness and leadership perception ratings would show divergence between participants in the emotion suppression and non-suppression conditions. The logic for these differing patterns of ratings over time is the greater influence of schematic processing on ratings from those in the emotion suppression condition compared to the non-suppression condition. Ego-depleted participants are likely to be more strongly guided by their initial categorizations compared to the non-ego-depleted participants; for the female

leader, they would be more resistant to providing higher leadership ratings that correspond to the female leader's increasing leadership behaviors relative to the other managers in the business videos, resulting in a divergence of ratings across the suppression conditions. As for the male leader, ego-depleted participants are likely to start with higher leadership ratings and consistently provide high ratings that align with their initial categorization of the male as a leader. This results in the convergence of ratings across the suppression conditions. Furthermore, it was anticipated that the rate of convergence between the emotion suppression and non-suppression conditions will be higher for the male leader, compared to the rate of divergence between the two emotion suppression conditions for the female leader. Alternatively stated, for the non-suppression condition, the slope for the regression line across time would be steeper for both leadership perceptions and leader effectiveness for the male leader than the female leader.

Figure 1

Hypothesized Leader Ratings by Emotion Suppression Condition over Time



Note. The time points reflect the perceiver's perceived leadership ratings after observing each of the business videos. The blue dots represent ratings for the male leader and the red dots for the female leader. The empty dots indicate ratings from the no emotion suppression condition, and filled dots from the emotion suppression condition. The y-axis represents leader ratings for both leader effectiveness and leadership perceptions (i.e. GLI ratings).

Hypothesis 7: There will be an emotion suppression by leader gender by time three-way interaction for *leader effectiveness and leadership perception ratings* such that:

7a: At time 1, the ratings of the male leader will be *higher* for the emotion suppression condition relative to the no emotion suppression condition

7b: At time 4, the ratings of the female leader will be *lower* for the emotion suppression condition relative to the no emotion suppression condition

7c: There will be an emotion suppression by time interaction for the female leader due to a slowly diverging pattern of ratings between the emotion suppression and no emotion suppression condition

7d: There will NOT be an emotion suppression by time interaction for the male leader because of the rapidly converging pattern of leadership ratings between the emotion suppression and no emotion suppression condition

Method

Recruitment

Participants initially targeted for recruitment of this study were undergraduates from the Psychology department of a large Mid-Atlantic university. They were recruited through postings on the department's SONA Experiment Management website, in-class announcements, and electronic flyers. However, recruitment efforts were soon expanded to the entire university population to achieve the sample size goal of 100 males and females each. Announcements were made through multiple listservs, including emails sent to undergraduates from the Psychology department, the university's Office of Undergraduate Research, the Graduate School, Virginia-Maryland College of Veterinary Medicine, and various other interest groups on campus. While these efforts were sufficient to meet the recruitment goal of female participants, additional male

participants were still needed for the study. Thus, individuals were recruited beyond the university setting through postings on my personal Facebook and Instagram accounts. The snowball sampling method was also used; research participants were encouraged to spread information about the study to all those who were eligible. All individuals were required to be at least 18 years old, proficient in English, and have a working computer and webcam to participate. Participants were compensated with a $10 Amazon gift card in exchange for approximately 1-hour of their time.

Participants

As a result of the previously outlined recruitment efforts, the study sample consisted of 193 individuals. The gender breakdown was 100 females and 93 males. Although the number of males was less than the recruitment goal of 100 participants, the time constraints to complete the study dictated the decision to end data collection. Participants ranged from the ages 18 to 36 (M = 23.25, SD = 4.36). The racial breakdown of self-identified race was 103 Whites (53.4%), 45 Asians (23.3%), 20 Hispanic/Latinx (10.6%), 12 Blacks (6.2%), 1 Native-American (0.5%) and 12 Multiracial (6.2%) participants. Out of the 93 males, 59 were recruited from within the university and 34 were not. Furthermore, the sample consisted of undergraduates, postgraduates (i.e. graduate, veterinary, and medical students), and individuals who were working full-time. The breakdown by gender and student status is as follows: for the 100 females, there were 72 undergraduates, 27 postgraduates, and 1 working full-time. For the 93 males, there were 31 undergraduates, 42 postgraduates, and 20 working full-time. The number of participants across the leader gender and emotion suppression conditions was balanced with 97 (50.3%) in the male leader condition and 96 in the female leader condition (49.7%). There were 96 (50.3%) in the no suppression condition and 97 (49.7%) in the suppression condition.

Design

The study design is a 2 (Emotion Suppression: Suppression/ No Suppression) x2 (Leader Gender: Male/Female) x2 (Participant Gender: Male/Female) ANOVA balanced factorial design. All participants were randomly assigned to one of the resulting eight (2x2x2) conditions. A power analysis was computed through the G*Power 3.1 program (Faul et al., 2009) to estimate that 199 participants are needed to achieve 0.8 power (with Cohen's f=0.20, and alpha= .05) for an ANOVA analysis of the highest order interaction predictions.

Procedure

The entire study procedure was conducted online through Qualtrics and Zoom. A Zoom meeting link was sent out to participants according to the timeslot that they signed up for. First, the procedure was explained to participants by the experimenter on Zoom. Participants were reminded that their facial expressions would be recorded as they watched the video stimuli. After the instructions were provided, they had the opportunity to ask any clarification questions. If they did not have any issues, participants were asked to provide their informed consent by clicking on the "I Agree" button on Qualtrics. Only once all questions and concerns were addressed and participants provided their informed consent did the experiment begin.

Participants first carried out the emotion suppression task. They watched a 3-minute Jackson Pollock documentary to establish a baseline mood before watching a 9-minute video compilation of funny scenes from the series "The Office". Before both videos, instructions flashed on the screen according to the assigned emotion suppression condition, telling participants whether they should suppress their emotions or watch the videos naturally. The participants' expressions while watching both videos were recorded through Zoom, and saved on the cloud in the experimenter's Canvas account. Next, participants watched a 16-minute

collection of four vignettes depicting the interaction between four managers working together on various business-related tasks. There are two versions of the business videos: one where an individual named "Bob" is the leader, and the other where an individual named "Sue" is the leader. Participants watched only one of the two versions, depending on their assigned Leader Gender condition. Prior to watching the video, participants were shown a screenshot of the video including the names of the four individuals, and an indication of whether Bob or Sue is the leader. They were also informed that they would provide ratings about the video after each of the four business videos and also after all the videos have played. After each business video, participants were asked to rate their perceptions of Bob/Sue as a leader and also rate the leader's effectiveness. Once the participant watched all the business videos, the video recording stopped. Then, participants completed two questionnaires regarding their recognition of leader behaviors and ratings of both the leader's competence and warmth. The presentation of the scales were randomized. Participants then completed a measure on their emotion regulation strategies used during the study. To find out if any participants were experiencing Zoom/video calling fatigue, they also reported what they had been doing one hour prior to the experiment. Finally, participants were provided with a debriefing statement summarizing the goals of the study and given the opportunity to ask the experimenter if they had any remaining questions. The entire procedure took approximately 60 minutes.

Business Video Stimuli

The four business videos used in this study are part of a broader set of leader emergence videos created by Hanges et al. (1997). The video series consists of nine workplace vignettes, where four White managers (two male, two female) are discussing various work-related issues. Each vignette covered a different theme (e.g. scheduling, budget, supplies) and had a length of

around 4-5 minutes each. There were two duplicate versions of these nine videos, one featuring "Bob" as the male leader, and the other featuring "Sue" as the female leader. Thus, aside from the leader's gender, both versions are the same with the leaders speaking and behaving identically.

Since the purpose of this study is to evaluate leadership, the current study only included the last four vignettes (scenes 6 - 9) where the leader has already emerged. Thus, two stimuli sets were used, one where the male leader executes the most leadership behaviors in the team meeting, and the second where the female leader executes the most leadership behaviors. As the original purpose of the videos was for depicting leader emergence, the number of leadership behaviors displayed by the leader relative to others increases from vignettes six to nine. In total, 21 agentic behaviors and 20 communal behaviors are depicted across the four vignettes. Appendix B contains the list of agentic and communal attributes and the breakdown by video of the frequencies of agentic/communal leaders behaviors exhibited by each team member.

Independent Variables

Emotion Suppression

To avoid any initial psychological differences between groups, the instructions were displayed right before the screening of every video clip, as done by other researchers (e.g., Demaree et. al., 2006). Participants were video-recorded through Zoom, and the participants' web camera light served as a reminder and motivating factor for them to maintain their emotion regulation strategy. The saved video footage of participants' emotional expressions was analyzed using an adapted version of a behavior coding system from Gross and Levenson (1993; 1997) where the frequency of various facial, mouth, and body movements were coded by independent

coders. Additionally, the perceived overall emotional expressivity for each participant was rated on a 5-point Likert scale from 1 (*No Reaction at All*) to 5 (*Extremely Reactive*).

Suppressing emotional expressions while watching videos is a common manipulation that has been demonstrated as robust in previous experiments (e.g., Gross & Levenson, 1993; Vohs & Schmeichel, 2003; Gross & John, 2003). In the emotion suppression condition, participants were instructed to suppress their emotions and maintain a neutral facial expression. Instructions for the suppression condition are as follows:

You will soon watch a short video clip. It is important that you pay attention and watch it carefully, immersing yourself in the scene you will be watching. Please do your best to suppress any expression of emotions you may feel while watching the videos. Try to maintain a neutral facial expression, so that a person who is watching you would not know that you are feeling anything at all. You are reminded that you are being video recorded, and are free to stop the experiment at any time. When you have finished the video clip, please alert the experimenter.

As for the participants in the no-suppression condition, they were not instructed to control their emotions prior to watching the videos. Instructions for the no-suppression condition are as follows:

You will soon watch a short video clip. It is important that you pay attention and watch it carefully, immersing yourself in the scene you will be watching. Please watch the videos as you would on your own, as naturally as possible. You are reminded that you are being video recorded, and are free to stop the experiment at any time. When you have finished the video clip, please alert the experimenter.

Modulation Strategy Questionnaire. As a manipulation check for the emotion suppression condition, participants in the emotion suppression condition were asked several

questions to assess if they were complying with the regulation instructions or utilizing a different strategy. The questionnaire was administered after participants completed all other leader-related measures. Following the procedure from Martjin et. al (2002), participants answered three statements accessing their perceived success in hiding their emotions on a Likert rating of 1(*Strongly Disagree*) to 5(*Strongly Agree*). Additionally, participants across both the emotion suppression and control condition endorsed two items regarding the emotional control exerted during the task with a similar Likert rating (Martjin et. al., 2002).

Dependent Variables

Repeated Measures

Leader Effectiveness. After each of the four business videos, participants responded to a 4-item measure on perceived leadership effectiveness adapted from Hais et. al. (1997). The male/female leader observed in the video was rated on a Likert scale from 1(*Strongly Disagree*) to 5(*Strongly Agree*). Sample items include "This business manager is very effective as a leader" and "This business manager has good leadership qualities".

Leadership Perceptions. Participants' leadership perceptions, as reflected by their confidence in their categorization of the leader, were measured using the Global Leadership Impression scale (GLI; Cronshaw & Lord, 1987; Lord et al., 1984). The GLI consists of 5 items that are rated on a Likert scale from 1(*Not at All*) and 5 (*Extreme Amount*). Example items include "To what extent does {Sue/Bob} fit your image of a leader?" and "How typical was {Sue/Bob} of a leader?." The GLI has demonstrated good reliability, with a reported reliability alpha coefficient of .87 (Cronshaw & Lord, 1987).

Behavior Recognition

To assess memory for leadership behaviors exhibited by the male/female leader across the four business videos, participants completed a 24-item behavior recognition questionnaire. This measure consists of 12 agentic and 12 communal behaviors, where half of the behaviors for each of the two behavior dimensions occur (i.e. 6 agentic and 6 communal) while the other half do not occur. Each of the items was rated on a Yes/No scale. Following past research (Martell & Evans, 2005; Hansbrough et al., 2020), participants' false alarm rates were calculated by the proportion of communal and agentic behaviors that they identified but did not actually occur in the videos. Recognition behavior accuracy was calculated by the ratio of true rates (true positives, true negatives) divided by total number of behaviors on the measure (i.e. 24 behaviors).

Leader Competence and Warmth Ratings

Following Fiske and colleagues' Stereotype Content Model (SCM; Fiske, 2018; Fiske et al., 2002), evaluation of the male/female leader from the business videos was assessed according to the dimensions of warmth (i.e. trustworthiness, friendliness) and competence (i.e. capability, assertiveness). Following DeRue et al. (2015), Mayer and Davis' (1999) Benevolence and Ability subscales were adapted to measure participants' ratings of the leader's performance in terms of warmth and competence respectively. The Benevolence subscale contains 5 items related to concern for others, whereas the Ability subscale includes 6 items pertaining to task and leadership competence. All items are scored on a 5-point Likert scale from 1(*Strongly Disagree*) to 5(*Strongly Agree*). The reported reliability for both scales is adequate, with $\alpha = .85$ for Benevolence and $\alpha = .91$ for Ability (DeRue et al., 2015).

Analyses

Hypotheses 1a and 1b, 2, 3, and 7 regarding leader effectiveness and leadership perception ratings were tested using multilevel modeling where time was treated as the level-1 predictor and Emotion Suppression, Leader Gender, and Perceiver Gender were treated as level-2 predictors. In addition to testing the predicted mean differences across conditions for these four hypotheses, the slope differences across the different conditions over time were tested for hypotheses 7c and 7d. The omnibus test for Hypotheses 1c, 1d, 4, 5, and 6 related to leader behavior recognition, competence, and warmth was a 2 (Emotion Suppression) x 2 (Leader Gender) x 2 (Perceiver Gender) ANOVA with follow-up *a priori* contrasts to test predicted mean differences.

Results

Data Cleaning

Before conducting analyses, data quality was checked in terms of 1) compliance with emotion suppression manipulation instructions and 2) participants' overall false alarm rates on the Behavior Recognition measure. Six participants appeared to incorrectly suppress their emotions even though they were assigned to the no suppression condition and instructed to watch the T1 video stimuli naturally. After a closer inspection of their responses on both the Modulation Strategy Questionnaire that measured self-reported effort in emotion suppression and the video recordings of their emotional expressions, five participants were reassigned to the emotion suppression condition[1]. For the remaining participant, it was unclear whether the individual was exerting effort to suppress their emotions or not; as a result, the participant was

[1] To ensure findings were not influenced by the reassignment of the 5 participants to the emotion suppression condition, post-hoc exploratory analyses were conducted for all hypotheses on the sample without these participants (N=184). However, most of the findings did not change, so the 5 participants were kept in the sample.

dropped from the sample. Additionally, participants' responses on the Behavior Recognition measures were evaluated as a form of an attention check. There were three participants who had a total false alarm rate of one; these participants did not discriminate between the recognition items and incorrectly identified all leadership behaviors that did not occur in the videos. All three participants were removed from further analyses.

Thus, the final sample consisted of 189 participants, with a gender breakdown of 100 females and 89 males. All four participants who were removed were males who were recruited from the university. As for the student status breakdown, the remaining male participants were 29 undergraduates, 41 postgraduates, and 19 working full-time. The distribution of participants across conditions remained somewhat balanced; there were 96 (50.8%) in the male leader condition and 93 in the female leader condition (49.2%) and 88 (46.6%) in the no suppression condition and 101 (53.4%) in the suppression condition.

Descriptive Statistics

The means, standard deviations, and correlations for all the variables measured after the four business videos are reported in Table 1. Both the Leader Competence and Warmth measures demonstrated excellent reliability, with coefficient alpha values of .94 and .92 respectively.

Table 1

Means, Standard Deviations, Correlations, and Reliabilities between Variables of Interest

	Mean	*SD*	1	2	3	4	5
1. Leader Competence	3.68	.71	(.94)				
2. Leader Warmth	3.52	.76	.75**	(.92)			
3. Sensitivity	.63	.11	-.02	-.05	(-)		
4. FAR_A	.40	.27	.09	.04	-.60**	(-)	
5. FAR_C	.60	.24	.32**	.38**	-.50**	.29**	(-)

Note. $N = 189$. FAR_A represents false alarm rates on the Behavior Recognition measure for agentic behaviors, and FAR_C represents false alarm rates for communal behaviors. Cronbach's alpha are included in parentheses.

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

The corresponding descriptive statistics for Leader Effectiveness and Leadership Perception measures that were administered after each of the four business videos are reported in Table 2. Both measures also showed strong reliability, ranging from α = .94-.98 for Leader Effectiveness and α = .90-.95 for Leadership Perceptions.

Table 2

Correlations of Leader Effectiveness and Leadership Perception Ratings across four Business Videos

	Mean	*SD*	1	2	3	4	5	6	7	8
1. LE_T1	3.67	0.77	(.94)				.			
2. LE_T2	3.70	0.92	.32**	(.97)						
3. LE_T3	3.78	0.87	.42**	.28**	(.97)					
4. LE_T4	3.74	0.99	.25**	.28**	.42**	(.98)				
5. LP_T1	3.53	0.74	85**	.32**	.42**	.22**	(.90)			
6. LP_T2	3.54	0.88	.25**	.94**	.30**	.25**	.29**	(.94)		
7. LP_T3	3.71	0.81	.38**	.24**	.94**	.37**	.42**	.30**	(.95)	
8. LP_T4	3.77	0.87	.25**	.27**	.42**	.95**	.22**	.25**	.39**	(.95)

Note. N = 189. LE represents Leader Effectiveness ratings across videos 1-4. LP represents Leadership Perceptions ratings. Cronbach's alphas are included in parentheses.

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

Additionally, to examine if Leader Effectiveness and Leadership Perception ratings varied by emotion suppression, the correlations within the emotion suppression and no suppression conditions are reported in Table 3. The pattern of correlations were similar to the overall sample; ratings of Leader Effectiveness and Leadership Perception were highly correlated within each time point, but modestly correlated within each of the two measures across the four time points.

Table 3

Correlations of Leader Effectiveness and Leadership Perception Ratings across four Business Videos for each Emotion Suppression condition

	Mean1(SD)	*Mean2(SD)*	1	2	3	4	5	6	7	8
1. LE_T1	3.72 (0.85)	3.62 (0.69)	-	.28**	.44**	.31**	.82**	.23*	.41*	.30**
2. LE_T2	3.70 (0.92)	3.70 (0.92)	.37**	-	.23*	.13	.23*	.95**	.20*	.10
3. LE_T3	3.69 (0.96)	3.86 (0.77)	.42**	.32**	-	.40**	.46**	.26**	.95**	.42**
4. LE_T4	3.59 (1.09)	3.88 (0.88)	.23*	.43**	.42**	-	.29**	.13	.40**	.96**
5. LP_T1	3.58 (0.83)	3.50 (0.66)	.87**	.41**	.41**	.19	-	.24*	.45**	.27**
6. LP_T2	3.54 (0.86)	3.55 (0.89)	.28**	.92**	.35**	.37**	.34**	-	.25*	.11
7. LP_T3	3.63 (0.89)	3.78 (0.74)	.36**	.29**	.94**	.34**	.40**	.35**	-	.43**
8. LP_T4	3.68 (0.90)	3.84 (0.84)	.21*	.45**	.41**	.95**	.18	.41**	.35**	-

Note. N1 = 88 for the no suppression condition. Mean1 represents mean ratings from the no suppression condition. *N2* = 101 for the emotion suppression condition. Mean2 represents mean

ratings from the emotion suppression condition. LE represents Leader Effectiveness ratings across videos 1-4. LP represents Leadership Perceptions ratings. Correlations from the no suppression condition are reported in the bottom triangle and emotion suppression condition in the upper triangle.

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

Emotion Suppression Manipulation Checks

The effectiveness of the emotion suppression manipulation was assessed by two methods: 1) participants' responses on the Modulation Strategy Questionnaire regarding the effort exerted in emotional control and 2) video recordings of participants' emotional expressions that were behaviorally coded.

A t-test was conducted to compare responses on the Modulation Strategy Questionnaire across the emotion suppression and no suppression conditions. Participants in the emotion suppression condition reported exerting greater emotional control (M=4.28, SD=0.53) compared to those in the no suppression condition (M=2.06, SD=1.02), with $t(187) = 19.02$ $p = .00$, Cohen's $d = 2.77$.

Behavioral Coding of Emotional Expressions

Two undergraduate research assistants were trained to watch the video recordings of participants and rate them on their emotional expressivity based on an adapted version of Gross and Levenson's Emotional Expressive Coding System (1993; 1997). The coders evaluated participants' emotional expressions based on their body, facial, and mouth movements before providing ratings on overall emotional expressivity. Prior to coding the videos, both coders were trained using a codebook that was developed for this purpose. Adapted from the Facial Action Coding System (FACS; Ekman et al., 2002), the codebook listed a range of visible

emotional expressions that were organized into categories of body, facial, and mouth movements. In the early stages of the coding process, the coders met frequently to discuss coding-related issues and to ensure they were both using similar decision criteria when deciding what to code as an emotional expression or not.

To avoid potential leader gender and perceiver gender effects, both coders independently coded five videos that were randomly selected from each of the possible eight conditions, totaling up to 40 videos. Furthermore, coders were not informed which of the 20 videos were participants in the emotion suppression and which of the remaining 20 videos were those in the no suppression condition. Excellent inter-rater reliability was achieved between both coders, with $r = .90$. To assess the effectiveness of the emotion suppression manipulation, ratings of overall emotional expressivity were compared across the conditions. Participants in the suppression condition (M=1.73, SD=0.98) were perceived as *less* expressive than those in the no suppression condition (M=2.75, SD=1.12), with $t(38) = 3.08$, $p = .00$, Cohen's $d = 0.97$. Additionally, as past research demonstrates that males suppress their emotions more often than females (Gross & John, 2003), the emotional expressiveness across gender was also compared. In this sample, males (M=2.38, SD=1.21) did not differ from females (M=2.10, SD=1.12) in terms of their overall emotional expressivity ratings with $t(38) = 0.75$, $p = .46$.

Thus, ratings from both the Modulation Strategy Questionnaire and the behavioral coding indicate that the emotion suppression manipulation was successful.

Hypothesis Testing

All repeated measures hypotheses for Leader Effectiveness and Leadership Perceptions ratings that were measured after each of the four business videos are discussed first followed by

findings for overall ratings of Leader Competence and Warmth, and Behavior Recognition responses that were collected after watching all four videos.

Repeated Measures Analyses

All hypotheses related to Leader Effectiveness and Leadership Perceptions were tested on Mplus (Version 8; Muthén and Muthèn, 2017). Participants were treated as between-group Level-2 variables (N = 189) and their ratings after each of the four business videos were treated as within-group Level-1 variables (N = 756). All models tested were estimated using the BAYES option on Mplus, which utilizes a Bayesian estimation method. As a result, the p-value significance of estimates are evaluated using one-tailed tests and 95% credibility intervals are reported. To enable straightforward interpretation of the categorical predictors (i.e. Leader Gender: -1 for male, 1 for female, Perceiver Gender: -1 for male, 1 for female, and Emotion Suppression: -1 for no suppression, 1 for suppression), all regression coefficients are reported in their unstandardized form. The intraclass correlation values for both Leader Effectiveness ($ICC1$ = .32) and Leadership Perceptions ($ICC1$ = .30) demonstrate that there was sufficient variability of the outcome variables both at the between-group and within-group levels, indicating that multilevel modeling is the appropriate method for data analysis.

Leader Effectiveness and Leadership Perceptions Collapsed Over Time

Leader Gender Effect. Hypotheses 1a and 1b predicted a leader gender main effect on leadership ratings, such that the female leader would be rated less effective and perceived as less leader-like than the male leader overall, across the four time points. To test Hypothesis 1a and 1b, two multilevel models were tested with Leader Gender entered as the Level 2 predictor and Leader Effectiveness and Leadership Perceptions as the respective outcome variables. Model estimates are reported in Table 4.

Table 4

Prediction of Leader Effectiveness and Leadership Perceptions from Leader Gender for Hypothesis 1

Level and Variable	Leader Effectiveness	Leadership Perceptions
Level 2		
Intercept (γ_{00})	3.73** (0.05)	3.64** (0.04)
Leader Gender (γ_{01})	0.13** (0.04)	0.13** (0.04)
Variance Components		
Within-person variance (σ^2)	0.54	0.49
Residual variance (τ_{00})	0.24	0.19
Additional information		
ICC	.32	.30
R^2 between group	.07	.08

Note. Level 1 sample size = 756 ratings, Level 2 sample size = 189. Males coded as -1, females coded as 1 for Leader Gender.

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

For the Leader Effectiveness model, leader ratings were greater for the female leader (*M*=3.85, *SD*=0.54) than the male leader (*M*=3.59, *SD*=0.67), γ_{01} = 0.13, *SD* = 0.04, *p* = .001, 95% CI [0.06, 0.21]. Hypothesis 1a was not supported. Similar results were found for the Leadership Perceptions model, where higher leadership perception ratings were predicted by the female leader (*M*=3.77, *SD*=0.53) compared to the male leader (*M*=3.51, *SD*=0.58), γ_{01} = 0.13, *SD* = 0.04, *p* = .001, 95% CI [0.06, 0.20]. Hypothesis 1b was also not supported. Contrary to predictions, the female leader was not rated less favorably; instead, she was given greater ratings on both leader effectiveness and leadership perceptions than the male leader.

Leader Gender by Perceiver Gender Interaction. Hypothesis 2a predicted an interaction between Leader Gender and Perceiver Gender such that female perceivers (compared

to male perceivers) would rate the female leader as more effective than the male leader overall, across four time points. The main effect terms of leader and perceiver gender, along with their interaction term were entered as Level-2 predictors in the model, with Leader Effectiveness as the outcome variable.

Table 5

Prediction of Leader Effectiveness and Leadership Perceptions from Leader Gender and Perceiver Gender for Hypothesis 2

Level and Variable	Leader Effectiveness	Leadership Perceptions
Level 2		
Intercept (γ_{00})	3.72** (0.05)	3.64** (0.04)
LG (γ_{01})	0.13** (0.04)	0.13** (0.13)
PG (γ_{02})	0.01 (0.04)	0.00 (0.04)
LG x PG (γ_{03})	0.05 (0.05)	0.03 (0.04)
Variance Components		
Within-person variance (σ^2)	0.54	0.49
Residual variance (τ_{00})	0.25	0.20
Additional information		
ICC	.32	.30
R^2 between group	.09	.10

Note. Level 1 sample size = 756 ratings, Level 2 sample size = 189. LG represents Leader Gender and PG represents Perceiver Gender. Males coded as -1, females coded as 1 for Leader and Perceiver Gender.

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

As reported in Table 5, the leader gender and perceiver gender interaction effect was not significant. Hypothesis 2a was not supported. However, the leader gender main effect was significant. To better understand the findings, a simple slopes test was carried out; the effect of leader gender on leader effectiveness ratings was tested at the two levels of perceiver gender. For

female perceivers, the female leader (*M*=3.92, *SD*=0.51) was associated with higher leader effectiveness ratings than the male leader (*M*=3.55, *SD*=0.65), *simple slope* = .18, p = .00, 95% CI [0.08, 0.29]. In contrast, for male perceivers, no relationship was found between the leader's gender and leader effectiveness.

For Hypothesis 2b, the multilevel model tested included Leader Gender, Perceiver Gender, and the two-way interaction term as the Level-2 predictors and Leadership Perceptions as the outcome variables. Hypothesis 2b also predicted that female perceivers (relative to males) would perceive the female leader as more leader-like than the male leader. Again, the interaction effect was not significant (see Table 5) thus providing no evidence for Hypothesis 2b. Similar to the Leader Effectiveness model, the leader gender main effect was significant in this model too. A simple slopes test was conducted looking at the effect of leader gender on leadership perception ratings across the two levels of perceiver gender. Results paralleled findings for Leader Effectiveness; for female perceivers, the female leader (*M*=3.80, *SD*=0.52) had higher leadership perception ratings than the male leader (*M*=3.49, *SD*=0.54), *simple slope* = .16, p = .00, 95% CI [0.06, 0.25]. However, no significant association was found for male perceivers.

Taking both the Leader Effectiveness and Leadership Perceptions findings together, the results suggest that while male perceivers did not differentiate between the male and female leader when providing their leader ratings, female perceivers evaluated the female leader more favorably than the male leader. Although the interaction term between Leader Gender and Perceiver Gender was not significant in the model, this provides limited support to the predictions of Hypothesis 2a and 2b whereby female perceivers did give the female leader higher leader ratings than the male leader, relative to male perceivers.

Leader Gender by Emotion Suppression Interaction. Hypothesis 3a predicted an interaction between Leader Gender and Emotion Suppression on Leader Effectiveness ratings, such that for the emotion suppression condition (compared to the no suppression condition), ratings for the female leader would be lower than the male leader. Leader Gender, Emotion Suppression, and their interaction term were entered as Level-2 predictors in the model, with Leader Effectiveness as the outcome variable. Regression estimates are reported in Table 6.

Table 6

Prediction of Leader Effectiveness and Leadership Perceptions from Leader Gender and Emotion Suppression for Hypothesis 3

Level and Variable	Leader Effectiveness	Leadership Perceptions
Level 2		
Intercept (γ_{00})	3.72** (0.05)	3.64** (0.04)
LG (γ_{01})	0.13** (0.04)	0.13** (0.04)
ES (γ_{02})	0.05 (0.05)	0.03 (0.04)
LG x ES (γ_{03})	0.00 (0.05)	-0.01 (0.04)
Variance Components		
Within-person variance (σ^2)	0.54	0.49
Residual variance (τ_{00})	0.25	0.20
Additional information		
ICC	.32	.30
R^2 between group	.09	.10

Note. Level 1 sample size = 756 ratings, Level 2 sample size = 189. LG represents Leader Gender and ES represents Emotion Suppression. Males coded as -1, females coded as 1 for Leader Gender. No suppression was coded as -1 and suppression as 1.

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

The only significant predictor was the leader gender main effect term, where the female leader had higher leader effectiveness ratings than the male leader across both the emotion suppression and no suppression conditions, γ_{01} = 0.13, *SD* = 0.04, *p* = .002, 95% CI [0.06, 0.21]. This echoes the leader gender effect from Hypothesis 1a. The interaction term was not significant, thus there was no support for Hypothesis 3a. However, the Leader Gender main effect aligns with findings from the simple slopes test, which examined the relationship between leader gender and leader effectiveness ratings at the two different levels of emotion suppression. For the no suppression condition, the female leader was associated with higher leader effectiveness ratings (*M*=3.81, *SD*=0.59) than the male leader (*M*=3.54, *SD*=0.76), *simple slope* = .13, p = .02, 95% CI [0.02, 0.24]. Similar simple slope effects are found for the suppression condition (*simple slope* = .14, p = .02, 95% CI [0.03, 0.23]), where the female leader (*M*=3.90, *SD*=0.49) is rated with higher leader effectiveness ratings than the male leader (*M*=3.63, *SD*=0.59).

Hypothesis 3b also predicted a Leader Gender and Emotion Suppression interaction effect on Leadership Perception ratings, where ratings for the female leader (relative to the male leader) were expected to be lower in the emotion suppression than the no suppression condition. The two main effects and the interaction term between Leader Gender and Emotion Suppression were added as Level-2 predictors in the model with Leadership Perceptions as the outcome variable. As was seen for Hypothesis 3a, the interaction term was not significant but the leader gender main effect was, (γ_{01} = 0.13, *SD* = 0.04, *p* = .001, 95% CI [0.06, 0.20]). As a result, Hypothesis 3b was not supported. The simple slope test reflected the leader gender main effect findings; for those in the no suppression condition, the female leader was associated with higher leader effectiveness ratings (*M*=3.75, *SD*=0.57) than the male leader (*M*=3.47, *SD*=0.63), *simple*

slope = .14, p = .01, 95% CI [0.04, 0.24]. Similar simple slope effects are found for the suppression condition (*simple slope* = .12, p = .02, 95% CI [0.03, 0.21]), where the female leader (*M*=3.79, *SD*=0.50) is rated with higher leader effectiveness ratings than the male leader (*M*=3.55, *SD*=0.54). Again, contrary to expectations, the female leader was rated more favorably than the male leader on both Leader Effectiveness and Leadership Perception ratings. Furthermore, this effect was observed across both no suppression and emotion suppression conditions.

Leader Effectiveness and Leadership Perceptions Over Time

Time by Leader Gender by Emotion Suppression Interaction. Hypotheses 7a to 7d related to the three-way interaction between Leader Gender and Emotion Suppression over Time for both Leader Effectiveness and Leadership Perception ratings. The focus of this hypothesis is whether the trajectory of both types of leader ratings varies by the leader's gender and emotion suppression condition across the four business videos. Hence, to test this hypothesis, Time was entered as a Level-1 predictor and represented in the model as the slope effect on the relevant outcome variable (i.e. Leader Effectiveness *or* Leadership Perceptions). Time was treated as a random slope and allowed to vary across participants. The three other Level-2 predictors included in the model were Leader Gender, Emotion Suppression, and their interaction term. Finally, the cross-level interaction terms entered in the model included two two-way interaction terms which were Time x Leader Gender and Time x Emotion Suppression, and one three-way interaction term which was Time x Leader Gender x Emotion Suppression.

The model-predicted values from the multilevel model analysis of Leader Effectiveness over time across the Leader Gender and Emotion Suppression conditions are reported in Table 7 and plotted in Figure 2. To provide additional perspective, the raw mean leader ratings for

Leader Effectiveness by Leader Gender and Emotion Suppression conditions across each of the four time points are reported in Table 8 and plotted in Figure 3.

Table 7

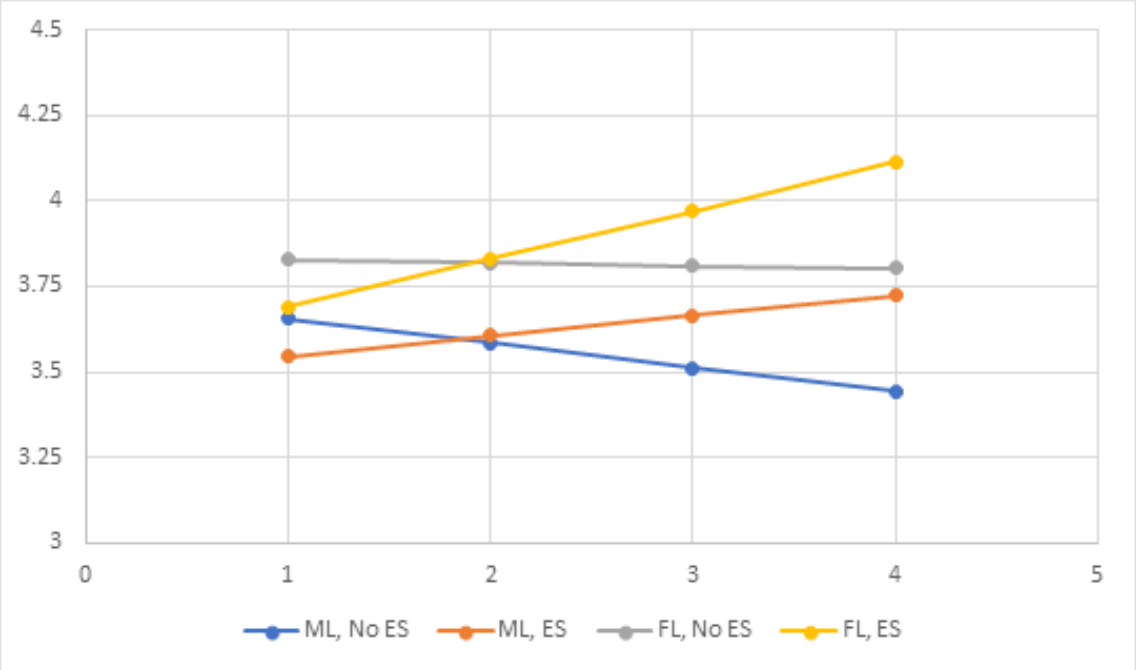
Model-predicted Leader Effectiveness Ratings by Leader Gender and Emotion Suppression Conditions over Time

Condition	Time 1	Time 2	Time 3	Time 4
ML, No ES	3.66	3.59	3.51	3.44
ML, ES	3.55	3.61	3.67	3.72
FL, No ES	3.83	3.82	3.81	3.80
FL, ES	3.69	3.83	3.97	4.12

Note. N = 756 ratings. Time represents the four time points when leader ratings were collected after each of the business videos. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Figure 2

Model-predicted Leader Effectiveness Ratings by Leader Gender and Emotion Suppression condition over Time



Note. N = 756 ratings. To enable easier interpretation of the graph, the y-axis (which reflects Leader Effectiveness ratings) was truncated from the original scale from 1 to 5 to 3 to 4.5. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Table 8

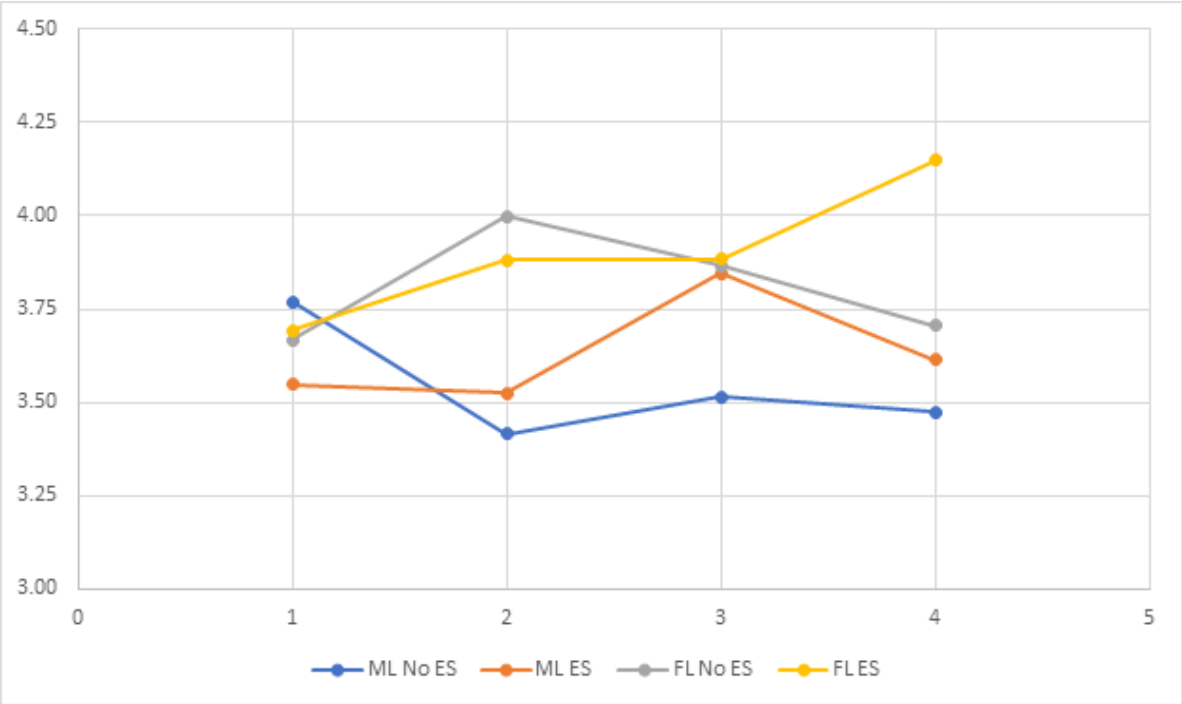
Raw Mean Values of Leader Effectiveness Ratings by Leader Gender and Emotion Suppression Conditions over Time

Condition	Time 1	Time 2	Time 3	Time 4
ML, No ES	3.77	3.42	3.52	3.47
ML, ES	3.55	3.52	3.84	3.61
FL, No ES	3.67	4.00	3.87	3.70
FL, ES	3.69	3.88	3.88	4.15

Note. N = 756 ratings. Time represents the four time points when leader ratings were collected after each of the business videos. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Figure 3

Raw Mean Values of Leader Effectiveness Ratings by Leader Gender and Emotion Suppression condition over Time



Note. N = 756 ratings. To enable easier interpretation of the graph, the y-axis (which reflects Leadership Perception ratings) was truncated from the original scale from 1 to 5 to 3 to 4.5. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

For Leadership Perceptions, the model-predicted values are reported in Table 9 and plotted in Figure 4. The corresponding raw mean ratings are included as Table 10 and plotted as Figure 5.

Table 9

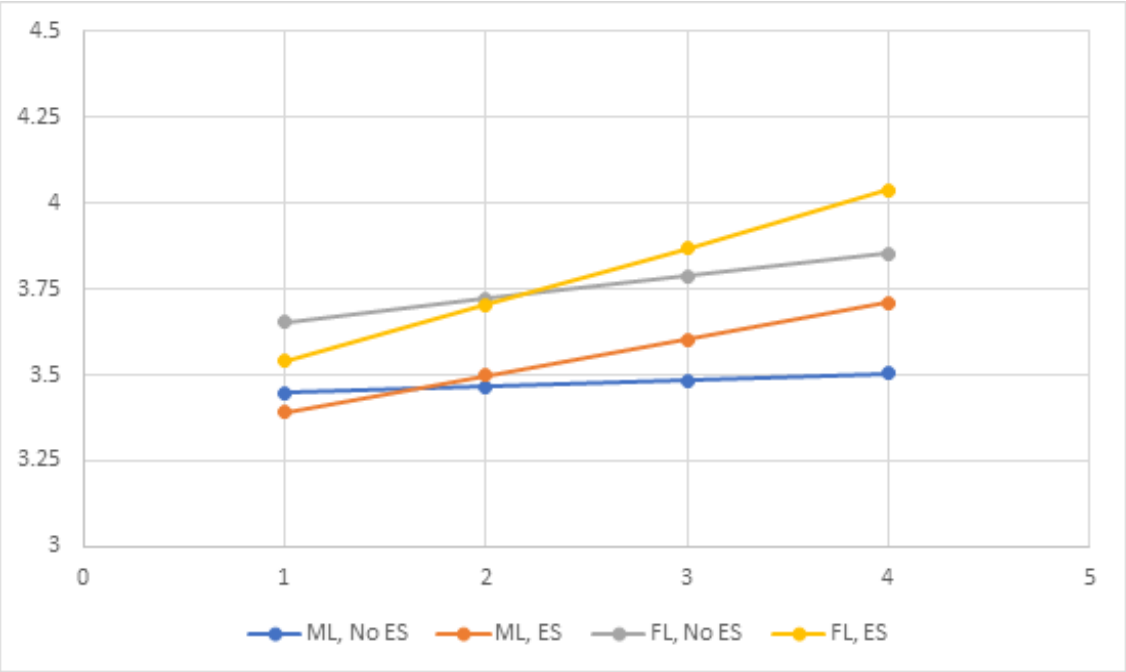
Model-predicted Leader Perceptions Ratings by Leader Gender and Emotion Suppression Conditions over Time

Condition	Time 1	Time 2	Time 3	Time 4
ML, No ES	3.45	3.47	3.48	3.50
ML, ES	3.39	3.50	3.60	3.71
FL, No ES	3.65	3.72	3.79	3.85
FL, ES	3.54	3.71	3.87	4.04

Note. $N = 756$ ratings. Time represents the four time points when leader ratings were collected after each of the business videos. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Figure 4

Model-predicted Leadership Perception Ratings by Leader Gender and Emotion Suppression condition over Time



Note. N = 756 ratings. To enable easier interpretation of the graph, the y-axis (which reflects Leadership Perception ratings) was truncated from the original scale from 1 to 5 to 3 to 4.5. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Table 10

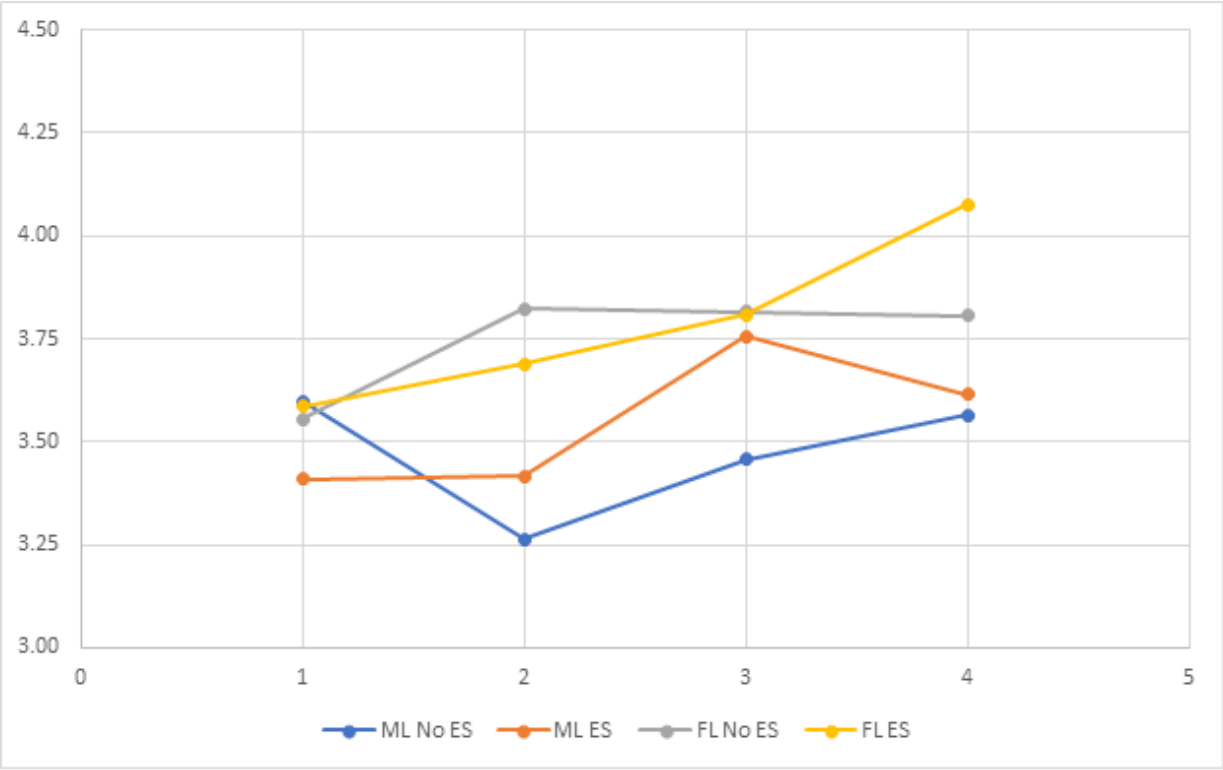
Raw Mean Values of Leadership Perception Ratings by Leader Gender and Emotion Suppression Conditions over Time

Condition	Time 1	Time 2	Time 3	Time 4
ML, No ES	3.60	3.26	3.46	3.56
ML, ES	3.41	3.42	3.76	3.61
FL, No ES	3.56	3.82	3.81	3.81
FL, ES	3.59	3.69	3.81	4.08

Note. N = 756 ratings. Time represents the four time points when leader ratings were collected after each of the business videos. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Figure 5

Raw Mean Values of Leadership Perception Ratings by Leader Gender and Emotion Suppression condition over Time



Note. N = 756 ratings. To enable easier interpretation of the graph, the y-axis (which reflects Leadership Perception ratings) was truncated from the original scale from 1 to 5 to 3 to 4.5. ML represents male leader, FL represents female leader. No ES represents the no emotion suppression condition while ES represents emotion suppression condition.

Table 11

Prediction of Leader Effectiveness and Leadership Perceptions from Leader Gender and Emotion Suppression over Time for Hypothesis 7

Level and Variable	Leader Effectiveness	Leadership Perceptions
Level 2		
Intercept (γ_{00})	3.68** (0.06)	3.51** (0.05)
Intercept of Time (γ_{10})	0.03 (0.03)	0.09 (0.03)
LG (γ_{01})	0.08 (0.05)	0.09 (0.05)
ES (γ_{02})	-0.06 (0.06)	-0.05 (0.05)
LG x ES (γ_{03})	-0.01 (0.05)	-0.02 (0.05)
Time x LG (γ_{11})	0.04 (0.03)	0.03 (0.02)
Time x ES (γ_{12})	0.07* (0.03)	0.05* (0.02)
Time x LG x ES (γ_{13})	0.01 (0.02)	0.00 (0.02)
Variance Components		
Within-person variance (σ^2)	0.50	0.46
Residual variance (τ_{00})	0.21	0.18
Time slope variance (τ_{10})	0.02	0.01
Additional information		
ICC	.32	.30
R^2 within group	.06	.06
R^2 between group	.04	.04
R^2 between group for Time	.17	.18

Note. Level 1 sample size = 756 ratings, Level 2 sample size = 189. Time represents the four time points when leader ratings were collected after each of the business videos. LG represents Leader Gender and ES represents Emotion Suppression. Males coded as -1, females coded as 1 for Leader Gender. No suppression was coded as -1 and suppression as 1. Time coded as 0 to 3 for time 1 – time 4.

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

The model estimates from the multilevel model analysis for Hypothesis 7 are reported in Table 11. The three-way interaction term was not significant. Thus, Hypotheses 7a – 7c were not supported. Technically, Hypothesis 7d was supported because there was no interaction predicted between Time and Emotion Suppression conditions for the male leader. However, without the occurrence of the complementary interaction between Time and Emotion Suppression for the female leader, support for Hypothesis 7d has little meaning. To provide more insight into the results and the pattern of findings, each of the hypotheses for both types of leader ratings is discussed in more detail below.

For Hypothesis 7a, it was predicted that Leader Effectiveness and Leadership Perception ratings would be higher for the male leader for the emotion suppression compared to the no suppression condition at Time 1 (i.e. after the first business video). To test this prediction within the context of multilevel modeling, a simple slopes test was conducted on the Emotion Suppression effect while constraining the value of Leader Gender to the male leader, and the value of Time to time 1. The resulting simple slopes equation included the Emotion Suppression main effect and all interaction terms with Emotion Suppression (i.e. ES x T, ES x LG, ES x T x LG). The simple slopes test of the Emotion Suppression effect on leader ratings for the male leader and time 1 was not significant. There was no indication that the male leader received higher Leader Effectiveness ratings or Leadership Perception ratings after the first business video.

For Hypothesis 7b, it was predicted that Leader Effectiveness and Leadership Perception ratings would be lower for the female leader for the emotion suppression compared to the no suppression condition at Time 4 (i.e. after the fourth business video). This prediction was tested by conducting a simple slopes test on the Emotion Suppression effect while constraining the

value of Leader Gender to the female leader, and the value of Time to time 4. Thus, the resulting simple slopes equation contained the same terms as the equation tested for Hypothesis 7a, but with different Leader Gender and Time constraints. Similar to Hypothesis 7a, the simple slopes test of the Emotion Suppression effect on leader ratings for the female leader at Time 4 was not significant for both Leader Effectiveness (*simple slope* = .31, p = .04, 95% CI [0.02, 0.61]) and Leadership Perceptions (*simple slope* = .09, p = .12, 95% CI [-0.04, 0.23]). Though the Leader Effectiveness estimate was close, it did not cross the significance threshold of .025 for one-tailed *p*-values.

For Hypothesis 7c, it was predicted that for the female leader, there would be an emotion suppression by time interaction effect due to the diverging pattern of leader ratings over time across the emotion suppression and no suppression conditions. A simple slopes test was conducted to examine the Time x Emotion Suppression interaction effect on Leader Effectiveness ratings while constraining values for Leader Gender to the female leader. The simple slope effect was significant (*simple slope* = .16, *p* = .01, 95% CI [0.03, 0.27]), however, the pattern of divergence was the opposite of what was hypothesized; unexpectedly, ratings of the female leader by those in the emotion suppression condition (*M*=3.90, *SD*=0.49) were higher than the no suppression condition (*M*=3.81, *SD*=0.59). In contrast to findings from the Leader Effectiveness model, for Leadership Perception ratings, the simple slope effect of the interaction between Time x Emotion Suppression was not significant (*simple slope* = .05, *p* = .06).

In comparison to the predictions made for the female leader in Hypothesis 7c, for Hypothesis 7d, an interaction effect between Time and Emotion Suppression was *not* predicted for the male leader due to the expected convergence of leader ratings between the emotion suppression and no suppression conditions. The simple slopes test was conducted on the Time x

Emotion Suppression interaction effect while constraining values for Leader Gender to the male leader. Similar to what was observed for the female leader in Hypothesis 1c, the simple slopes test of the Time x Emotion Suppression interaction effect was significant for Leader Effectiveness (*simple slope* = .13, p = .02, 95% CI [0.02, 0.23]) but not Leadership Perceptions (*simple slope* = .05, p = .07, 95% CI [-0.01, 0.10]). Leader Effectiveness ratings of the male leader was higher for those in the emotion suppression condition (M=3.63, SD=0.59) compared to the no suppression condition (M=3.54, SD=0.76). Contrary to predictions, the trajectory pattern of Leader Effectiveness and Leadership Perception ratings (see Figure 2 and 4 respectively for model-predicted values) for the male leader followed what was observed with the female leader, whereby ratings increased at a higher rate for the emotion suppression condition than the no suppression condition.

Leader Competence and Warmth Ratings

Leader Gender Effect. Hypotheses 1c and 1d also predicted that gender bias in leadership judgments would manifest such that the female leader would have (1c) lower leader competence ratings and (1d) lower leader warmth ratings than the male leader. Independent samples t-tests were conducted to test both hypotheses. Contrary to expectations, results indicated that the female leader had higher leader competency ratings (M=3.93, SD=0.59) compared to the male leader (M=3.43, SD=0.73), t(187) = -5.08, p = .00, Cohen's d = .74. Hypothesis 1c was not supported. The female leader also had higher warmth ratings (M=3.75, SD=0.69) compared to the male leader (M=3.31, SD=0.78), t(187) = -4.15, p = .00, Cohen's d = .60. Although the evidence supported Hypothesis 1d, this finding must be tempered by the fact that the female leader was judged to be more competent than the male leader.

Leader Gender by Emotion Suppression by Perceiver Gender Interaction. Both Hypothesis 4a and 4b were tested using a 2x2x2 between-subjects full factorial ANOVA. Hypothesis 4a predicted a three-way interaction between Leader Gender, Perceiver Gender, and Emotion Suppression, such that within the emotion suppression condition, male participants (relative to female participants) would rate the male leader as more competent than the female leader. For Hypothesis 4b, Leader Warmth was used as the dependent variable instead. Hypothesis 4b predicted a three-way interaction between Leader Gender, Perceiver Gender, and Emotion Suppression such that within the emotion suppression condition, female participants (relative to male participants) would rate the female leader as warmer than the male leader.

For Leader Competence ratings, the three-way interaction was not significant, $F(1, 181) = 1.88$, $p = .17$, partial $\eta 2 = .01$. Thus, Leader Competence ratings did not vary differently across emotion suppression conditions by leader and perceiver gender. Hypothesis 4a was not supported. Given the findings from the independent groups t-test of leader gender effect on Leader Competence ratings, it is not surprising that the Leader Gender main effect was significant, $F(1, 181) = 23.76$, $p = .00$, partial $\eta 2 = .12$, with the female leader being judged as more competent than the male leader.

For Leader Warmth ratings, the three-way interaction was also not significant, with $F(1, 181) = .44$, $p = .51$, partial $\eta 2 = .00$. Thus, Leader Warmth ratings did not vary differently across emotion suppression conditions by leader and perceiver gender. Hypothesis 4b was not supported. Again, a Leader Gender main effect was found, $F(1, 181) = 15.72$, $p = .00$, partial $\eta 2 = .08$, with the female leader rated as more warm than the male leader. Additionally, the Emotion Suppression main effect was significant, $F(1, 181) = 4.35$, $p = .04$, partial $\eta 2 = .02$. A follow-up pairwise comparison was conducted using the Bonferroni correction; those in the

emotion suppression condition provided higher Leader Warmth ratings (*M*=3.62, *SE*=0.73) compared to those in the no suppression condition (*M*=3.41, *SE*=0.78).

Behavior Recognition

Emotion Suppression Effect. Hypothesis 5 predicted an emotion suppression effect on overall recognition behavior accuracy for the leadership behaviors, such that those who suppressed their emotions would have lower accuracy than those who did not suppress their emotions. Initially, behavior recognition accuracy was calculated by obtaining the proportion of correct answers (correctly identifying behaviors that did i.e. true hits, and did not occur i.e. true negatives) from the total number of recognition items. However, recognition accuracy and false alarm rates were highly correlated; recognition accuracy had an r = -.67 with false alarm rates for agentic behaviors and r = -.58 for communal behaviors.

Calculating Sensitivity using A'. Thus, a more formal signal detection measure of accuracy was utilized in the attempt to reduce the dependencies between accuracy and false alarm rates. Drawing from Signal Detection Theory (SDT; Stanislaw & Todorov, 1999), accuracy was operationalized instead as *sensitivity* which represents the ability to distinguish signals (true hits) from noise (false alarms). Although only true hits and false alarms are used in the computation of sensitivity, the true negative rates and false negative rates are reflected given that both are determined by the corresponding true hit rate and false positive rate. The metric of *A'* (Pollack & Normal, 1964) was calculated as a nonparametric measure of sensitivity[2]. *A'* was calculated using formula below:

$$A' = .5 + \frac{(H - F)(1 + H - F)}{4H(1 - F)} \; when \; H \geq F$$

[2] Hypothesis testing was repeated on additional SDT measures such as *d'* and *c*, that measure sensitivity and bias respectively. However, findings were similar to previous analyses hence are not reported in the current study.

$$A' = .5 - \frac{(F-H)(1+F-H)}{4F(1-H)} \; when \; H < F$$

where H represents hit rates and F represents false alarm rates.

A' values range from 0 to 1, with 1 reflecting maximum sensitivity/accuracy. Compared to recognition accuracy, the sensitivity measure was less correlated to false alarm rates. Sensitivity had an r = -.60 with false alarm rates for agentic behaviors and r = -.50 for communal behaviors.

Hypothesis 5 was tested with a 1-tailed independent samples t-test; findings indicated that participants in the emotion suppression condition were less accurate in recognizing the leadership behaviors (M=.68, SD=0.17) compared to those in the no suppression condition (M=.72, SD=0.13), using Welch's t-test correction for unequal variances with t(184.35) = 2.22, p = .01, Cohen's d = .32. Hence, Hypothesis 5 was supported.

Leader Gender by Emotion Suppression Interaction. Hypothesis 6a predicted a two-way interaction between Leader Gender and Emotion Suppression on false alarm rates, such that those in the emotion suppression condition would have a greater false alarm rate for agentic behaviors for the male leader compared to the female leader. The tested 2x2 full factorial ANOVA on the agentic false alarm rate did not yield a significant two-way interaction, F(1, 185) = 0.001, p = .97, partial $\eta 2$ = .00. Hypothesis 6a was not supported. Leader Gender was the only significant main effect in the model, F(1, 185) = 4.59, p = .03, partial $\eta 2$ = .02. A follow-up pairwise comparison using Bonferroni's correction indicated that the false alarm rate for agentic behaviors was higher for the female leader (M=.44, SE=0.03) compared to the male leader (M=.35, SE=0.03).

For Hypothesis 6b, it was predicted that those in the emotion suppression would have a greater false alarm rate for communal behaviors for the female leader compared to the male

leader. The 2x2 ANOVA for the false alarm rate for communal behaviors did not reveal a significant two-way interaction effect, $F(1, 185) = 0.008$, $p = .93$, partial $\eta 2 = .00$. Hypothesis 6b was also not supported. However, the main effects in the model were significant for both Leader Gender $F(1, 185) = 6.08$, $p = .02$, partial $\eta 2 = .03$ and Emotion Suppression $F(1, 185) = 5.17$, $p = .02$, partial $\eta 2 = .03$. Follow-up pairwise comparisons using Bonferroni's correction indicated that false alarm rates for communal behaviors were higher for the female leader (M=.65 SE=0.03) than the male leader (M=.56, SE=0.02) and for the emotion suppression condition (M=.64, SE=0.02) than the no suppression condition (M=.56, SE=0.03).

Research Questions

Three research questions were proposed to further understand how recognition accuracy of the leader behaviors in the video would vary by both leader and perceiver gender. Research Question 1 pertained to recognition accuracy of leader behaviors carried out by the female leader, and if accuracy would be higher for agentic behaviors because they are schema-inconsistent compared to communal behaviors which would be schema-consistent. A paired t-test was carried out for individuals in the female leader condition (N = 93). Results showed that sensitivity scores (as measured by *A'*) were higher for agentic behaviors (M=0.73, SD=0.18) compared to the no suppression condition (M=0.57, SD=0.24), with $t(92) = 5.48$, $p = .00$, Cohen's $d = .67$. The ability to correctly identify behaviors that occurred in the videos from those that did not was higher for agentic behaviors by the female leader than communal behaviors. However, a closer examination of sensitivity scores for agentic and communal behaviors across both leader gender conditions revealed an overall greater accuracy for agentic behaviors (M=0.75, SD=0.20) relative to communal behaviors (M=0.60, SD=0.24). The paired t-test was significant at $t(188) = 7.39$, $p = .00$, Cohen's $d = .79$. It is difficult to determine if the higher

recognition accuracy rate for agentic behaviors by the female leader was influenced by the inconsistency of the behaviors with the female leader schema, resulting in stronger memory traces, or if agentic behaviors, in general, were easier to remember and identify compared to communal behaviors.

For Research Question 2, it was asked if false alarm rates for agentic behaviors would vary by both leader and perceiver gender, such that male (relative to female) participants would have greater false alarm rates for the male compared to the female leader. A 2x2 full factorial ANOVA was conducted with false alarm rates for agentic behaviors as the dependent variable. The interaction term between leader and perceiver gender was not significant $F(1, 185) = 1.59$, $p = .21$. Leader Gender was the only significant effect in the model, with $F(1, 185) = 4.91$, $p = .03$, partial $\eta 2 = .03$. A follow-up pairwise comparison using Bonferroni's correction demonstrated that false alarm rates were greater for the female leader (M=0.44, SE=0.03) compared to the male leader (M=0.35, SE=0.03). The findings suggest that false alarm rates did not vary by the participant's gender at different rates for the male or female leader. Surprisingly, agentic behaviors were falsely identified for the female leader by both male and female participants.

Research Question 3 was regarding the interaction between leader and perceiver gender on false alarm rates for communal behaviors; specifically, if female (relative to male) participants would have higher false alarm rates for the female leader compared to the male leader. The tested 2x2 full factorial ANOVA did not reveal a significant interaction effect $F(1, 185) = 0.29$, $p = .87$. However, Leader Gender was significant $F(1, 185) = 6.02$, $p = .02$, partial $\eta 2 = .03$. A follow-up pairwise comparison with Bonferroni's correction revealed a higher false alarm rate for the female leader (M=0.65, SE=0.03) compared to the male leader

(*M*=0.56, *SE*=0.03). Similar to findings for agentic behaviors, there was no difference in false alarm rates by the perceiver's gender; instead, there was a higher false alarm rate overall for the female leader than the male leader.

Exploratory Analyses

Additional exploratory analyses were conducted to better understand the unexpected but dominant female leader effect that was observed in both the Leader Effectiveness and Leadership Perception ratings collected after each of the four business videos and the Leader Competence and Warmth ratings measured after all four videos. In particular, it was examined if participants' current or past experience of having a female supervisor at work potentially influenced their judgments of the female leader compared to the male leader. 110 participants (63 males, 47 females) reported working with a female supervisor, either currently or in the past, while 79 participants (26 males, 53 females) did not. Thus, a majority of male participants (70.8%) did/do have a female supervisor while the split was more balanced for female participants.

The correlations between the female supervisor variable and all four leader judgments were examined. The female supervisor variable was only significantly correlated with Leader Effectiveness at time 1 (i.e. after the first business video) (r = -.17) and with Leader Warmth ratings (r = -.17). These negative correlations can be interpreted such that individuals who did not have a female supervisor currently or in the past provided higher Leader Effectiveness and Leader Warmth ratings than those who did. Since it brings little meaning to interpret the relationship between the female supervisor variable and leader effectiveness ratings for time 1 without considering the other three leader effectiveness ratings after subsequent videos, the discussion is focused on the Leader Warmth ratings instead.

To further probe the influence of the female supervisor variable on Leader Warmth, it was entered as a covariate in the 2 (Leader Gender) x2 (Emotion Suppression) x2 (Perceiver Gender) full-factorial ANOVA that was tested for Hypothesis 4b. Although the female supervisor was a significant covariate in the model $F(1, 180) = 4.20$, $p = .04$, partial $\eta 2 = .03$, the pattern of findings did not change; both Leader Gender ($F(1, 180) = 15.02$, $p = .00$, partial $\eta 2 = .08$) and Emotion Suppression ($F(1, 180) = 4.86$, $p = .03$, partial $\eta 2 = .03$) remained significant main effects in the model. This suggests that while past or current experience working with a female supervisor is an important variable to consider, it did not account for the leader gender effect observed. Furthermore, perceiver gender was still not a significant predictor in the model, indicating that experience with a female supervisor was not the reason leader warmth ratings did not differ by the perceiver's gender. Finally, the female supervisor variable did not provide additional insight as to why the female leader was rated higher on Leader Effectiveness, Leadership Perceptions, and Leader Competence, contrary to predictions.

Discussion

The purpose of the current study was to investigate bias in leadership ratings against female leaders compared to male leaders, and if that bias varies as a function of both the perceiver's gender and engagement in emotion suppression. Emotion suppression was used as the depleting method to induce ego depletion, whereby an ego-depleted individual was expected to be less able to carry out effective self-regulation and engage in effortful processing when making judgments. Leadership judgments were measured through four different ratings; leadership effectiveness and leadership perceptions ratings were collected at four different time points throughout the study, and leader competence and warmth ratings were collected one time at the end of the study. In addition, behavior recognition accuracy of agentic and communal

leadership behaviors was measured to assess if recognition accuracy of male/female leader behaviors would differ due to emotion suppression. Due to the mismatch between the female gender stereotype and the leader prototype, it was hypothesized that the female leader would be evaluated more negatively than the male leader; ego depletion was predicted to amplify these biases such that ego-depleted individuals would provide relatively higher levels of biased leader judgments and recognition accuracy compared to non-ego-depleted individuals. Furthermore, perceiver gender was investigated as a potential moderating factor that could affect how leadership information is processed and subsequent leadership judgments are made. Specifically, it was examined if male perceivers would rate the male/female leader based on an agentic-oriented understanding of leadership while female perceivers would be guided by a communal-oriented understanding of leadership. To my knowledge, this study is the first to examine gender bias in leadership evaluations through the lens of ego depletion. Another contribution of the study is the combined use of repeated measures and overall leadership ratings to better understand how and when ego depletion effects manifest in leadership judgments.

Leader Gender Main Effect

A pattern of results that persisted through most findings in this study was the unexpected leader gender effect, where the female leader was rated more favorably in all four leader judgments that were collected both throughout the study and overall, at the end of the study. The female leader was rated as more effective, more leader-like, more competent, and warmer than the male leader. The female leader was given higher leader ratings by both male and female perceivers, and across both the no suppression and emotion suppression conditions. This finding was contrary to the foundational assumption that this study was based on; that is, there exists a bias against female leaders, and they would be evaluated more negatively compared to male

leaders. In addition to that, the leader gender main effect was also observed with behavior recognition ratings. Recognition accuracy did not change across what was considered schema-consistent or schema-inconsistent behaviors for male and female leaders; instead, false alarm rates were higher for both agentic and communal behaviors for the female leader compared to the male leader.

The study results are a departure from the existing research literature that highlights prejudices faced by female leaders due to the incongruence between the communal nature of female stereotypes and the agentic-oriented stereotypes of a successful leader (e.g., Eagly & Karau, 2002; Eagly & Carli, 2007; Rudman et al., 2012; Schein, 1973). Though speculative, several possible explanations for the female leader effect are presented to better understand this pattern of findings.

Changing Nature of Gender and Leader Stereotypes

The first explanation for this pattern of findings is that perhaps gender and leader stereotypes are changing. There may be more overlap between the female stereotype and expectations of a successful leader compared to what has been reported in past studies. Duehr and Bono (2006) found some evidence of this shift occurring; compared to studies conducted in the 70s and the 80s that demonstrated a “think-manager-think-male" effect, both male and female managers in the 2006 study described women managers as being similar to a successful manager in general. Furthermore, participants in the Duehr and Bono (2006) study also rated women in general as more agentic when compared to male managers from Heilman et al.’s (1989) findings. In a more recent study on gender stereotypes, Hentschel et al. (2019) found that female (but not male) raters rated men and women similarly on leadership competence. Not only are gender stereotypes possibly changing, but the nature of leader prototypes could be shifting

too. Powell and Butterfield (2015) conducted a longitudinal study where they measured how masculine/feminine individuals' descriptions of a good manager evolved using the short form of the Bem Sex-Role Inventory (short BSRI; Bem, 1981). Across four studies over four decades, the researchers found that leader prototypes became increasingly "undifferentiated" and had a reduced emphasis on both masculine and feminine traits. Hence, as the definitions of gender stereotypes and leader prototypes become more nuanced, there is a need to consider additional explanations for differences in leadership evaluations beyond the role incongruity argument.

Additionally, research demonstrating the changing nature of gender stereotypes and leader expectations could also be indicative of an overall generational shift in how gender and leadership are viewed. The current sample was comprised exclusively of Millennials (born 1982 – 1996; Dimock, 2019) and Generation Zers (born after 1996; Dimock, 2019) who have grown up with an increasing amount of representation of female leadership, relative to experiences of previous generations. This is reflected by slightly more than half the sample (N=110) who reported having worked with a female supervisor currently or in the past. Thus, it is likely that they would be generally more open to the idea of women in authority. Moreover, while there has not been sufficient research on Generation Z attitudes, Millennials have been shown to endorse more egalitarian gender roles and approve of women having equal roles in the workplace and of working mothers in general (Donnelly et al., 2016).

Different Motivation Underlying Ratings of a Male/Female Leader

However, even if gender differences in leadership prototypes are decreasing, it is highly unlikely that such changes have resulted in females being perceived as more leader-like and more effective than male leaders. As such, changes in leader prototypes by gender cannot fully explain the powerful reversal of the hypothesized leader gender effect in the current study. As

has been seen in laboratory studies of racial bias where non-conservative Whites show bias in favor of Blacks compared to Whites (e.g., Byrd et al., 2015; Nail et al., 2003), this “bending over backward” effect may be migrating to experimental research on gender bias. Similar to how individuals do not want to appear racially bigoted and believe themselves to be aligned with personal egalitarian ideals, it is possible that the motivation to avoid displaying any semblance of gender bias pushed participants to rate the female leader more positively than perhaps initially perceived. In contrast, the other participants in the male leader condition likely did not experience the same motivation to increase their original ratings in favor of the male leader.

Another potential explanation for why the male leader received poorer evaluations compared to the female leader is the nature of “The Office” videos used for the T1 task. The various scenes portrayed from “The Office” were highly humorous and likely to deplete self-regulatory resources of those who suppressed their emotions while watching them; however, an unintended consequence of using these clips is that participants could have been primed by the funny but ineffective leadership displayed by the male protagonist (i.e. Michael Scott played by Steve Carell). This may have influenced how participants perceived and evaluated the male leader from the business videos. To discern if “The Office" videos did impact subsequent leader ratings, a follow-up study should be conducted using a different set of clips for the T1 task that is also comedic but portrays an ineffective female leader instead (e.g., Parks and Recreation, VEEP). If the female leader is rated more negatively than the male leader in the follow-up study, this would suggest that “The Office" videos were a contributing factor to the dominant female leader effect observed in this study.

Additionally, though anecdotal, conversations with a subset of participants who were willing to share their reactions and personal experiences after the experiment provided further

insight into why the female leader was rated more favorably than the male leader. Some individuals expressed dislike for the male leader who reminded them of other male leaders with whom they had interacted in the past. Based on their experience, the male leaders adopted a laissez-faire approach to leadership, consistently delegating work to others without providing clear direction on the next steps forward. This aligns with research conducted by Ritter and Lord (2007) who found that evaluations of new leaders could be influenced by mental representations of previous leaders if the target leader closely resembled a past leader. Thus, these male leaders could have functioned as negative exemplars and unfairly biased leader judgments against the male leader observed in the business videos compared to the female leader.

Business Video Stimuli

A third possible factor contributing to the observed female leader effect could be due to the portrayal of leadership in the business videos themselves. There are two unpublished dissertation studies that also used the same Hanges et al. (1997) business videos to assess leadership differences between the male and female leader. Using the same four vignettes also utilized in this study, Shah (2017) found that the female leader was rated more warmly than the male leader, a finding that was predicted and replicated in this study. Shah (2017) also predicted that the male leader would be rated more competent than the female leader, which corresponds with what was hypothesized in this study. However, in contrast to the current findings, no differences in leader competence ratings were found between the male and the female leader (Shah, 2017). In the other dissertation study, Robson (2007) used the Hanges et al. (1997) business videos to examine leadership perceptions of the female and male leader. Though different vignettes within the video set were used, the female leader effect was still found in this study whereby the female leader was given higher leader effectiveness ratings than the male

leader (Robson, 2007). Considering the lack of a male leader effect for leader competence (Shah, 2017) and the rating of the female leader as more effective than the male leader (Robson, 2007), the results from both studies suggest that finding a female leader effect (or in other words, lack of a preference for the male leader) within this set of business videos is not without precedent.

Furthermore, the videos by Hanges et al. (1997) were constructed so that the leader displayed an almost equal amount of agentic and communal behaviors across the four vignettes. Past research shows that negativity against female leaders can be mitigated when their communality is explicitly highlighted (Heilman & Okimoto, 2007). In this study, the female leader was given high ratings on both competence and warmth. Hence, being recognized for her communal leadership behaviors could have reduced penalties the female leader could have incurred for engaging in agentic behaviors, thus influencing perceptions that she was a good leader overall.

Role of Perceiver Gender

Although the perceiver's gender did not play as significant of a role in findings compared to the leader's gender, there was some preliminary support for the prediction that female perceivers would rate the female leader as more effective and leader-like than the male leader, relative to male perceivers. In contrast, differences in leader effectiveness and leadership perception ratings across the male or female leader did not occur for male perceivers. This corresponds with Paustian-Underdahl et al.'s (2014) finding that female leaders were evaluated as more effective than male leaders when rated by a majority-female sample. Parallel to findings from the current study where the male perceivers did not show a preference for the male leader, male leaders were not perceived as more effective than female leaders when rated by a sample consisting of majority males (Paustian-Underdahl et al., 2014). Furthermore, similar to the leader

effectiveness and leadership perceptions measures used in this study, the leadership effectiveness measures included in Paustian-Underdahl et al.'s (2014) meta-analysis were not agentic or communal in nature but reflected general perceptions of effectiveness.

A perceiver gender effect was not observed for both the agentic-oriented (i.e. leader competence) and communal-oriented (i.e. leader warmth) leadership ratings. As discussed before, the female leader was rated as more competent and warm by both male and female perceivers. These findings suggest that for this sample, activated leader prototype networks did not vary by the perceiver's gender; male perceivers did not demonstrate a more masculine understanding of leadership, neither by providing higher competence ratings themselves nor by rating the male leader as more competent. Similarly, female perceivers did not demonstrate a more feminine understanding of leadership by providing higher warmth ratings to the leader compared to male perceivers.

Two research questions were proposed to explore if information processing and recognition differed by perceiver gender for schema-consistent behaviors (i.e. agentic behaviors by the male leader for the male perceiver, communal behavior by the female leader for the female perceiver). However, parallel to findings related to the leader competence and leader warmth ratings, no perceiver gender effect was observed for the false alarm rates for both agentic and communal leadership behaviors. There was no evidence indicating that the recognition of behaviors varied in schema-consistency/inconsistency according to the perceiver's gender. Hence, the findings suggest that individuals did not show gendered expectations for how a leader should behave as their ratings of leader competence and warmth did not vary by their gender or leader's gender. This aligns with Powell and Butterfield's (2015) assertions that leader prototypes are becoming less associated with either masculine or feminine traits.

Ego Depletion Effect

Despite doubt in the broader self-control/self-regulation literature regarding the robustness of the ego depletion effect, the emotion suppression manipulation was successful in the current study; the ego depletion effect was observed and affected individuals' processing capabilities. As hypothesized, compared to the no suppression condition, those who exerted self-control and suppressed their emotions were less accurate in their behavior recognition ratings. More specifically, the sensitivity score for overall recognition behavior accuracy was lower for ego-depleted individuals relative to non-ego-depleted individuals, indicating that those who were depleted were less able to distinguish between the leadership behaviors that happened in the business videos compared to the ones that did not. In line with expectations, emotion suppression negatively impacted the ability to process all situational details of the business videos, leading to decreased memory of the leadership behaviors compared to the non-depleted individuals. This pattern was also observed for recognition of communal behaviors, where ego-depleted individuals had higher false alarm rates for communal behaviors in the videos compared to non-ego-depleted individuals.

Additionally, those who were ego-depleted were more lenient in their leadership ratings compared to those who were not ego-depleted. This was observed for Leader Warmth ratings, whereby participants who suppressed their emotions provided higher warmth ratings to both the male and female leader compared to the control condition. Instead of expending the necessary resources to engage in effortful thought processing, ego-depleted individuals seemed to show less discernment when providing their ratings and were less strict in their leader evaluations. Interestingly, this pattern of increased leniency and lower accuracy was not observed for leader competence ratings or false alarm rates for agentic behaviors respectively. Compared to the

concrete, task-oriented behaviors that reflect leader competence and agentic behavior recognition items, leader warmth and communal behavior recognition items are relational-oriented and more abstract in nature. It is likely that the greater room for subjective interpretation of the communal-oriented leadership items allowed for the ego depletion effects to manifest more clearly compared to agentic-oriented items.

The current study also examined ego depletion effects over time by tracking the trajectory of leader ratings over four time points. An interesting finding was that the rate of increase (i.e. slope) in leader effectiveness ratings for the female leader was higher for ego-depleted than non-ego-depleted individuals. Assuming that this sample had an underlying positive bias for the female leader and was neutral about the male leader (as reflected by the preference for the female leader but still overall above-midpoint ratings for both the male and female leader), then the ego depletion effect can be interpreted as biasing ratings such that ego-depleted individuals became increasingly lenient over time in evaluating the female leader's effectiveness compared to non-ego-depleted individuals. However, this finding should be interpreted with caution as it was only observed for leader effectiveness ratings for the female leader. A similar pattern of ratings was not found for leadership perception ratings for the female leader nor for leader effectiveness/leadership perception ratings for the male leader.

Limitations

Although the initial goal was to recruit a homogenous sample consisting of Psychology undergraduate students, recruitment efforts were soon expanded to achieve the target subgroup sample sizes. The final sample was comprised of individuals from a wide range of ages, fields of study, and national origin. Furthermore, compared to the females in the sample, the male participants were more diverse in terms of their geographic location within the United States and

work experience. While it is not a limitation of the study per se, the heterogeneity of the overall sample and the differing levels of diversity between the male and female subgroups are characteristics of the current study sample that do not align with the original recruitment goals. Additional exploratory analyses were carried out to compare findings across the males who were recruited from the university (*N*=55) and those who were not (*N*=34), but no meaningful differences were observed across the subgroups. Thus, it is difficult to determine if the study findings are an accurate and generalizable representation of individuals' perceptions of female versus male leaders, or should be treated as "noise" and the inadvertent outcome of having this specific sample of participants.

Another study limitation was the audio and video quality of the Hanges et al. (1997) business videos. Although efforts were made to digitally enhance the videos prior to the study, participants still reported having some difficulty in comprehending some parts of the videos. While they could still get the overall gist of the conversations, this could have been a factor that potentially impacted participants' subsequent evaluations of the leader.

Additionally, covariate analyses to control for potentially relevant attitudinal variables could not be carried out as originally anticipated. A separate online study was planned to collect relevant information such as individuals' attitudes on sexism and gender roles in society, as well as their self-reported gender schemas. Participants were supposed to participate in this online study before having the option of signing up for the current study. However, due to the previously discussed recruitment challenges, the decision was made to just focus on recruiting for the focal study; nonetheless, data collection for the online study remained active for anyone interested in participating. The final sample size of participants who took part in both studies (*N*=58) was not large enough to allow for any meaningful covariate analyses.

Finally, due to the COVID-19 pandemic, the experiment was not carried out in person in the laboratory as was initially planned. Instead, the procedure was modified to be carried out completely online through Qualtrics and Zoom. Participants were monitored through Zoom as they were completing the study; however, it is impossible to know how engaged they were and if their participation effort is comparable to what it would have been if they were physically in the laboratory with the experimenter. Moreover, conducting this study virtually modified the original plan to collect participants' continuous ratings of the leader throughout the four business videos. This would have provided deeper insight into participants' leadership perceptions throughout the videos compared to only measuring their leader ratings at the end of each of the four videos.

Implications and Future Directions

The current study results carry several implications. First, the positive evaluations of the female leader compared to the male leader suggests that perceptions of females in leadership may be changing. This finding is in juxtaposition with the existing research literature on gender and leadership, which highlights the negative biases in leadership evaluations faced by female leaders relative to male leaders. To ascertain if the positive female leader effect was unique to this study or an indication of a general shift in how female leaders are perceived, it is important to revisit the research on gender stereotypes and leader prototypes.

Foremost is the need to re-examine what constitutes effective leadership, as it is defined in the current times. Offermann and Coats (2018) recently replicated Offermann et al.'s (1994) study to examine the stability of individuals' implicit leadership theories (ILT). Based on data that was collected in 2014, Offermann and Coats (2018) found that the list of attributes generated by participants that they believed to be characteristic of a leader was mostly similar to ILTs measured back in 1994. Notably, masculinity was identified as an ILT factor across both the

1994 and 2018 studies. However, considering the societal events and changes that have occurred especially in the last few years (i.e. election of the first female vice president of the United States, strong leadership shown by female leaders during the COVID-19 pandemic (Sergent & Stajkovic, 2020), greater representation of Millennials and Generation Z'ers in the student/working adult population), it would be worthwhile to engage in additional leader prototype research. In particular, it would be informative to replicate Lord et al.'s (1984) seminal study on leadership categorization theory by measuring the category representativeness and accessibility of each leader attribute and how that affects leadership judgments. Adopting an inductive approach to determine which attributes are central to individuals' perceptions of an ideal leader would lend insight on whether leader prototypes have changed, and by how much. Furthermore, the gendered expectations tied to each attribute in a leader prototype should be assessed; what has been defined as more agentic/masculine or communal/feminine leader behaviors may no longer be associated with a particular gender, as suggested by Powell and Butterfield's (2015) research.

In the business videos used in the current study, race could not be studied as a relevant factor as everyone featured in the videos was White, including both the male and female leader as well as the two other business managers. As the representation of leaders from diverse backgrounds hopefully continues to increase, it is important to specifically address the unique challenges that leaders may face as the outcome of their various subgroup identities. Unfortunately, the existing research is sparse concerning leadership and the intersectionality of race *and* gender identities (e.g., Chávez & Mitchell, 2020). Thus, an interesting and increasingly relevant avenue of research is to explore how the intersection of a leader's multiple identities could influence the leader's perceived effectiveness ratings. Intersectionality is a

complex issue; comparing evaluations of a Black male leader with a White male leader is different than comparing the same Black male leader with an Asian female leader. Assessment of the leader might also vary according to the race and gender identities of the perceiver too. Due to the highly contextual and nuanced nature of these research questions, future research efforts could benefit from a more holistic and person-centered approach to addressing this topic.

In terms of ego depletion research, this study demonstrates that emotion suppression is a viable method for depleting self-control resources. The process of evaluating a leader is an effortful task that requires regulatory resources, and ego depletion negatively affected individuals' ability to provide accurate ratings of the leader. When tired or depleted, individuals would not likely provide ratings that are an accurate representation nor informative of the leader's performance. According to the self-control strength model, self-control resources can be repleted after a certain amount of time has passed (Baumeister & Vohs, 2016b; Muraven et al., 1998). Hence, it would be interesting to examine if the accuracy of leader ratings would improve if there was a larger interval of time between the T1 depleting task and the T2 task when ratings are provided. Additionally, to further probe the robustness of the ego depletion effect, it would be informative to examine if leader ratings would be similarly affected if a cognitive or behavioral-based self-control resource depleting task was used instead of emotion suppression.

Conclusion

The general finding that the female leader was evaluated more favorably than the male leader in all aspects of leadership made it impossible to assess *a priori* predictions about gender bias against female leaders. As a result of the increased exposure to female leaders both in personal experiences and in society in general, individuals may be changing how they perceive and evaluate female leaders. Thus, it is important to revisit leader prototype and gender

stereotype research to determine if the current findings are a signal that perceptions are truly changing, and the incongruence between the female stereotype and expectations of a successful leader may not as large as previously studied. Instead of assuming the nature of pre-existing biases, it would be informative to directly measure individuals' perceptions of effective leadership and assess if that is associated with a specific gender.

While the perceiver's gender was not found to be a significant factor that influenced perceptions of the leader's effectiveness by gender, there are other contextual variables that should be investigated and potentially controlled for. For example, anecdotal conversations with some participants indicate that their individual experiences as a follower could have influenced their evaluations of a leader in general. The negative/positive experiences that individuals have with a male/female leader could shed more insight into how leader judgments are made.

Further, this study demonstrates a reversal of the expected bias in leader ratings in favor of the female leader. It may be that experimental gender bias research has reached a tipping point where rendering judgments about females provides a strong demand characteristic to avoid being perceived as sexist. As such, the goal of not demonstrating gender bias against the female leader creates a response bias that supplants faithful responses to experimental manipulation.

Additionally, an ego depletion effect was observed in this study; leadership judgments of ego-depleted individuals were more lenient and less accurate than non-ego-depleted individuals. Though the nature of the bias was in the opposite direction where the female leader was rated was evaluated more positively (not more negatively) than the male leader, the study findings still demonstrate that being in a state of ego-depletion allows biases to manifest more clearly in subsequent leader ratings. Thus, there should be an awareness of the limits of our self-control

and processing capabilities; our ability to provide unbiased and informative evaluations of leaders decreases when we are depleted.

In conclusion, the findings were unexpected but provide some cautious optimism that the negative bias against female leaders may be reducing. It is necessary to conduct additional research to determine if the same pattern of results would be replicated in future samples. Regardless, the current study demonstrates that bias in leadership evaluations still exists, as the male and the female leader were evaluated differently despite behaving in exactly the same way. Not only are we are prone to allowing our biases to negatively affect the objectivity of our judgments, but our susceptibility to bias increases when we lack the self-control resources to render fair and accurate assessments of others.

References

Abraham, E., & Hauenstein, N. (2020). *Not if, but when do we show bigotry? Ego depletion, moderators, and boundary conditions.* [Manuscript submitted for publication]. Psychology Department, Virginia Tech.

Ayman, R., Korabik, K., & Morris, S. (2009). Is transformational leadership always perceived as effective? Male subordinates' devaluation of female transformational leaders. *Journal of Applied Social Psychology*, *39*(4), 852–879. https://doi.org/10.1111/j.1559-1816.2009.00463.x

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252–1265. https://doi.org/10.1037/0022-3514.74.5.1252

Baumeister, R. F., & Vohs, K. D. (2016a). Misguided effort with elusive implications. *Perspectives on Psychological Science, 11*(4), 574-575. https://doi.org/10.1177/1745691616652878

Baumeister, R. F., & Vohs, K. D. (2016b). Strength model of self-regulation as limited resource: Assessment, controversies, update. In M. O. James & P. Z. Mark (Eds.), *Advances in experimental social psychology* (Vol. 54, pp. 67-127). San Diego, CA: Academic Press.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351-355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Beal, D. J., & Gabriel, A. S. (2018). Looking within: An examination, combination, and extension of within-person methods across multiple levels of analysis. *The Handbook of*

Multilevel Theory, Measurement, and Analysis., 305–327. https://doi.org/10.1037/0000115-014

Beal, D. J., & Trougakos, J. P. (2013). *Episodic intrapersonal emotion regulation: Or, dealing with life as it happens.* In A. A. Grandey, J. M. Diefendorff, & D. E. Rupp (Eds.), *Organization and management. Emotional labor in the 21st century: Diverse perspectives on emotion regulation at work* (p. 31–55). Routledge/Taylor & Francis Group.

Bem, S. L. (1981). *A manual for the Bem Sex Role Inventory*. Palo Alto, CA: Consulting Psychologist.

Braun, S., Peus, C., & Frey, D. (2018). Connectionism in action: Exploring the links between leader prototypes, leader gender, and perceptions of authentic leadership. *Organizational Behavior and Human Decision Processes*, *149*, 129–144. https://doi.org/10.1016/j.obhdp.2018.10.003

Brescoll, V. L. (2011). Who Takes the Floor and Why: Gender, Power, and Volubility in Organizations. *Administrative Science Quarterly*, *56*(4), 622–641. https://doi.org/10.1177/0001839212439994

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition.* (Vol. I, pp. 1-36). Erlbaum.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-Role Stereotypes: A Current Appraisal. *Journal of Social Issues*, *28*(2), 59–78. https://doi.org/10.1111/j.1540-4560.1972.tb00018.x

Brown, D., Marchioro, C., Tan, J., & Lord, R. (1998). Individual difference variables and leadership perceptions: Linear or catastrophic relationship? Paper presented at the 13th

Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Burns, K. C., Isbell, L. M., & Tyler, J. M. (2008). Suppressing emotions toward stereotyped targets: The impact on willingness to engage in contact. *Social Cognition*, *26*(3), 276–287. https://doi.org/10.1521/soco.2008.26.3.276

Byrd, D. T. L., Hall, D. L., Roberts, N. A., & Soto, J. A. (2015). Do Politically Non-conservative Whites "Bend Over Backwards" to Show Preferences for Black Politicians? *Race and Social Problems*, *7*(3), 227–241. https://doi.org/10.1007/s12552-015-9153-6

Carli, L. L., & Eagly, A. H. (2018). Leadership and gender. In J. Antonakis & D. V. Day (Eds.), *The nature of leadership*. (pp. 244-271). SAGE Publications, Inc.

Catalyst. (2020, November). *Women CEOs of the S&P 500.* https://www.catalyst.org/research/women-ceos-of-the-sp-500/

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General*, *144*(4), 796–815. https://doi.org/10.1037/xge0000083

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. Springer-Verlag.

Chávez, K., & Mitchell, K. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, *53*(2), 270-274. doi:10.1017/S1049096519001744

Côté, S. (2005). A social interaction model of the effects of emotion regulation on work strain. *Academy of Management Review*, *30*(3), 509–530. https://doi.org/10.5465/AMR.2005.17293692

Crocker, J., Hannah, D. B., & Weber, R. (1983). Person memory and causal attributions. *Journal of Personality and Social Psychology*, *44*(1), 55–66. https://doi.org/10.1037/0022-3514.44.1.55

Cronshaw, S. F., & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. *Journal of Applied Psychology*, *72*(1), 97–106. https://doi.org/10.1037/0021-9010.72.1.97

Dasborough, M. T., & Ashkanasy, N. M. (2002). Emotion and attribution of intentionality in leader-member relationships. *Leadership Quarterly*, *13*(5), 615–634. https://doi.org/10.1016/S1048-9843(02)00147-9

Deal, J.J., Stevenson, M.A. Perceptions of Female and Male Managers in the 1990s: Plus ça change . . .. *Sex Roles* **38,** 287–300 (1998). https://doi.org/10.1023/A:1018741318216

DeRue, S. S., Nahrgang, J. D., & Ashford, S. J. (2015). Interpersonal perceptions and the emergence of leadership structures in groups: A network perspective. *Organization Science*, *26*(4), 1192–1209. https://doi.org/10.1287/orsc.2014.0963

Demaree, H. A., Schmeichel, B. J., Robinson, J. L., Pu, J., Everhart, D. E., & Berntson, G. G. (2006). Up-and down-regulating facial disgust: Affective, vagal, sympathetic, and respiratory consequences. *Biological Psychology*, *71*(1), 90-99. https://doi.org/10.1016/j.biopsycho.2005.02.006.

Devine, P. G., & Elliot, A. J. (1995). Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited. *Personality and Social Psychology Bulletin*, *21*(11), 1139–1150. https://doi.org/10.1177/01461672952111002

Diefendorff, J. M., & Gosserand, R. H. (2003). Understanding the emotional labor process: A control theory perspective. In *Journal of Organizational Behavior* (Vol. 24, Issue 8, pp. 945–959). John Wiley & Sons, Ltd. https://doi.org/10.1002/job.230

Diefendorff, J. M., & Richard, E. M. (2003). Antecedents and consequences of emotional display rule perceptions. *Journal of Applied Psychology*, *88*(2), 284–294. https://doi.org/10.1037/0021-9010.88.2.284

Dimock, M. (2019, January 17). *Defining generations: Where millennials end and generation z begins.* Pew research center. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

Ditto, P. H., & Lopez, D. F. (1992). Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

Donnelly, K., Twenge, J. M., Clark, M. A., Shaikh, S. K., Beiler-May, A., & Carter, N. T. (2016). Attitudes Toward Women's Work and Family Roles in the United States, 1976–2013. *Psychology of Women Quarterly*, *40*(1), 41–54. https://doi.org/10.1177/0361684315590774

Duehr, E. E., & Bono, J. E. (2006). Men, women, and managers: are stereotypes finally changing? *Personnel Psychology*, *59*(4), 815-846. https://doi.org/10.1111/j.1744-6570.2006.00055.x

Dunning, D., & Sherman, D. A. (1997). Stereotypes and Tacit Inference. *Journal of Personality and Social Psychology*, *73*(3), 459–471. https://doi.org/10.1037/0022-3514.73.3.459

Eagly, A. H. (2007). Female leadership advantage and disadvantage: Resolving the contradictions. *Psychology of Women Quarterly*, *31*(1), 1–12. https://doi.org/10.1111/j.1471-6402.2007.00326.x

Eagly, A. H., & Carli, L. L. (2007). Women and the labyrinth of leadership. *Harvard Business Review*, *85*, 62–67.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the Effectiveness of Leaders: A Meta-Analysis. *Psychological Bulletin*, *117*(1), 125–145. https://doi.org/10.1037/0033-2909.117.1.125

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the Evaluation of Leaders: A Meta-Analysis. In *Psychological Bulletin* (Vol. 111, Issue 1, pp. 3–22). https://doi.org/10.1037/0033-2909.111.1.3

Eberly, M. B., & Fong, C. T. (2013). Leading via the heart and mind: The roles of leader and follower emotions, attributions and interdependence. *Leadership Quarterly*, *24*(5), 696–711. https://doi.org/10.1016/j.leaqua.2013.05.003

Eden, D., & Leviatan, U. (1975). Implicit leadership theory as a determinant of the factor structure underlying supervisory behavior scales. *Journal of Applied Psychology*, *60*(6), 736–741. https://doi.org/10.1037/0021-9010.60.6.736

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System.* Research Nexus division of Network Information Research Corporation.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275-298. https://doi.org/10.1146/annurev-psych-122216

Epitropaki, O., & Martin, R. (2004). Implicit Leadership Theories in Applied Settings: Factor Structure, Generalizability, and Stability Over Time. *Journal of Applied Psychology*, *89*(2), 293–310. https://doi.org/10.1037/0021-9010.89.2.293

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*, 1149-1160.

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, *23*(C), 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Foti, R. J., & Hauenstein, N. M. (1993). Processing demands and the effects of prior impressions on subsequent judgments: Clarifying the assimilation/contrast debate. *Organizational Behavior and Human Decision Processes*, *56*(2), 167-189. https://doi.org/10.1006/obhd.1993.1050

Foti, R. J., Knee, R. E., & Backert, R. S. G. (2008). Multi-level implications of framing leadership perceptions as a dynamic process. *Leadership Quarterly*, *19*(2), 178–194. https://doi.org/10.1016/j.leaqua.2008.01.007

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a Within-Subject, Repeated-Measures Ego-Depletion Paradigm: Inconsistent Results and Future Recommendations. *Social Psychology*, *49*(5), 271–286. https://doi.org/10.1027/1864-9335/a000348

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is Ego Depletion Real? An Analysis of Arguments. *Personality and Social Psychology Review*, *23*(2), 107–131. https://doi.org/10.1177/1088868318762183

Gabriel, A. S., Diefendorff, J. M., Bennett, A. A., & Sloan, M. D. (2017). It's About Time: The Promise of Continuous Rating Assessments for the Organizational Sciences. *Organizational Research Methods*, *20*(1), 32–60. https://doi.org/10.1177/1094428116673721

Gabriel, A. S., Koopman, J., Rosen, C. C., Arnold, J. D., & Hochwarter, W. A. (2020). Are coworkers getting into the act? An examination of emotion regulation in coworker exchanges. *Journal of Applied Psychology*, *105*(8), 907–929. https://doi.org/10.1037/apl0000473

Gailliot, M. T., Baumeister, R. F., Dewall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., Brewer, L. E., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, *92*(2), 325–336. https://doi.org/10.1037/0022-3514.92.2.325

Gill, M. J. (2004). When information does not deter stereotyping: Prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 619–632. https://doi.org/10.1016/j.jesp.2003.12.001

Glasø, L., & Einarsen, S. (2008). Emotion regulation in leader - Follower relationships. *European Journal of Work and Organizational Psychology*, *17*(4), 482–500. https://doi.org/10.1080/13594320801994960

Gloor, J. L., Morf, M., Paustian-Underdahl, S., & Backes-Gellner, U. (2020). Fix the Game, Not the Dame: Restoring Equity in Leadership Evaluations. *Journal of Business Ethics*, *161*(3), 497–511. https://doi.org/10.1007/s10551-018-3861-y

Gordijn, E. H., Hindriks, I., Koomen, W., Dijksterhuis, A., & Van Knippenberg, A. (2004). Consequences of Stereotype Suppression and Internal Suppression Motivation: A Self-Regulation Approach. *Personality and Social Psychology Bulletin*, *30*(2), 212–224. https://doi.org/10.1177/0146167203259935

Grandey, A. A., & Melloy, R. C. (2017). The state of the heart: Emotional labor as emotion regulation reviewed and revised. *Journal of Occupational Health Psychology*, *22*(3), 407–422. https://doi.org/10.1037/ocp0000067

Gross, J. J. (1998a). Antecedent- and Response-Focused Emotion Regulation: Divergent Consequences for Experience, Expression, and Physiology. *Journal of Personality and Social Psychology*, *74*(1), 224–237. https://doi.org/10.1037/0022-3514.74.1.224

Gross, J. J. (1998b). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271-299. https://doi.org/10.1037/1089-2680.2.3.271

Gross, J. J., & John, O. P. (2003). Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Gross, J. J., & Levenson, R. W. (1993). Emotional Suppression: Physiology, Self-Report, and Expressive Behavior. *Journal of Personality and Social Psychology*, *64*(6), 970–986. https://doi.org/10.1037/0022-3514.64.6.970

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, *106*(1), 95–103. https://doi.org/10.1037/0021-843X.106.1.95

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546-573. https://doi.org/10.1177/1745691616652873

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis. *Psychological Bulletin*, *136*(4), 495–525. https://doi.org/10.1037/a0019486

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing … or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353–363. https://doi.org/10.1177/0361684316634081

Hais, S. C., Hogg, M. A., & Duck, J. M. (1997). Self-Categorization and Leadership: Effects of Group Prototypicality and Leader Stereotypicality. Personality and Social Psychology Bulletin, 23(10), 1087–1099. https://doi.org/10.1177/01461672972310009

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-Based Expectancies: Effects on Information Processing and Social Behavior. *Journal of Social Issues*, *46*(2), 35–60. https://doi.org/10.1111/j.1540-4560.1990.tb01922.x

Hanges, P., Lord, R., Day, D., Sipe, W., Gradwohl, W., & Brown, D. (1997). Leadership and gender bias: Dynamic measures and nonlinear modeling. Conference manuscript for the Society ofIndustrial and Organizational Psychology.

Hansbrough, T. K., Lord, R. G., & Schyns, B. (2015). Reconsidering the accuracy of follower leadership ratings. *Leadership Quarterly*, *26*(2), 220–237. https://doi.org/10.1016/j.leaqua.2014.11.006

Hansbrough, T.K., Lord, R. G., Schyns, B., Foti, R. J., Liden, R. C., & Acton, B. P. (2020). Do you remember ? Rater memory systems and leadership measurement. *Leadership Quarterly*. *https://doi.org/10.1016/j.leaqua.2020.101455*

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*(1), 25–38. https://doi.org/10.1037//0022-3514.37.1.25

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*(4), 3-26.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674. https://doi.org/10.1111/0022-4537.00234

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. http://login.ezproxy.lib.vt.edu/login?url=https://www.proquest.com/scholarly-

journals/sex-stereotypes-do-they-influence-perceptions/docview/1292282002/se-2?accountid=14826

Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has Anything Changed? Current Characterizations of Men, Women, and Managers. *Journal of Applied Psychology*, *74*(6), 935–942. https://doi.org/10.1037/0021-9010.74.6.935

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), 431–441. https://doi.org/10.1037/0021-9010.90.3.431

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, *92*(1), 81–92. https://doi.org/10.1037/0021-9010.92.1.81

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology*, *10*(JAN). https://doi.org/10.3389/fpsyg.2019.00011

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling.* University of California Press.

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, *102*(6), 1318–1335. https://doi.org/10.1037/a0026545

Hogue, M., & Lord, R. G. (2007). A multilevel, complexity theory approach to understanding gender bias in leadership. *Leadership Quarterly*, *18*(4), 370–390. https://doi.org/10.1016/j.leaqua.2007.04.006

Humphrey, R. H., Burch, G. F., & Adams, L. L. (2016). The benefits of merging leadership research and emotions research. *Frontiers in Psychology*, *7*(JUL), 1–12. https://doi.org/10.3389/fpsyg.2016.01022

Johnson, S. E., Mitchell, M. A., Bean, M. G., Richeson, J. A., & Shelton, J. N. (2010). Gender moderates the self-regulatory consequences of suppressing emotional reactions to sexism. *Group Processes and Intergroup Relations*, *13*(2), 215–226. https://doi.org/10.1177/1368430209344867

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. https://doi.org/10.1016/j.obhdp.2007.12.002

Jordan, P. J., & Lindebaum, D. (2015). A model of within person variation in leadership: Emotion regulation and scripts as predictors of situationally appropriate leadership. *Leadership Quarterly*, *26*(4), 594–605. https://doi.org/10.1016/j.leaqua.2015.05.004

Kafetsios, K., Athanasiadou, M., & Dimou, N. (2014). Leaders' and subordinates' attachment orientations, emotion regulation capabilities and affect at work: A multilevel analysis. *Leadership Quarterly*, *25*(3), 512–527. https://doi.org/10.1016/j.leaqua.2013.11.010

Kafetsios, K., Nezlek, J. B., & Vassilakou, Th. (2012). Relationships between leaders' and subordinates' emotion regulation, satisfaction and affect at work. *The Journal of Social Psychology*, *152*(4), 436–457. https://doi.org/10.1080/00224545.2011.632788

Kaiser, R. B., Hogan, R., & Craig, S. B. (2008). Leadership and the Fate of Organizations. *American Psychologist*, *63*(2), 96–110. https://doi.org/10.1037/0003-066X.63.2.96

Kark R., Eagly A.H. (2010) Gender and Leadership: Negotiating the Labyrinth. In: Chrisler J., McCreary D. (Eds.) *Handbook of gender research in psychology.* Springer. https://doi.org/10.1007/978-1-4419-1467-5_19

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), 616–642. https://doi.org/10.1037/a0023557

Kunda, Z., & Thagard, P. (1996). Forming Impressions from Stereotypes, Traits, and Behaviors: A Parallel-Constraint-Satisfaction Theory. In *Psychological Review* (Vol. 103, Issue 2, pp. 284–308). https://doi.org/10.1037/0033-295X.103.2.284

Kwon, I., &Milgrom, E. M. M. (2010). Working for female managers: Gender hierarchy in the workplace. *Stanford Institute for Economic Policy Research Discussion Paper.* Retrieved from https://siepr.stanford.edu/research/publications/working-female-managers-gender-hierarchy-workplace-0

Lord, R. G. (1985). Accuracy in Behavioral Measurement. An Alternative Definition Based on Raters' Cognitive Schema and Signal Detection Theory. *Journal of Applied Psychology*, *70*(1), 66–71. https://doi.org/10.1037/0021-9010.70.1.66

Lord, R. G., Brown, D. J., Harvey, J. L., & Hall, R. J. (2001). Contextual constraints on prototype generation and their multilevel consequences for leadership perceptions. *Leadership Quarterly*, *12*(3), 311–338. https://doi.org/10.1016/S1048-9843(01)00081-9

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. https://doi.org/10.1016/0030-5073(84)90043-6

Lord, R. G., & Maher, K. J. (1993). *Leadership and information processing: Linking perceptions and performance*. Routledge.

Madera, J. M., & Smith, D. B. (2009). The effects of leader negative emotions on evaluations of leadership in a crisis situation: The role of anger and sadness. *Leadership Quarterly*, *20*(2), 103–114. https://doi.org/10.1016/j.leaqua.2009.01.007

Martell, R. F., & Evans, D. P. (2005). Source-monitoring training: Toward reducing rater expectancy effects in behavioral measurement. *Journal of Applied Psychology*, *90*(5), 956–963. https://doi.org/10.1037/0021-9010.90.5.956

Martell, R. F., & DeSmet, A. L. (2001). A diagnostic-ratio approach to measuring beliefs about the leadership abilities of male and female managers. *Journal of Applied Psychology, 86*(6), 1223–1231. https://doi.org/10.1037/0021-9010.86.6.1223

Martijn, C., Tenbult, P., Merckelbach, H., Dreezens, E., & de Vries, N. K. (2002). Getting a grip on ourselves: Challenging expectancies about loss of energy after self-control. *Social Cognition*, *20*(6), 441-460. https://doi.org/10.1521/soco.20.6.441.22978

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*(1), 123–136. https://doi.org/10.1037/0021-9010.84.1.123

Mayer, J.D. & Salovey, P. (1997). What is emotional intelligence? In Salovey, P. and Sluyter, D.J. (eds), Emotional Development and Emotional Intelligence. (pp. 3–31). Basic Books.

McCusker, M.E., Foti, R.J., & Abraham, E. K. (2018). Leadership research methods: Progressing back to process. In R. E. Riggio (Ed.) *What's wrong with leadership?* (pp. 9-40). Routledge.

McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). Social, cultural and cognitive factors in stereotype formation. *Stereotypes as Explanations: The Formation of Meaningful Beliefs about Social Groups*, 1–15. https://doi.org/10.1017/CBO9780511489877

Muraven, M. (2008). Prejudice as Self-Control Failure. *Journal of Applied Social Psychology*, *38*(2), 314–333. https://doi.org/10.1111/j.1559-1816.2007.00307.x

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-Control as Limited Resource: Regulatory Depletion Patterns. *Journal of Personality and Social Psychology*, *74*(3), 774–789. https://doi.org/10.1037/0022-3514.74.3.774

Muthén, L. K., and Muthén, B. O. (2017). Mplus User's Guide, 8th Edition. Muthén & Muthén.

Nail, P. R., Harton, H. C., & Decker, B. P. (2003). Political Orientation and Modern Versus Aversive Racism: Tests of Dovidio and Gaertner's (1998) Integrated Model. *Journal of Personality and Social Psychology*, *84*(4), 754–770. https://doi.org/10.1037/0022-3514.84.4.754

Niven, K., Sprigg, C. A., & Armitage, C. J. (2013). Does emotion regulation protect employees from the negative effects of workplace aggression? *European Journal of Work and Organizational Psychology*, *22*(1), 88–106. https://doi.org/10.1080/1359432X.2011.626200

Niven, K., Totterdell, P., & Holman, D. (2013). *Emotional Labor at the unit-level.* In A. A. Grandey, J. M. Diefendorff, & D. E. Rupp (Eds.), *Organization and management. Emotional labor in the 21st century: Diverse perspectives on emotion regulation at work* (p. 121-144). Routledge/Taylor & Francis Group.

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. https://doi.org/10.1177/1046496491223005

Offermann, L. R., & Coats, M. R. (2018). Implicit theories of leadership: Stability and change over two decades. *Leadership Quarterly*, *29*(4), 513–522. https://doi.org/10.1016/j.leaqua.2017.12.003

Offermann, L. R., Kennedy, J. K., & Wirtz, P. W. (1994). Implicit leadership theories: Content, structure, and generalizability. *The Leadership Quarterly*, *5*(1), 43–58. https://doi.org/10.1016/1048-9843(94)90005-1

Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, *36*(7), 923–936. https://doi.org/10.1177/0146167210371949

Parks-Stamm, E. J., Heilman, M. E., & Hearns, K. A. (2008). Motivated to penalize: Women's strategic rejection of successful women. *Personality and Social Psychology Bulletin*, *34*(2), 237–247. https://doi.org/10.1177/0146167207310027

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), 1129–1145. https://doi.org/10.1037/a0036751

Perrone, J., & Vickers, M. H. (2004). Emotions as strategic game in a hostile workplace: An exemplar case. *Employee Responsibilities and Rights Journal*, *16*(3), 167–178. https://doi.org/10.1023/B:ERRJ.0000038651.06027.03

Phelan, J. E., & Rudman, L. A. (2010). Prejudice toward female leaders: Backlash effects and women's impression management dilemma. *Social and Personality Psychology Compass*, *4*(10), 807–820. https://doi.org/10.1111/j.1751-9004.2010.00306.x

Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science,* 1,125-126. https://doi.org/10.3758/BF03342823

Powell, G. N., & Butterfield, D. A. (1989). The "Good Manager": Did androgyny fare better in the 1980s? *Group & Organization Management*, *14*(2), 216–233. https://doi.org/10.1177/105960118901400209

Powell, G. N., & Anthony Butterfield, D. (2015). The role of androgyny in leader prototypes over four decades. *Gender in Management*, *30*(1), 69–86. https://doi.org/10.1108/GM-07-2013-0082

Powell, G. N., Butterfield, D. A., & Bartol, K. M. (2008). Leader evaluations: A new female advantage? *Gender in Management: An International Journal*, *23*(3), 156–174. https://doi.org/10.1108/17542410810866926

Powell, G. N., Anthony Butterfield, D., & Parent, J. D. (2002). Gender and managerial stereotypes: Have the times changed? *Journal of Management*, *28*(2), 177–193. https://doi.org/10.1016/S0149-2063(01)00136-2

Richards, J. M., & Gross, J. J. (2000). The Cognitive Costs of Keeping One's Cool. *Journal of Personality and Social Psychology, 79*(3), 410-424. doi: 10.1037/70O22-3514.79.3.410

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287-290. https://doi.org/10.1111/1467-9280.03437

Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*(4), 637–655. https://doi.org/10.1111/0022-4537.00233

Ritter, B. A., & Lord, R. G. (2007). The Impact of Previous Leaders on the Evaluation of New Leaders: An Alternative to Prototype Matching. *Journal of Applied Psychology*, *92*(6), 1683–1695. https://doi.org/10.1037/0021-9010.92.6.1683

Robson, V. (2007). *It's all about me: The role of the self in predicting leadership perceptions.* [Unpublished doctoral dissertation]. Virginia Tech.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Erlbaum.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26*(11), 1315–1328. https://doi.org/10.1177/0146167200263001

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. https://doi.org/10.1016/j.jesp.2011.10.008

Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. (2011). Think crisis-think female: The glass cliff and contextual variation in the think manager-think male stereotype. *Journal of Applied Psychology*, *96*(3), 470–484. https://doi.org/10.1037/a0022133

Schaubroeck, J. M., & Shao, P. (2012). The role of attribution in how followers respond to the emotional expression of male and female leaders. *Leadership Quarterly*, *23*(1), 27–42. https://doi.org/10.1016/j.leaqua.2011.11.003

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, *57*(2), 95–100. https://doi.org/10.1037/h0037128

Schein, V. E. (2001). A Global Look at Psychological Barriers to Women's Progress in Management. *Journal of Social Issues*, *57*(4), 675–688. https://doi.org/10.1111/0022-4537.00235

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. https://doi.org/10.1016/j.obhdp.2006.06.002

Sczesny, S. (2003). A Closer Look Beneath the Surface: Various Facets of the Think-Manager-Think-Male Stereotype. In *Sex Roles* (Vol. 49, Issues 7–8, pp. 353–363). https://doi.org/10.1023/A:1025112204526

Sczesny, S., Bosak, J., Neff, D., & Schyns, B. (2004). Gender stereotypes and the attribution of leadership traits: A cross-cultural comparison. *Sex Roles*, *51*(11–12), 631–645. https://doi.org/10.1007/s11199-004-0715-0

Sergent, K., & Stajkovic, A. D. (2020). Women's leadership is associated with fewer deaths during the COVID-19 crisis: Quantitative and qualitative analyses of United States governors. *The Journal of Applied Psychology*, *105*(8), 771–783. https://doi.org/10.1037/apl0000577

Shah, Y. J. (2017). *Perceptions of leaders: The role of leader prototypes and intervention to improve judgments of female leaders.* [Unpublished doctoral dissertation]. Virginia Tech.

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and Stereotyping: Cognitive Constraints on the Construction of Meaningful Social Impressions. *European Review of Social Psychology*, *11*(1), 145–175. https://doi.org/10.1080/14792772043000022

Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and stereotype maintenance processes: Attention, attribution, and individuation. *Journal of Personality and Social Psychology*, *89*(4), 607–622. https://doi.org/10.1037/0022-3514.89.4.607

Shondrick, S. J., & Lord, R. G. (2010). Implicit leadership and followership theories: Dynamic structures for leadership perceptions, memory, leader-follower processes. In G. P. Hodgkinson & J. K. Ford (Eds.). *International Review of Industrial and Organizational Psychology,* (Vol 25, pp. 1-33). John Wiley & Sons Ltd.

Shumski Thomas, J., Olien, J. L., Allen, J. A., Rogelberg, S. G., & Kello, J. E. (2018). Faking It for the Higher-Ups: Status and Surface Acting in Workplace Meetings. *Group & Organization Management*, *43*(1), 72–100. https://doi.org/10.1177/1059601116687703

Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, *2*, 219-220. https://doi.org/10.3758/BF03329252

Smith, A. N., Watkins, M. B., Burke, M. J., Christian, M. S., Smith, C. E., Hall, A., & Simms, S. (2013). Gendered Influence: A Gender Role Perspective on the Use and Effectiveness of Influence Tactics. *Journal of Management*, *39*(5), 1156–1183. https://doi.org/10.1177/0149206313478183

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

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. https://doi.org/10.1007/s10869-011-9210-0

Totterdell, P., Hershcovis, M. S., Niven, K., Reich, T. C., & Stride, C. (2012). Can employees be emotionally drained by witnessing unpleasant interactions between coworkers? A diary study of induced emotion regulation. *Work and Stress*, *26*(2), 112–129. https://doi.org/10.1080/02678373.2012.681153

Trichas, S., Schyns, B., Lord, R., & Hall, R. (2017). "Facing" leaders: Facial expression and leadership perception. *Leadership Quarterly*, *28*(2), 317–333. https://doi.org/10.1016/j.leaqua.2016.10.013

Troth, A. C., Lawrence, S. A., Jordan, P. J., & Ashkanasy, N. M. (2018). Interpersonal Emotion Regulation in the Workplace: A Conceptual and Operational Review and Future Research Agenda. *International Journal of Management Reviews*, *20*(2), 523–543. https://doi.org/10.1111/ijmr.12144

UN Women. (2019, June). *Facts and figures: Leadership and political participation.* https://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures

Van Overwalle, F., & Labiouse, C. (2004). A Recurrent Connectionist Model of Person Impression Formation. *Personality and Social Psychology Review*, *8*(1), 28–61. https://doi.org/10.1207/S15327957PSPR0801_2

Vial, A. C., Napier, J. L., & Brescoll, V. L. (2016). A bed of thorns: Female leaders and the self-reinforcing cycle of illegitimacy. *Leadership Quarterly*, *27*(3), 400–414. https://doi.org/10.1016/j.leaqua.2015.12.004

Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources. *Journal of Personality and Social Psychology*, *88*(4), 632–657. https://doi.org/10.1037/0022-3514.88.4.632

Vohs, K. D., & Schmeichel, B. J. (2003). Self-Regulation and the Extended Now: Controlling the Self Alters the Subjective Experience of Time. In *Journal of Personality and Social Psychology* (Vol. 85, Issue 2, pp. 217–230). https://doi.org/10.1037/0022-3514.85.2.217

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On The Role Of Encoding Processes In Stereotype Maintenance. *Advances in Experimental Social Psychology*, *27*(C), 177–254. https://doi.org/10.1016/S0065-2601(08)60406-2

Vonk, R. (1994). Trait inferences, impression formation, and person memory: Strategies in processing inconsistent information about persons. *European Review of Social Psychology*, *5*(1), 111–149. https://doi.org/10.1080/14792779543000039

Warning, R., & Buchanan, F. R. (2009). An exploration of unspoken bias: Women who work for women. *Gender in Management*, *24*(2), 131–145. https://doi.org/10.1108/17542410910938817

Xu, J., Liu, Y., & Guo, Y. (2014). The role of subordinate emotional masking in leader-member exchange and outcomes: A two-sample investigation. *Journal of Business Research*, *67*(2), 100–107. https://doi.org/10.1016/j.jbusres.2012.11.011

Yammarino, F. (2013). Leadership: Past, present, and future. *Journal of Leadership and Organizational Studies*, *20*(2), 149–155. https://doi.org/10.1177/1548051812471559

Appendix A: Informed Consent Form

Informed Consent: Emotion Regulation and Evaluations of Leaders
IRB # 20-336

Principal Investigator (PI):
Neil Hauenstein, Ph.D.
nhauen@vt.edu

Co-Investigator
Elsheba Abraham, M.S.
elsheba@vt.edu

Thank you for your interest in this study! Before beginning, please take your time to read the consent form below. If you understand the statements and freely consent to participate in the study, please click on the "I Agree" button at the bottom of the page.

Study Purpose
The overall purpose of this study is to understand how self-regulation of emotions can affect evaluations of a leader.

Procedure
This session will be conducted through Zoom by the experimenter. You will first be briefed about the study procedure, which will include watching three different videos and providing answers to several questionnaires regarding your evaluations of a leader. Once all your questions have been answered and you provide consent, then the session will begin.

First, you will watch a 3-minute video on Jackson Pollock followed by a 9-minute video compilation of scenes from “The Office”. While watching both videos, you will either be instructed to suppress your emotional expressions (i.e. keep a poker face) or to watch them normally. These instructions are assigned randomly, so you have a 50/50 chance of being in either condition. *During this segment, you will be video-recorded.*

Then, you will watch a 16-minute compilation of four videos of working adults collaborating on a business task. Before the business video begins, you will be shown a screenshot identifying the leader and 3 adults in the video. After each of the four videos, you will be asked to rate the leader on several items.
Once you watch all videos, you will then complete several other questionnaires regarding your evaluation of the leader. Finally, you are provided with a debriefing statement that elaborates on the research question and summarizes the study goals.

Participation in the study will take no longer than 1-hour and your results will be kept confidential.

Risks
There are no more than minimal risks involved in participation in this study. You may experience some discomfort because of the sensitive nature of some of the humor in "The Office" video clips.

If additional counseling or medical treatment is needed as a result of the study, you can seek help from the Cook Counseling Center of Schiffert Health Center on campus.

Benefits
There is potential for you to gain insight into how you evaluate the effectiveness of a leader. You may also find "The Office" video clips entertaining to watch.

No promise or guarantee of benefits has been made to encourage you to participate.

Compensation
You will be compensated for participating in this phase of the study by receiving a $10 gift card.

Confidentiality
Your responses will be kept strictly confidential. Your information will only be accessible to researchers listed on this protocol. All response data is de-identified and only contains a numerical ID. Data will be stored on computer and the Virginia Tech Google Drive requiring passwords, and only the researchers on this protocol will have access to it.

The Virginia Tech's Human Research Protection Program may view the study's data for auditing purposes. They are is responsible for the oversight of the protection of human subjects involved in research.

The results of this study may be presented in summary form at conferences, in presentations, academic papers, and as part of a thesis/dissertation.

Freedom to Withdraw
You are free to withdraw your consent and terminate your participation at any time. Withdrawing from the study will not cause you any penalty or trouble.

Questions Concerning Research
Should you have any questions about this research or its conduction, you may contact:

Researchers:

Neil Hauenstein, Ph.D.
540-231-5716
nhauen@vt.edu

Elsheba Abraham, M.S.
608-772-9795
elsheba@vt.edu

Dr. Roseanne Foti, Chair Psychology Department
rfoti@vt.edu

Should you have any questions or concerns about the study or the rights you hold as a research participant, or need to report an injury or event that occurred in relation to the research, you may contact the Virginia Tech Human Research Protection Program at 540-231-3732 or at irb@vt.edu.

Subject's Consent

If you have any questions prior to participation, please contact Elsheba Abraham at elsheba@vt.edu before you continue.

If you are 18 years of age or older, understand the statements above, and freely consent to participate in the study, click on the "I Agree" button to begin the experiment.

- o I Agree
- o I Do Not Agree

Appendix B: Agentic and Communal Traits in the Leader Behavior Videos

Communal and agentic leader attributes depicted in the videos.

Communal Leader Attributes	**Agentic Attributes**
Appreciative	Decisive
Considerate	Independent
Appeasing	Confident
Cooperative	Pushy
Praising	Competitive

Number of agentic (blue font) and communal (red font), leadership behaviors by leader and coworkers.

Scene	6	7	8	9	Total
Leader	6, 5	5, 5	5, 6	5, 6	21, 22
Coworker 1	4, 3	5, 3	2, 1	0, 1	11, 8
Coworker 2	3, 2	2, 0	0, 0	1, 1	6, 3
Coworker 3	2, 1	1, 1	1, 0	0, 2	4, 4

Appendix C: Emotional Expressions Codebook

Emotional Expressions Behavioral Coding Scheme

(adapted from Facial Action Coding System (FACS); Ekman, Friesen & Hager, 2002)

Coding Units of Behaviors

Category	Action Unit	Description
Body Movement	Moving the upper torso of the body	Shifting in the seat (e.g. moving backward away from the screen)
		Moving shoulders
	Hands are raised towards the face	Actual touching of face
	Places head in hands	
	Taking a deep breath	
	Head movement	Head tilts
Facial Movement	Laughing	
	Moving eyebrows	Eyebrows raised (like in surprise)
	Moving eyelids	Squinting
		Winking
		Closing the eyes
		Noticeable blinking (i.e. 2 blinks in succession or more
	Moving of nose	Wrinkling of nose
		Dilating/Compressing nostril
	Moving of cheeks	Cheek puff
		Cheek suck

Category	Action Unit	Description
Mouth Movement	Smiling	Corners of lips are pulled upwards, cheeks raised
	Opening of mouth	Lips parted
		Speaking to themselves (code a string of utterances just once)
		Blowing air through their mouth
	Jaw movement	Jaw drop (like in gasping)
		Clenching of jaw
		Moving the chin
	Playing with lips	Lip biting
		Lip licking
	Gulping/Swallowing	Any movement of the throat

Appendix D: Emotional Expressions Rating Sheet

Name:
Video: e.g. ID03 from 4:08-11:05

Emotional Expressive Coding Sheet
Adapted from Gross & Levenson (1993; 1997)

Please mark down every time you see any of the coding units of behavior described in the coding scheme:

Body Movement	
Facial Movement	
Mouth Movement	

Please rate your overall impression of the duration, frequency **and** intensity of the emotional/behavioral expressions below:

	No Reaction at All	Slightly Reactive	Moderately Reactive	Very Reactive	Extremely Reactive
Body movement	1	2	3	4	5
Facial movement	1	2	3	4	5
Mouth movement	1	2	3	4	5
Overall emotional expressivity	1	2	3	4	5

Note:

- think of "No reaction at all" as low duration and frequency, low intensity
- think of "Slightly reactive" as low duration and frequency, medium intensity
- think of "Moderately reactive" as medium duration and frequency, medium intensity
- think of "Very reactive" as medium duration and frequency, high intensity
- think of "Extremely reactive" as high duration and frequency, high intensity

Most reactive participant: IDXX; Least reactive participant: IDXX

Other behaviors not observed:

Appendix E: Leadership Perceptions scale[3]

Based on the business scenarios you watched, please respond to the following question regarding the leader's behaviors. Please respond to the statements below using a 5 point scale, where 1 = Not at All and 5 = Extreme Amount

1. To what extent does {Sue/Bob} fit your image of a leader?
2. How typical was {Sue/Bob} of a leader?
3. How much leadership did {Sue/Bob} exhibit?
4. To what extent does {Sue/Bob} engage in leadership behaviors?
5. If you had to choose a leader for a new task, how willing would you be to choose {Sue/Bob} as the leader?

[3] Global Leadership Impression scale (GLI; Cronshaw & Lord, 1987; Lord et al., 1984)

Appendix F: Perceived Leader Effectiveness scale[4]

Based on the business scenarios you watched, please rate your level of agreement using the 5 point scale, where 1 = Strongly Disagree and 5 = Strongly Agree

1. {Sue/Bob} is very effective as a leader.
2. {Sue/Bob} is a good leader.
3. {Sue/Bob} has good leadership qualities.
4. {Sue/Bob} behaves as a leader should.

[4] Adapted from Perceived Leadership Effectiveness scale (Hais et al., 1997)

Appendix G: Leader Behavior Recognition Measure

Please use the following scale of Yes/No to indicate if you observed the following behaviors in the video scenarios you saw.

1. The leader was quick to assign responsibilities for the new employee orientation (A, Yes)
2. The leader decided how to split the money between projects (A, Yes)
3. The leader assigned Dave the task of scheduling meetings to perform a job analysis (A, Yes)
4. The leader told Dave it was important he set up the two meetings right away (A, Yes)
5. The leader decided to hire Dr. Smith and asked a team member to contact him for the contracts (A, Yes)
6. The leader directed Libby to help Dave with the fax machine (A, Yes)
7. The leader said it was a good suggestion to bring in an outsider to review the recurring budget issue (C, Yes)
8. The leader demonstrated concern for a stressed-out employee and asked a co-worker to check in on the employee (C, Yes)
9. The leader praised Dave for his excellent job forecasting expenses (C, Yes)
10. The leader decided to host a dinner to build group spirit while at the conference (C, Yes)
11. The leader inquires what other group members' think about getting a fax machine that is in the middle of their identified price range (C, Yes)
12. The leader thanks Libby for her help with finding more information about other office suppliers (C, Yes)
13. The leader tells Libby to call Human Resources and inquire about the missing meeting notes (A, No)

14. The leader asks Dave to double-check Dr. Smith's credentials (A, No)
15. The leader says they called Office Max to order new desks (A, No)
16. The leader asks Dave to form a committee to brainstorm strategies to reduce department expenses (A, No)
17. The leader tells Libby to rebalance the budget (A, No)
18. The leader directs Libby to get from upper management a list of employees who will get a bonus this year (A, No)
19. The leader expresses concern for the staff being overworked (C, No)
20. The leader thanked Libby for compiling the necessary paperwork for the conference (C, No)
21. The leader demonstrated concern for a sensitive employee struggling with poor performance (C, No)
22. The leader asks Dave to send out invites for the new employee orientation event (C, No)
23. The leader inquires when other group members are taking vacations (C, No)
24. The leader expresses appreciation to the group members for finding additional funding for the Hanson Memorial and Higgins projects (C, No)

Note: A indicates agentic behaviors and C indicates communal behaviors. "Yes" indicates behaviors that did occur in the videos and "No" indicates behaviors that did not occur in the videos. All "Yes" behaviors occur in the videos only once. For ease of reference, items are presented by behavior type and occurrence. Items will be randomized when presented to participants with no mention of behavior type and occurrence

Appendix H: Leader Competence and Warmth Ratings[5]

Please evaluate the leader on the following dimensions using the scale provided. Use a 5 point scale, where 1 = Strongly Disagree and 5 = Strongly Agree.

1. {Sue/Bob} is very capable of performing her/his job.
2. {Sue/Bob} is known to be successful at the things she/he tries to do.
3. {Sue/Bob} has much knowledge about the work that needs done.
4. I feel very confident about {Sue/Bob}'s skills.
5. {Sue/Bob} has specialized capabilities that can increase team performance.
6. {Sue/Bob} is well qualified.
7. {Sue/Bob} would be very concerned about my welfare.
8. My needs and desires are very important to {Sue/Bob}.
9. {Sue/Bob} would not knowingly do anything to hurt me.
10. {Sue/Bob} really looks out for what is important to me.
11. {Sue/Bob} will go out of their way to help me.

Note: Items 1-6 represent competent attributes and items 7-11 represent warmth attributes

[5] Mayer and Davis (1999) Ability and Benevolence subscales

Appendix I: Modulation Strategy Questionnaire[6]

Please rate your agreement to the items below from 1 (Strongly Disagree) to 5 (Strongly Agree):

Perceived Success:

1. I succeeded in not showing my emotions
2. I continuously tried to hide my emotions
3. I had trouble not to show my emotions

Emotional Control:

1. I tried to control my emotions
2. I controlled myself

Note: Participants in the suppression condition will answer all items, while participants in the no suppression condition will only answer Emotional Control items.

[6] Adapted from Martijin et al.'s (2002) Modulation Strategy Questionnaire

Appendix J: Demographic Questionnaire

1. Please enter your age:
2. Please indicate your gender:
    a. Female
    b. Male
    c. If not listed above, please specify
3. Please indicate your race/ethnicity below:
    a. Asian
    b. Black or African-American
    c. White
    d. Hispanic/Latinx
    e. Native-American
    f. Multiracial
    g. If not listed above, please specify
4. Do you have a job?
    a. Yes
    b. No
5. If you do have a job, how many hours a week do you work?
6. Have you currently or in the past had a female boss or supervisor at work?
7. How did you spent the hour right before this experiment session?
    a. On Zoom or another video-hosting platform attending a live, synchronous session (e.g. class, meeting) - video camera ON
    b. On Zoom or another video-hosting platform attending a live, synchronous session (e.g. class, meeting) - video camera OFF
    c. On the computer doing anything else that did not require your live participation or engagement
    d. Not on the computer
    e. Other