

**WHEN THEY WERE GOOD THEY WERE VERY GOOD INDEED BUT WHEN
THEY WERE BAD THEY WERE HORRID! BIASING EFFECTS ON RATINGS OF
LEADERSHIP**

John Antonakis¹
Department of Management
University of Lausanne

Marika Angerfelt
Department of Management
University of Lausanne

Nagaraj Sivasubramaniam
A. J. Palumbo School of Business Administration
Duquesne University

PAPER PRESENTED AT

*WOMEN AS GLOBAL LEADERS CONFERENCE
MARCH 14-16, 2005 DUBAI, UAE.*

¹Corresponding author:
Faculty of Economics and Business Administration
University of Lausanne, BFSH-1
Lausanne, CH-1015
Switzerland
Tel: ++41 (0)21 692-3300
e-mail: john.antonakis@unil.ch

Authors' note:

We would like to thank Alice Eagly for her helpful advice and for introducing us to the amplification-ambivalence bias.

ABSTRACT

We examined rater heuristic judgment processes, manipulating leader sex, job sex-typing, and organizational performance outcomes. Participants rated a target individual's leadership style, personality, and whether organizational outcomes were attributable to the individual. As expected, the female leader was seen less favorably than the male was when organizational outcomes were bad and the context male sex-typed—unexpectedly, she was seen more favorably under good performance outcomes (the male leader was rated in the same way in a feminine context). Independent of other conditions, the female leader was rated as being more neurotic than was the male. Furthermore, leader personality ratings were predictive of leadership style. Finally, participants overblamed out-of-context leaders independent of performance outcomes. Our results provide new understandings regarding situational cue effects on leader ratings, concordant with the “shifting standards” theory.

Many writers, from the ancient Greeks (Plato, circa 360BCE/1991) to contemporary western scholars (Bass, 1980; Burns, 1978; Gardner, 1990; Kellerman, 2004; Vroom, 1976) agree that effective leadership is one of society's most valued functions. As succinctly stated by Bennis (2004, p. 331) "the quality of all our lives is dependent on the quality of our leadership." It would be desirable—economically as well as ethically—that leaders are evaluated such that those who reach top positions of political or economic power do so as a result of their ability and competence and, all else being equal, not because of their sex. But is this the case? It is well established that men dominate most of society's consequential activities, particularly high-status leadership positions (Eagly & Karau, 2002). Various reasons might exist for this phenomenon, ranging from the argument that biological factors account for the fact that males are superior—and thus deserving of the power they have—to the reasoning that differences in power and influence reflect socially constructed expectations and constraints that makes it easier for men to rise to top positions of leadership (Eagly & Carli, 2004).

The biological-differences argument appears to be implausible for a variety of reasons, which are beyond the scope of this paper (see Eagly & Carli, 2004; Wood & Eagly, 2002). An alternative position suggests that social role expectations spur or inhibit the advancement of individuals to positions of power (Eagly & Karau, 2002). Given that women leaders (a) exhibit prototypically good leader behavior to a slightly higher degree than do male leaders (see Antonakis, Avolio, & Sivasubramaniam, 2003; Eagly & Johannesen-Schmidt, 2001; Eagly, Johannesen-Schmidt, & van Engen, 2003) and (b) are as effective as men leaders are (Eagly, Karau, & Makhijani, 1995), one would logically expect that women occupy about the same proportion of positions of power that men do. They do not. A potential explanation regarding why women are prevented from getting to the top might, therefore, have something to do with how women leaders are perceived and evaluated, irrespective of what they actually

do or how competent they might be. Our goal was to investigate this potential source of rater bias.

ROLE CONGRUITY THEORY

According to Eagly and Carli (2004) the same behavior, competence, or outcome might not be evaluated the same way when exhibited by a male as compared to a female leader. If this reasoning is correct, cognitive distortion mechanisms must be operating that affect what perceivers pay attention to, how they evaluate it, and how they recall it. Role congruity theory of prejudice (Eagly & Karau, 2002) suggests that women are disadvantaged because perceivers imbue them with characteristics that are deemed incongruent for success in positions of leadership (Eagly & Carli, 2004). The ascription of characteristics stems from how social roles are distributed in society—individuals are supposed to behave in ways that are consistent with culturally-determined gender roles (Eagly et al., 1995; Heilman, 1983, 2001). Eagly and colleagues, and Heilman argue that women are associated with communal roles and are described as generally being compliant, unselfish, nurturing, sociable, agreeable, nice, and so forth. Contrarily, men are associated with agentic roles and are described as generally being aggressive, assertive, dominant, competent, and so forth.

Apart from the descriptive aspect of these social roles assigned to men and women, role congruity theory (Eagly & Karau, 1991) suggests that these descriptive expectations become injunctive norms (i.e., prescriptive rules). These injunctions dictate how an individual is supposed to behave, which makes it difficult for women to assume positions of power, particularly in male dominated domains (see also Heilman, 1983). Briefly, the characteristics of women are seen as incongruent with the characteristics expected of a leader, which have traditionally been, and still are today, defined in masculine terms (see Powell, Butterfield, & Parent, 2002). Thus, women are less likely to emerge as leaders because they are seen as less able (the *descriptive* expectation).

If, however, a woman demonstrates characteristics that are masculine—which are seen as more concordant with leadership roles—she violates injunctive roles associated with her sex because she is demonstrating undesirable behaviors (the *prescriptive* expectation). Thus, she is evaluated less favorably. As stated by Eagly and Carli (2004, p. 297), gender roles create a “double bind for female leaders who must demonstrate exceptional competence to be seen as equal in ability to men and must also avoid threatening others with the competence and lack of warmth.” If this theorizing is correct, stereotyping of this nature makes it difficult for women to emerge as leaders or to be evaluated on equal terms with men. As a consequence, disproportionately fewer women will be found in top leadership positions, potentially reinforcing the vicious circle that males should occupy these positions.

Our overarching purpose was to examine the veracity of these claims. We examined rater judgment processes of leaders to determine how raters are affected by leader sex as well as other situational cues. An understanding of causal mechanisms associated with evaluation biases has important implications for the design of leader training and evaluation systems. Inasmuch as research in this domain has been bountiful, certain questions about the impact of stereotypes remain unexplored, particularly with regard to the transformational-transactional leadership paradigm (Powell, Butterfield, Alves, & Bartol, 2004), which currently dominates the leadership research landscape (Antonakis, Cianciolo, & Sternberg, 2004; Hunt, 1999; Lowe & Gardner, 2000). We focused on how rater judgments about male and female leaders were jointly influenced by the context in which leadership is observed (male or female sex-typed) and the extent to which performance outcomes were good or bad. The questions we sought to answer were fourfold:

1. Are male and female leader behaviors judged equivalently?
2. Is organizational success or failure attributed the same way to male and female leaders?

3. Do raters make the same inferences regarding male and female leader's personality?
4. Are personality factors associated with successful leader behavior?

PERSON PERCEPTION IN LEADERSHIP PROCESSES

Leadership can be defined as “the nature of the influencing process—and its resultant outcomes—that occurs between a leader and followers and how this influencing processes is explained by the leader's dispositional characteristics and behaviors, follower perceptions and attributions of the leader, and the context in which the influencing process occurs (Antonakis et al., 2004, p. 5).

Two features in the above definition are essential for understanding the leadership process: follower judgments and the context. Judgments include evaluating what others are like (inferences) or the causes of outcomes (attributions) (Erickson & Krull, 1999). Many factors influence how social objects—particularly leaders—are perceived and how outcomes are assigned to causes (i.e., was the leader responsible for outcomes?). Contexts, as well as other situational signaling mechanisms (e.g., leader sex, organizational performance outcomes) act as triggers for category or schema-driven information processing (Brown, Scott, & Lewis, 2004).

A schema, which broadly includes prototypes (i.e., normative expectations), stereotypes, and scripts, is “a cognitive structure that represents knowledge about a concept or type of stimulus, including its attributes and the relations among those attributes” (Fiske & Taylor, 1991, p. 98). Leader prototypes are associated with their respective contexts (Lord, Brown, Harvey, & Hall, 2001; Lord, Foti, & De Vader, 1984) and influence various cognitive information processing functions (Brown et al., 2004). For example, a military leader is not expected to behave like a leader of a cross-functional scientific team. Because of prototypical

expectations, perceivers are able to determine which individuals should emerge as leaders and the criteria under which to judge their performance.

Leader prototypes are also associated with situational cues. For example, if raters have knowledge of purported leader outcomes (bad organizational performance) they use that outcome as an indicator of leadership (bad leadership). Bad performance is—in the absence of concrete information indicating otherwise (i.e., in information uncertain conditions, see Cantor & Mischel, 1979)—representative of bad leadership (Binning, Zaba, & Whattam, 1986). A heuristic judgment (i.e., mental shortcut) is made, classifying the leader as bad, regardless of whether the leader was actually bad or whether the leader was responsible for the bad outcomes. The “performance cue paradigm” has received ample empirical support (e.g., Bryman, 1987; Gioia & Sims, 1985; Larson, 1982; Lord, Binning, Rush, & Thomas, 1978; Rush, Thomas, & Lord, 1977; see Brown et al., 2004 for a review).

Leader sex

An important individual-difference cue that might trigger expectations and bias judgments in stereotypical ways is leader sex. Stereotypes, which are a type of schema, refer to expectations about a target individual wherein the perceiver imputes the target with characteristics ascribed to the group in which the individual belongs (Fiske & Taylor, 1991). Biases in judgments occur because perceivers process information in a way that is consistent with the expectation (e.g., women are bad leaders) by ignoring schema inconsistent information and focusing on and recalling only schema consistent information, even with respect to unobserved but schema consistent attributes (Cantor & Mischel, 1977; Fiske, 1995; Fiske & Taylor, 1991; also Feldman, 1981 for applications to performance appraisal).

Although judgment heuristics are useful in some instances, particularly as cognitive resource saving devices (Fiske & Taylor, 1991) they can create biased outcomes, particularly in contexts that trigger sex-role expectations (Heilman, 1983). Furthermore, once stereotypes

are activated they are difficult to correct even when raters are aware that stereotypes are invalid or when they are motivated to be accountable for their judgments (Nelson, Acker, & Manis, 1996). Apart from the impact of stereotypes on information processing, stereotypes can also create self-fulfilling prophecies because of cuing mechanisms (e.g., follower stereotyping leads to negative body language, which gives signals to others and the leader, that the leader is ineffective, Butler & Geis, 1990). Targets who are being stereotyped feel threatened, thus creating self-limiting behavior that perturbs their performance (known as stereotype threat, see: Hyde & Kling, 2001; Steele, 1997).

The way psychologists have determined whether raters exhibit stereotyped information processing is to expose participants to a target leader behaving in a particular way and to manipulate, among others, the leader's sex. If participants react differently to the same behavior as a function of the leader's sex, then one can determine that participants are not processing the same information about the target individual in the same way. A meta-analytic review demonstrated, as social role theory or the "Lack-of-Fit" model (Heilman, 1983) would predict that raters do indeed evaluate women and men leaders differentially, even if they have exhibited precisely the same behavior (Eagly, Makhijani, & Klonsky, 1992). Specifically, women leaders are generally evaluated more negatively than male leaders are, particularly in stereotypically masculine contexts (discussed further below) and when exhibiting autocratic or directive leadership styles. Interestingly, men were not stereotyped if leading in a feminine way which suggested that they have more latitude in the range of behaviors that they portray.

Sex typing of the job

Similar to the tenets of social role theory as well as attribution theory (see Kelley & Michela, 1980), Heilman (1983) argued that an individual in an atypical context is not expected to do well. The typicality of the sex-context link depends on the characteristics that are normally required for a particular position, which will also be reflected in the sex of the

average job holder (see also Cejka & Eagly, 1999). Thus, expectations of performance depend on “the fit between the perception of an individual's attributes and the perception of the job's requirements in terms of skills and abilities” (Heilman, 1983, p. 278). This Lack-of-Fit model predicts that when fit is poor, failure is expected; when fit is good, success is expected.

Applied to leadership—and in particular contexts that are masculine oriented—perceivers will expect masculine characteristics as determinants of leader success. Because women's stereotypical characteristics do not fit with achievement oriented characteristics (independence, competitiveness, self-confidence, ambition) necessary for leadership expectations of failure will ensue leading to biased evaluations of performance. These biases include negative evaluations, devaluation of performance, denying credit for organizational success, and discrimination in selection decisions, performance appraisal and reward allocations (Heilman, 1983, 2001). Important to note here is that both men and women can be stereotyped if found to be in an uncongenial context.

Butterfield and Ginnell (1999, p. 336) noted that “context is extremely important in mitigating sex effects,” and urged researchers to include contextual factors in research designs. Controlling for context does confirm the predictions of the Lack-of-Fit model. In general, for equally qualified candidates or for equal performance, women tend to receive lower ratings in masculine sex-typed jobs and men receive lower ratings in feminine sex-typed jobs. This phenomenon (i.e., the context-sex interaction) has been observed in leader emergence (see meta-analysis of Eagly & Karau, 1991), leader effectiveness (see meta-analysis of Eagly et al., 1995), evaluation of leader behavior (see meta-analysis of Eagly et al., 1992), simulated employment contexts (see meta-analysis of Davison & Burke, 2000), and hiring decisions (Cohen & Bunker, 1975), among others. Although some of the meta-analyses are dated, recent studies continue to provide evidence for this prediction (e.g., see Heilman, Wallen, Fuchs, & Tamkins, 2004; Ritter & Yoder, 2004).

Thus, we expect context congeniality to interact with the leader's sex to determine ratings of leadership and leadership effectiveness. However, this effect will be further moderated by organizational performance signals that raters associate with the leader. We therefore tested the following general hypothesis:

*Hypothesis 1: Ratings of leadership will be jointly affected by a three-way leader sex*sex-typing of job*organizational performance cue interaction.*

For ease of reference, we have depicted the three-way interaction and the series of hypothesis derived from it in Figure 1.

[Insert Figure 1 about here]

We expected the performance cues to have differential effects on men and women leaders as a function of context. Although performance cues should be associated with good leader behavior, perceivers might discount the cues if the cue is not representative of the target leader. Thus, the female leader should be rated more severely than will be the male leader in a masculine typed context whereas the man would be treated in a more tempered manner. The reason for this occurrence is because performance expectations for a man will be higher than those of a woman—he should thus be evaluated more leniently, irrespective of a good or bad performance cue. Conversely, performance cues will have a significant effect on the woman in a masculine context, because, as attribution theory would predict, a bad performance signal will confirm the perceivers initial low expectations. Thus, she will be rated much lower than the man will be. The woman leader will also be rated lower than a man leader under good performance signals simply because this performance will be devalued relative to the man's performance. Thus, ratings under good and bad performance signals will be accentuated for a woman leader. This amplification or “booster” effect should not be evident in ratings of male leader because a bad performance signal will be seen as a “blip”—a chance event that should be discounted. We therefore expect that the difference in ratings

under good or bad performance signals for male leaders will be attenuated in masculine contexts.

The form of the interaction effect for the male sex-typed context was tested with the following hypotheses:

Hypothesis 1a: In the presence of good or bad performance cues the female leader will be rated lower than will the male leader on prototypically effective leadership.

Hypothesis 1b: The performance cue will have a larger effect for the female leader than for the male leader on ratings of prototypically effective leadership.

The Lack-of-Fit model would make precisely the same predictions for a male leader in a female context. Thus, for ratings of leaders in a feminine context we tested the following hypotheses:

Hypothesis 1c: In the presence of good or bad performance cues the male leader will be rated lower than will the female leader on prototypically effective leadership.

Hypothesis 1d: The performance cue will have a larger effect for the male leader than for the female leader on ratings of prototypically effective leadership.

Following the above theorizing, and given that prototypically effective leadership is by definition opposite to prototypically ineffective leadership (i.e., they are negatively correlated), we expected out-of-context leaders to be rated as exhibiting more prototypically ineffective behaviors. Furthermore, and following the above theorizing, we expected the performance cue to have greater effects on the out-of-context leader. Thus, for the male sex-typed context we tested the following hypotheses:

Hypothesis 1e: In the presence of good or bad performance cues the female leader will be rated higher than will the male leader on prototypically ineffective leadership.

Hypothesis 1f: The performance cue will have a larger effect for the female leader than for the male leader on ratings of prototypically ineffective leadership.

Again, concordant with the above reasoning, the Lack-of-Fit model makes precisely the same predictions for a male leader in a female context. We thus tested the following hypotheses:

Hypothesis 1g: In the presence of good or bad performance cues the male leader will be rated higher than will the female leader on prototypically ineffective leadership.

Hypothesis 1h: The performance cue will have a larger effect for the male leader than for the female leader on ratings of prototypically ineffective leadership.

Ratings of leader effectiveness are consequences of prototypically good leader behavior. Thus, we expect the patterns of findings to be aligned in the same manner as the findings regarding leader behavior. Using the precise reasoning above, we therefore tested the following general hypothesis.

*Hypothesis 2: Ratings of leader effectiveness outcomes will be jointly affected by a three-way leader sex*sex-typing of job*organizational performance cue interaction.*

The form of the interaction effect for the male sex-typed context was tested with the following hypotheses:

Hypothesis 2a: In the presence of good or bad performance cues the female leader will be rated lower than will the male leader on leader effectiveness outcomes.

Hypothesis 2b: The performance cue will have a larger effect for the female leader than for the male leader on ratings of leader effectiveness outcomes.

Again, the Lack-of-Fit model makes the same prediction for a male leader in a female context. Thus, for ratings of leaders in a feminine context we tested the following hypotheses:

Hypothesis 2c: In the presence of good or bad performance cues the male leader will be rated lower than will the female leader on leader effectiveness outcomes.

Hypothesis 2d: The performance cue will have a larger effect for the male leader than for the female leader on ratings of leader effectiveness outcomes.

Attributions of Performance

According to Heilman (1983), women and men are not attributed success in an equivalent manner—for equivalent performance men are generally viewed as more skillful. Following the precepts of attribution theory, The Lack-of-Fit model predicts that expected performance will be attributed to individual disposition; however, unexpected performance will be attributed to external or temporary factors (e.g., chance, easiness of the task, etc.). Thus, Heilman argues that the successful performance of a woman in a male sex-typed job will not be credited to her dispositions but to situational factors—in this way, the perceiver maintains cognitive consistency and the initial expectation is reified.

Put another way, Foschi (2000) stated that “those who are considered to be of lower status will have their successful performances scrutinized (since these are inconsistent with status) and then assessed by a stricter standard than similar performance by higher status actors” (p. 25). Because good performance of a low status individual is unexpected (i.e., not in the norm), more evidence of competence will be required before a perceiver equates that performance with that of a higher status individual. In this way, low status individuals will be held to higher standards of evidence—a one-off demonstration of competency simply cannot be representative of the individual’s disposition.

There is a long line of research showing that the performance of out-of-context individuals and in particular women is devalued (see meta-analysis by Swim & Sanna, 1996). In line with attribution theory, these findings generally demonstrate that above average performance on a sex-congruent task would likely be attributable to internal factors, whereas the same performance on a sex-incongruent task would be likely attributed to external factors (e.g., chance, see Deaux & Emswiller, 1974).

To summarize, raters do not expect good performance from an out-of-context individual and thus are more likely to attribute good performance to external factors and bad

performance to internal factors. Therefore, the bad performance of an in-context individual will be attributed to external factors. The following general hypothesis was tested:

*Hypothesis 3: Attribution of organizational performance will be jointly affected by a three-way leader sex*sex-typing of job*organizational performance cues interaction.*

This form of predicted interaction is the standard cross-over interaction. Based on the above hypothesis, we tested the following hypotheses for a male sex-typed context:

Hypothesis 3a: In the presence of good performance cues, organizational performance will be attributed to external factors in the case of a female leader whereas organizational performance will be attributed to internal factors in the case of a male leader.

Hypothesis 3b: In the presence of bad performance cues, organizational performance will be attributed to external factors in the case of a male leader whereas organizational performance will be attributed to internal factors in the case of a female leader.

The male leader should be similarly stereotyped in a feminine context, as the Lack-of-Fit model would predict. We therefore tested the following hypotheses:

Hypothesis 3c: In the presence of good performance cues, organizational performance will be attributed to external factors in the case of a male leader whereas organizational performance will be attributed to internal factors in the case of a female leader.

Hypothesis 3d: In the presence of bad performance cues, organizational performance will be attributed to external factors in the case of a female leader whereas organizational performance will be attributed to internal factors in the case of a male leader.

Ratings of personality

Ratings of leader personality, as a function of leader sex, sex-typing of the job, and organizational performance signals is, to our knowledge, an understudied topic. Our predictions here thus follow from general theorizing regarding the impact of stereotyping, as well as the empirical findings of Heilman (1984), Heilman, Block, and Martel (1995), and

Heilman, Block, Martell, & Simon (1989). These results indicated that personality ratings of men and women job applicants or of men and women leaders do not differ significantly or have less of an impact in the presence of information rich performance cues (i.e., indicators of successful performance). However, even in those conditions, women still tend to be characterized as being less rational and more hostile (i.e., more neurotic or emotionally unstable in general) toward others (Heilman et al., 1995). The conditions that we have created could be characterized as information rich in terms of the performance signals; however, we did not include much individuating information (see *Materials and Manipulations* section below). Given that dispositional inferences are linked to descriptive stereotyping (what women or men are like) and that individuating information (or performance signals) should reduce the impact of descriptive stereotyping (Gill, 2004), we expect that our results should parallel those of Heilman and colleagues. Thus, the predicted three-way interactions discussed above should not hold in this case. However, performance signals, should theoretically, be linked to dispositional inferences, independent of sex and job sex-typing. That is, under good (bad) performance signals, the leader will be seen to have prototypically good (bad) personality traits. In the absence of relevant theory, we did not make predictions regarding two-way interactions.

Furthermore, given recent meta-analytic results we expected ratings of personality, based on the big five framework (i.e., neuroticism or emotional instability, extraversion, openness, agreeableness, conscientiousness) to be predictive of ratings of leadership and effectiveness across all conditions (see Bono & Judge, 2004; Judge, Bono, Ilies, & Gerhardt, 2002). Ratings of participants regarding leader personality should be related in a consistent manner to leader behavior as a function of performance outcomes but irrespective of the sex of the leader, and sex typing of the job. That is, participants should be affected by the manipulated variables in rating the target leader's behavior in a particular condition; however,

they will be equally affected in how they rate the target's personality. These two ratings should be intuitively aligned to parallel empirical results regarding the leadership-personality link. For example, a leader associated with good (bad) outcomes will be seen as demonstrating prototypically good (bad) leader behaviors, which would be predicted by high (low) extraversion, high (low) openness, and low (high) neuroticism. Because we used business contexts, we did not expect agreeableness and conscientiousness to be predictive of leadership (Judge et al., 2002).

Hypothesis 4a: Independent of other treatments, the female leader will be viewed as more emotionally unstable (i.e., neurotic) than will be the male leader.

Hypothesis 4b: Leaders rated in the presence of good (bad) performance signal will be described as more (less) extraverted, more (less) open, and more (less) emotionally stable than will leaders described in the presence of a bad performance signals, independent of leader sex and job sex type.

Hypothesis 4c: Across manipulated conditions, ratings of personality will be aligned with ratings of leadership such that extraversion, openness, and emotional stability will be positively related to prototypically good leadership and negatively related to prototypically bad leadership.

Sex of rater

A recent meta-analytic review established that rater sex does not systemically influence whether raters will provide favorable ratings of targets with the same sex as the rater (Davison & Burke, 2000). Another meta-analysis showed that, although male raters tend to devalue female leaders, rater sex is not a significant covariate in the case of college students (Eagly et al., 1992)—the population used in this study. Finally, as noted by Heilman et al. (2004), gender stereotypic norms seem to be universally held and do not depend on the sex of the rater. Thus, we did not expect rater sex to significantly interact with the treatments.

METHOD

Overview

The overarching goal of our experiment was to determine whether the same target leader would be viewed differently on various leadership measures (i.e., leadership style, leader outcomes, responsibility attribution, and personality). We used a 2x2x2 between-subjects (factorial) design manipulating: (a) the leader's sex, (b) organizational performance signals, (b) the sex-typing of the job.

We chose to use the popular vignette approach, which affords a high degree of experimental control (Eagly et al., 1992). Using this approach entails providing a written description of a target leader while crossing the three independent variables. Thus, participants were randomly assigned to one of eight conditions. If raters do pay attention to the manipulated variables and if they are affected by them as we have theorized then this experimental approach would allow us to estimate the extent to which the manipulated factors were causally related to the dependent variables.

Participants

Three hundred and forty nine participants were recruited from a medium sized university in the northeastern US. These students were undergraduate business majors, the majority of whom (70%) were freshmen and sophomores. Their mean age was 20.28 years (sd = 1.39) and their average work experience was 4.64 months (sd = 11.48). The sex of the participants was roughly equal, $\chi^2(1) = 2.09$, $p > .10$, that is, 57.5 % male and 42.5 % female. Approval to conduct the experiment was sought and given by the university's ethics board.

Materials and manipulations

The materials were developed to depict a realistic leadership situation in which a director, Robin, was discussing a typical operations problem with his/her management team.

The vignette was one page long, single-spaced. Eight conditions were developed in which the following three independent variables were fully crossed:

1. *subject sex* (either a male or a female leader)
2. *job sex type* or context congeniality (the leader was either the director of a “UBC Bank” or “Fashion Today,” a department store exclusively for women customers)
3. *organizational performance outcomes/cues* (either rising or falling profits), depicted once prominently in graphical form (i.e., net income trends from 1997 to 2004) as well as descriptively in four different ways (e.g., “Net profits have been going down at a steady rate between 2002 and 2003”). A projection was also given for the current fiscal year (e.g., “Profits [in 2004] will be half of those of 1997”). Important here was to induce raters to associate organizational performance with the leader’s tenure. Thus, an evident trend change in profits (heading up or down) occurred the year immediately after the director was appointed (i.e., in 2000).

The materials were pilot tested on 24 MBA students on an English-speaking program at a European university. Upon debriefing, these students reported that the situation was realistic and that the instructions and requirements seemed reasonable. Furthermore, the purpose of the experiment was not obvious. Finally, the students reported that the context that was depicted was concordant with the sex-typing expectations that we intended to provoke.

Procedures

Students participating in the experiment were informed that we were testing the reliability of a leadership questionnaire and a short personality measure. To be able to answer the questionnaire, they were informed that they first had to read a case study about an individual that they would later rate. Students first read the case study and then gave it back to the experimenter prior to completing the measures. In this way, we ensured that the ratings

they provided reflected what they encoded and retrieved from memory about the target. Once finished, the participants were fully debriefed.

Measures

Leadership and leader outcomes: we used the MLQ (Form 5X) (Bass & Avolio, 1995), to measure the factors constituting the full-range leadership model. This questionnaire is the best known measure of transformational and transactional leadership (Hunt, 1999; Lowe, Kroeck & Sivasubramaniam, 1996; Yukl, 1999).

Four items each were used as indicators of the leader style and between two to four items for outcomes factors. Five factors (i.e., attributed idealized influence, behavioral idealized influence, inspirational motivation, intellectual stimulation, individualized consideration) represented transformational leadership, which is considered prototypically effective. One factor (contingent rewards,) represented an active constructive form of transactional leadership, which is considered prototypically effective. Active management-by-exception represented an active corrective transaction leadership, which may be considered as prototypically ineffective depending on the context (Antonakis & House, 2002). The final transactional measure, passive management-by-exception, is considered as prototypically ineffective. The full-range model also includes laissez-faire leadership, the most prototypically ineffective form of leadership (see Bass, 1998, for full descriptions of the scales). Also measured by the MLQ are three outcome measures of leader effectiveness: the general effectiveness of the leader, the extent to which the leader makes followers exert extra effort, and the extent to which followers are satisfied with the leader.

Respondents judge how frequently the leader demonstrates the behavior in the item described using a Likert-type response scale, which includes “not at all,” “once in while,” “sometimes,” “fairly often,” and “frequently if not always.” The scale has a magnitude estimation ratio of 0:1:2:3:4 corresponding to the above descriptors (Bass, Cascio, &

O'Connor, 1974). A recent study based on a total of about 9,500 respondents rating actual leaders (in a variety of contexts but in predominantly business settings) and using confirmatory factor analysis provided support for the dimensionality of the theorized nine-factor MLQ model (see Antonakis, Avolio, & Sivasubramaniam, 2003).

Attribution of organizational responsibility: two items, which can be viewed as repeated measures of the same underlying construct, were used to gauge whether raters made an internal or external attribution regarding responsibility of organizational performance. Given that the leader was depicted in a team environment, and to make this measure realistic and unobtrusive, we asked participants to rate the likelihood that the director (internal attribution) or the team (external attribution) was responsible for the company's performance over the last two years.

Personality: we used five single-item measures to gauge the five-factor personality model (i.e., neuroticism, extraversion, openness, agreeableness, conscientiousness, Costa & McCrae, 1992). Participants were asked to judge the extent to which the target fit the personality description, for example, "Robin is emotionally unstable (i.e., is prone to being anxious, hostile, depressed, timid, impulsive, and vulnerable)."

Single-item measures are useful for providing broad description of personality while reducing completion time and rater fatigue (see Bernard, Walsh, & Mills, 2005). As mentioned by Langford (2003, p. 1139) the "predictive validity of abridged and single-item measures is similar to, and can sometimes equal or even exceed, the predictive validity of multi-item measures." We are not making the claim that our measures would be highly reliable or predictive. However, and given the terse description of the leader we provided, it would be trying to rate a target leader on an extended NEO-PI five factor type survey, which takes about 40 minutes to complete, and which has very specific questions that would be impossible to answer in the experimental conditions we created.

RESULTS

Manipulation checks

Measures were included to determine if the between-subjects treatments had the intended effects. To ascertain the salience of the sex manipulation, participants were asked to indicate whether Robin was a man or a woman. Participants chose the correct option in the respective condition, suggesting that the treatment was successfully perceived: $\chi^2(1) = 212.74$, $p < .001$. For the performance signal, participants were asked to indicate whether profits were rising or falling over the last two years. Again, this treatment was successful, $\chi^2(1) = 295.96$, $p < .001$. Finally, participants were asked to indicate whether the target leader was a director of a steelworks factory, food-processing company, women's department store, bank, travel agency, or a car dealership. Results indicated once more that the treatment had the intended effect, $\chi^2(5) = 277.74$, $p < .001$.

Ratings of leadership and leader outcomes

Interfactor correlations and factor reliabilities combined for all conditions are listed in Table 1. Because the reliability of the management-by-exception active scale was unacceptably low, it was not included in further analysis. Interafactor correlations, and correlations of factors to outcomes were in line with previous research based on real leaders (see Antonakis & House, 2002; Lowe et al., 1996). Thus, it appears that the leadership situations depicted in the vignettes were relatively realistic.

[Insert Table 1 about here]

To determine whether the sex of the participants was related to any of the 17 dependent variables (8 leader style, 3 outcome, 1 attribution, 5 personality), we treated the sex of participant as a between-subjects factor and estimated a four-way ANOVA model. With three exceptions¹, results indicated that the main effect of sex as well as the interaction of sex

with the treatments factors (two-way, three-way and four-way) was not significant. Following conventional practice (e.g., Heilman, et al., 2004) we thus combined responses of male and female participants for the dependent variables that were unaffected by participant sex.

Means and standard deviations of the dependent variables are listed in Table 2.

Because the patterns of responses for the transformational leader scales were very similar and for reasons of parsimony and to improve reliability, we combined them to form a “transformational leadership” composite factor (12 items, $\alpha = .90$). For the same reasoning, we combined the passive management-by-exception and laissez-faire leadership to form a “passive-avoidant” composite factor (6 items, $\alpha = .72$).

Multivariate analysis of variance indicated that the predicted three-way interaction was significant for ratings of leadership style when using the seven leader scales as dependent variables: Wilk’s $\Lambda = .95$, $F(7,335) = 2.50$, $p < .05$, partial $\eta^2 = .05$. Thus, we do not report main effects and two-way interactions because they are qualified by the three-way interaction. Similarly, MANOVA indicated the three-way interaction was significant for ratings of leadership outcomes: Wilk’s $\Lambda = .98$, $F(2,339) = 3.68$, $p < .05$, partial $\eta^2 = .02$. As regards the parsimonious scales of transformational and passive-avoidant leadership results were similar: Wilk’s $\Lambda = .97$, $F(2,340) = 6.20$, $p < .01$, partial $\eta^2 = .04$. Hypotheses 1 and 2 were supported.

Univariate tests indicated that the three-way interaction held for the following five leader scales²: attributed idealized influence, $F(1, 341) = 6.18$, $p < .05$, partial $\eta^2 = .02$, r^2 for the full regression model, with main effects and all interactions = .03; intellectual stimulation $F(1, 341) = 3.60$, $p = .06$, partial $\eta^2 = .01$, r^2 for the full regression model = .03; individualized consideration, $F(1, 341) = 4.46$, $p < .05$, partial $\eta^2 = .01$, r^2 for the full regression model, with main effects and all interactions = .02; passive management-by-exception, $F(1, 341) = 4.97$, $p < .05$, partial $\eta^2 = .01$, r^2 for the full regression model = .12; and laissez-faire leadership, $F(1, 341) = 11.91$, $p < .001$, partial $\eta^2 = .03$, r^2 for the full regression model = .07. Similarly,

univariate tests indicated a significant effect on the two parsimonious leader scales, transformational leadership, $F(1, 341) = 5.45$, $p < .05$, partial $\eta^2 = .02$, r^2 for the full regression model = .03; and passive-avoidant leadership, $F(1, 341) = 9.85$, $p < .01$, partial $\eta^2 = .03$, r^2 for the full regression model = .11. Means for the dependent variables in each of the conditions are reported in Table 2.

[Insert Table 2 about here]

As regards the leader outcomes scales, univariate tests also indicated that only the extra effort outcome measure was affected by the three way-interaction³, $F(1, 340) = 5.54$, $p < .05$, partial $\eta^2 = .02$, r^2 for the full regression model = .05. The above results provide further, yet qualified, support for Hypothesis 2.

Next, we plotted the three-way interactions to better understand how ratings were affected by the independent variables. For the leader scales, we report only the parsimonious factors given that the interaction plots were similar for the subfactors.

[Insert Figure 2 about here]

As indicated by the patterns of ratings in Figure 2, ratings of transformational leadership did not follow our theorizing entirely. As predicted, and with respect to the form of the interaction, the female leader was rated lower than was the male when participants were exposed to bad performance cues in a masculine context, partially supporting Hypothesis 1a. In the same condition and irrespective of performance cues, ratings of the male leader remained relatively unaffected, supporting the “tempering” effect argument we made above and partially supporting Hypothesis 1b. Thus, three of the four points of the interaction form were in line with our theorizing. Although we predicted an amplification effect for the female leader (Hypothesis 1b), the effect was substantially larger than what we predicted. The female leader was rated higher than was the male when participants were exposed to good performance cues in a masculine context.

Lending credence to the stability and salience of these findings, and—quite remarkably—the male leader in the feminine context was evaluated in exactly the same way that the female leader was (i.e., amplified ratings with a “rebound” effect for the male leader whereas more tempered ratings for the female leader). These results provide partial support for Hypotheses 1c and 1d.

Turning to ratings of passive-avoidant leadership, and as depicted in Figure 3, the same telltale tempering-amplification effects were evident for in-context and out-of-context leaders respectively. These results provide partial support for Hypothesis 1e, 1f, 1g, and 1h.

[Insert Figure 3 about here]

The above results were probed with simple effects analyses. The multivariate simple effects of the performance cue factor within each level of combination of the other factors indicated that the cue only had a significant effect on ratings of transformational and passive-avoidant leadership for the female leader when in the bank, Wilk's $\Lambda = .96$, $F(2,340) = 6.69$, $p < .01$, partial $\eta^2 = .04$. Ratings for the male leader were unaffected in the bank context. Univariate simple effects analyses indicated that the significant effects were evident for ratings of transformational as well as passive-avoidant leadership for out-of-context leaders only. These results provide further support for Hypothesis 1b.

Conversely the cue factor only had a significant effect on ratings of transformational and passive-avoidant leadership for the male leader when in the store Wilk's $\Lambda = .94$, $F(2,340) = 11.84$, $p < .01$, partial $\eta^2 = .07$. This affect was not evident for the female leader in the store. Again, univariate simple effects analyses indicated that the significant effects were evident for ratings of transformational as well as passive-avoidant leadership for out-of-context leaders only. Hypothesis 1d was further supported.

To complete this series of analyses, we plotted the interaction concerning the leader effectiveness outcome measure (i.e., subordinate extra effort). We have not included the plot here because its form is the same as that depicted in Figure 2.

As with Hypothesis 1, the same patterns of results were again evident. Ratings of out-of-context leaders were amplified and received a “rebound” effect, whereas ratings of in-context leaders were more restrained. The form of the interaction was surprisingly similar to the previous interactions both for the feminine and the masculine context. As for the previous findings, these results provided partial support for Hypotheses 2a, 2b, 2c, and 2d. These results were further probed with simple effects analyses. In line with the previous results, univariate simple effects analyses indicated that the cue was salient only on ratings of out of context leaders. Further support was provided for Hypotheses 2b and 2d.

Performance attributions

Recall that participants were asked to rate the likelihood that the director (internal attribution) or the team (external attribution) were responsible for the company’s performance over the last two years. To determine whether participants made an internal or external attribution, and to facilitate interpretation of this effect, we subtracted external attributions ratings from ratings of internal attributions concerning each leader. Positive values thus indicated that participants attributed organizational results to the leader. Negative values indicate that participants attributed organizational results to the team (note: using repeated-measures ANOVA would produce the same substantive results). The means and standard deviations of the performance attribution variable are listed in Table 2.

Hypothesis 3 was not supported. Instead we found a main effect for organizational performance cues, $F(1, 339) = 13.47, p < .001, \text{partial } \eta^2 = .04$, indicating that, independent of the other treatments, leaders were attributed (blamed) for bad organizational outcomes and teams were attributed (credited) for good organizational outcomes. Furthermore, we found a

two-way leader sex*job-sex type interaction $F(1, 339) = 6.18, p < .05, \text{partial } \eta^2 = .02$. The r^2 for the full regression model was .06. This interaction is depicted graphically in Figure 5.

[Insert Figure 4 about here]

These results indicated that out-of-context leaders are held more accountable than are in-context leaders. Because in all instance leaders are blamed more than teams are for bad organizational outcomes, these results indicate that (a) out-of-context leaders are more likely to be blamed for bad performance, as partially predicted by Hypotheses 3b and 3d, and (b) teams are more likely to be given credit for good outcomes when led by in-context leaders.

Ratings of personality

As expected, the female leader was seen as more neurotic (emotionally unstable, $\underline{m} = 1.58$) than was the male leader ($\underline{m} = 1.28$), $F(1, 336) = 5.73, p < .05, \text{partial } \eta^2 = .02$, supporting Hypothesis 4a. Results also indicated that performance cues had the expected main effects such that leaders rated in the presence of good cues were seen as (a) more extraverted ($\underline{m} = 2.26$) than “bad-cue” leaders ($\underline{m} = 1.84$), $F(1, 336) = 9.56, p < .001, \text{partial } \eta^2 = .03$, (b) more open ($\underline{m} = 1.98$) than “bad-cue” leaders ($\underline{m} = 1.71$), $F(1, 336) = 3.83, p = .05, \text{partial } \eta^2 = .01$, and (c) less neurotic ($\underline{m} = 1.30$) than “bad-cue” leaders ($\underline{m} = 1.54$), $F(1, 336) = 3.53, p = .06, \text{partial } \eta^2 = .01$. Additionally, “good-cue” leaders were rated higher on conscientiousness ($\underline{m} = 2.84$) than were “bad-cue” leaders ($\underline{m} = 2.39$), $F(1, 336) = 13.85, p < .001, \text{partial } \eta^2 = .04$. These results provide partial support for Hypothesis 4b.

In addition, none of the two-way interactions were significant nor were any of the three-way interactions. However, marginally significant three-way interactions were found for ratings of openness and conscientiousness whose form followed the “tempered-rebound” patterns described previously, thus qualifying the above main-effects results.

The results regarding the personality-leadership style link are reported in Table 4. We used multiple regression analysis to predict leadership style from the five-factor personality

model. Because the predictors in this analysis were outcomes of the manipulated variables, they were not independent from them. Therefore, it would not make any sense to use the personality predictors as covariates of the treatment variables. Furthermore, the personality measures were gathered after the leader measures were. Thus, the personality variables might have depended on how subjects rated the leader, suggesting that the direction of causality for these analysis flows from leadership to personality and not vice-versa. Furthermore, because the independent and dependent variables were both outcomes of the same treatment, these analyses should be interpreted with caution.

Across all conditions, results indicated that the big five personality factors predicted substantial amounts of variance in the dependent measures. As expected, extraversion was positively related to transformational and contingent reward leadership and leader outcomes; however, it was not predictive of passive-avoidant leadership. Also, openness was positively related to transformational and contingent reward leadership and leader outcomes. It was also negatively related to passive management-by-exception. Neuroticism was only positively related to the passive-avoidant leader measures. These results provide partial support for Hypothesis 4c. Unexpectedly, agreeableness was positively related to three of the transformational leadership measures as well as to the outcome measures. Conscientiousness was positively related to four transformational leadership measures, contingent reward leadership and leader outcomes. It was also negatively related to laissez-faire leadership and passive-avoidant leadership.

DISCUSSION

The major findings of this study, particularly regarding the "rebound" effect, can be summed with an updated, leader-appropriate, and gender inclusive version of Henry Wadsworth Longfellow's nursery rhyme, "There was a little girl"

Original verse

There was a little girl
Who had a little curl

Modern version

There were two leaders
A man and woman

Right in the middle of her forehead
 When she was good
 She was very good indeed
 But when she was bad she was horrid!

Who were simply not in their context
 When they were good
 They were very good indeed
 But when they were bad they were horrid!

As our results indicated, raters use cognitive distortion mechanisms when rating leaders.

Although exposed to precisely the same stimulus materials, participants used schematic-driven "fill-in-the-blank" procedures that depended on the treatment combinations to which they were exposed.

We correctly predicted that male and female leaders exhibiting the same behaviors would not be judged equivalently. In the presence of bad organizational performance signals, leaders who are out-of-context were stereotyped such that ratings of their leadership style, vis-à-vis an in-context leader, would be devalued. The surprising, but theoretically interpretable finding was that out-of-context leaders were overvalued when raters were exposed to good organizational performance signals. We discuss this finding in detail below, because of its potential contribution to the extant literature.

We briefly discuss the other findings first, which albeit interesting from a scientific replication perspective, do not have the theoretical implications of our "rebound" phenomenon. As concerns attributions of organizational performance, organizational success or failure was not attributed in the same way to male and female leaders. Following our predictions, out-of-context leaders were held accountable for bad organizational outcomes; however, in-context leaders were not given credit for organizational success. The latter finding is intriguing and puzzling. Apart from sample-specific idiosyncrasies that might have influenced these results, perhaps raters viewed "high fit" contexts as being on "autopilot." That is, all things being equal, the leader knows what he or she has to do in that particular context. So do the followers. Given that work teams have more followers than leaders, rationally and economically speaking, participants might have simply "added the numbers," attributing good performance mostly to the team. Indeed, the team works collectively more

than the leader does, thus they should be credited mostly for organizational outcomes. These findings should be explored in future research.

As regards personality inferences, differential evaluations were again evident. Female leaders were seen as more neurotic (emotionally unstable) than males were across all conditions, in line with previous findings. Results also showed that organizational performance cues tainted the dispositional inferences that were made. Under good performance cues leaders were seen to exhibit prototypically "good" dispositions (higher extraversion, higher openness, higher emotional stability) more often than leaders who were associated with bad performance cues.

Finally, personality factors were associated with successful leader behavior in way that generally paralleled previous studies. Across all conditions, openness and extraversion were consistently associated with effective leader behavior and outcomes. Neuroticism, agreeableness, and conscientiousness also showed some predictive validity.

Implications for theory

The most intriguing finding in our study was the overvaluation of an out-of-role leader. The polarization (accentuation) of ratings for out-of-context targets is interpretable as corroborated by a small body of literature. We found three independent theoretical models that provide very similar explanations to this polarization phenomenon: equity theory, amplification-ambivalence bias, and the shifting standards model. Important to note here is that the majority of work that has been carried out on stereotyping has used and tested conventional models, which predict that ambiguous performance or equal performance (in our case bad organizational outcomes) would be seen more negatively for out-of-context individuals. In the case of good performance, most theoretical models would predict that the performance of out-of-context targets would be devalued; however, the precise opposite

occurred, which—as Heilman, Martell, and Simon (1988) put it—is “totally inconsistent with ideas about stereotypes underlying the sex discrimination process.”

We believe that the reason why this unexpected finding occurred was because the performance manipulations we used led participants to believe that outcomes were exceptionally good (which, based on how we presented them seems to be the case). Outstanding performance is the exception to the rule, which when observed for an individual in an out-of-context situation will be overrewarded. That is, when behavior or performance signals are such that they "exceed a critical threshold, the break with the stereotyped can cause a boomerang [rebound] effect" (Heilman, 2001, p. 666). This phenomenon—an underinvestigated and relatively recent addition to the stereotyping literature (Eagly & Chaiken, 1993; Heilman et al., 1988)—has been observed in various settings including leadership performance (Jacobson & Effertz, 1974), performance in an emergency situation (Taynor & Deaux, 1973, 1975), ratings of vocational competence (Abramson, Goldberg, Greenberg, & Abramson, 1977), evaluation of job applicants (Heilman et al., 1988), and employee performance ratings (Bigoness, 1976; Hamner, Kim, Baird, & Bigoness, 1974).

Equity theory provides one potential explanation (see Taynor & Deaux, 1973, 1975). Equity, in social exchange processes exists when the ratio of inputs (e.g., sex, age, ethnicity, effort, performance, etc.) to outputs (payments, rewards, status, etc.) of a target individual equals the ratio of inputs to outputs of a referent individual (Adams, 1965). Applied to our case, equity occurs when exceptional performance is compensated appropriately as compared with the referent individual (note: the referent could be the person with whom the perceiver is transacting or a third party with whom the perceiver compares himself/herself). Because outputs can be constrained by certain inputs that are voluntary (e.g., performance) or involuntary (e.g., sex), performance should be rewarded more highly to the extent that an involuntary constraint is a limiting factor on an individual's output (Taynor & Deaux, 1975).

In other words, if a target's sex (e.g., a woman in a male sex-typed job) seems to be a limiting factor or a "handicap" of sorts to the perceiver then the perceiver will upwardly adjust outputs to "reward" the target for performance that would be equal to that of an individual that is not operating under the same constraint (e.g., a male leader in a male sex-typed job).

A second theoretical approach, that has been labeled the amplification-ambivalence bias has been advanced by Katz and colleagues (see Katz & Glass, 1979; Katz, Glass, & Cohen, 1973; Katz, Hass, & Wackenhut, 1986; Katz & Hass, 1988). This approach suggests that minority individuals elicit ambivalent feelings. On one hand, perceivers feel hostile and aversive toward minorities because these individuals have deviant qualities that are contrary to the qualities expected of a prototypical individual. This perspective is particularly prevalent in cultures valuing individualism. In such cultures, individual responsibility and self-reliance is stressed (Katz et al., 1986), which suggests that the minorities are themselves responsible for the situation in which they are. Theoretically, situational factors are ignored and the perceiver makes the fundamental attribution error (or correspondence bias, see Gilbert & Malone, 1995) by equating outcome (lower status) with disposition (lower ability/motivation/etc.). On the other hand, because of holding simultaneous egalitarian values, perceivers feel sympathy and compassion for minorities because minority individuals are disadvantaged (suggesting lenient ratings in contexts where the minority figure provides unquestionably good performance). Thus, ratings of such individuals will elicit extremely polarized responses as a function of situational factors (i.e., organizational performance cues, as in our experiment). In our case, if individuals perform exceptionally well, they are overrewarded. If individuals perform exceptionally poorly, then they are overly blamed.

The third perspective has been advanced by Biernat and colleagues (see Biernat, 2003 for a review) who have focused on counterstereotypical or contrast effects. From this perspective, out-of-role individuals are not judged on the same standard as in-role individuals.

Because of stereotypic expectations (e.g., "men are better leaders than women are"), perceivers will judge women on a lower subjective (within-category) standard relative to men, even though evaluations using objective (cross-category) performance criteria might lead to a different result. Thus, even though, objectively speaking two individuals are described in the same way, the perceiver will judge the out-of-context individual on a lower standard and the in-context individual on a higher standard. In this way, and particularly in the case where outcome criteria are indubitably exceptional and prototypically good, the performance of the out-of-context individual will be inflated as compared with the in-context individual

The above models all come to the same conclusion, all stemming from expectancy based information processing. Judgments about target individuals who are not in context are influenced by stereotypes, which make salient certain expectations about the target individual. If performance outcomes are exceptionally and unexpectedly good (or bad) the individual will be judged as having produced superior (inferior) performance relative to an in-context individual. Thus, low expectations make successful performance even more remarkable. As succinctly stated by Abramson et al., "when an individual achieves a level of success not anticipated, his/her achievement tends to be magnified rather than diminished. After all, it matters little what the platypus says, the wonder is that it can say anything at all" (p. 123).

Practical implications

The most important practical implication concerns the development of systems that will reduce potential biases regarding the evaluation of leaders. Apart from personality ratings, our results generally demonstrated that men and women leaders benefit or are punished equally, depending on context congeniality. Thus, those who bruit it about that women are systematically disadvantaged must have it wrong. However, as Heilman (2001) has cautioned, even when demonstrating good performance, women's career progressions might continue to be jeopardized because they are still violating prescriptive stereotypes and thus will be liked

less than will be males (Heilman et al., 2004). More importantly, and assuming that men and women are treated in an egalitarian manner, “because few organizational [leadership] positions are believed to require predominantly feminine skill for success, and these positions are rarely upper levels ones, women bear the brunt of sex bias in efforts to advance up the organizational ladder” (Heilman, 1983, p. 294). To get to the top women will have to show exceptionally good performance. Thus, the proportion of women who might get there will be small. This consequence may explain why female leaders generally exhibit better behaviors than male leaders—those that made it must be really good (Eagly & Carli, 2004).

To begin to change this dispensation, organizational systems, from schools to business and governments, should take an active stance in promoting egalitarian practices, creating awareness about biases in ratings, and in engaging in "sensemaking" activities that are gender inclusive. Furthermore, more female leader role models are required so that individuals begin to associate females with leadership positions in what were previously male-dominated vocations. To reiterate, gender role expectations are culturally constructed (Hofstede, 1991). They should thus be amenable to gradual change. To stress the point that culturally-induced prototypes affect progression of women to power, take the case of Sweden, which is considered a "feminine" society (i.e., gender role expectations are fuzzy and individuals can engage in a variety of activities or vocations without being penalized). Women make up 45% of representatives in parliament (Global Database of Quotas for Women, 2004). However, in the U.S., which can be characterized as mostly masculine, only 14 women serve in the senate and 63 in the congress, representing about 15% of total elected representatives (US Senate, 2004).

Apart from the ethical consequences of promoting equality in the workplace there are economic implications. If selection systems are not taking the best candidates, organizational systems might not operate effectively. Interestingly, a recent report examining 353 of the

Fortune 500 companies suggested that companies with the highest proportion of top management women executives had a Return on Equity (ROE) that was 35.1% higher than the companies with the lowest proportion (Catalyst, 2004). Certainly, a variety of factors and causal models could explain this phenomenon. However, the fact that (a) women leaders tend to exhibit active-constructive leader behaviors more often than do men (who in turn tend to exhibit passive-reactive behaviors more often than do women, Antonakis et al., 2003; Eagly et al., 2003) and (b) these active-constructive behaviors are more strongly correlated to objective and subjective performance outcomes than passive reactive behaviors (see Lowe, Kroeck, & Sivasubramaniam, 1996) suggests that companies with more diversified top management teams might actually be more effective than companies that are male dominated at the top echelons.

Limitations and suggestions for future research

One limitation is that we used stimulus material that was presented in writing. Our results however, were similar to studies that used more information rich media. The use of students in an artificial setting is also a concern. To the extent that they partially reflect how current and future decision makers might make judgments about leaders in the real world our experimental approach was needed, particularly in situations where nonconscious information processes, untainted from extraneous factors, require examination (see Brown & Lord, 1999). Finally, experimental results are more congruent with field experiments than is generally thought (Anderson, Lindsay, & Bushman, 1999).

Even if these effects are realistic, they are limited in that they can be qualified as small. Although scholars have been critical of the importance of such small effects or the fact that antecedent and moderator conditions still need to be better established in studying stereotyping phenomena (cf. Swim, Borgida, & Maruyama, 1989), it is important to note that effects such as these should be viewed in terms of their long-term practical consequences.

Similarly to the way bank interest is compounded, small sex differences have larger and larger effects as they influence proportions to larger degrees (i.e., from lower to higher hierarchical managerial levels, see Martell, Lane, & Emrich, 1996). For example, if only 1% (or 5%) variance of ratings—qualified as small effects—is attributed to sex, the proportion of women at top organizational levels will be reduced to 35% (29%).

Finally, future research should examine how raters respond to amount of organizational performance cues as manipulated factors. That is, in one condition, raters should be exposed to cues that are numerous in nature, whereas in the other condition raters should be exposed to fewer, possibly one cue. If, as we have theorized, the indubitable nature of the cues augmented by repeated reference to them caused the amplification-ambivalence bias then reducing the frequency of reference to the cues should eliminate this effect. Future research could also examine how manipulating the level (exceptional versus average performance) of the cue affects outcomes.

Conclusion

A quote often attributed to Maureen Reagan succinctly summarizes the status quo about apparent equality between the sexes, as well as the unequal treatment out-of-context individuals receive: "I will feel equality has arrived when we can elect to office women who are as incompetent as some of the men who are already there⁴."

Of course, we are not suggesting that we need to promote incompetent women to positions of power nor are we implying that many men in positions of power are incompetent. What we are suggesting, though, is that veritable equal treatment should be promoted. Equal evaluation can only occur when we better understand the sources of biases and how they can be minimized.

TABLE 1: ALPHA RELIABILITIES AND FACTOR CORRELATIONS OF MLQ SCALES AND OUTCOMES

	IIA	IIB	IM	IS	IC	CR	MBEP	LF	EFF	EE	SAT
1. IIA	.78¹										
2. IIB	.62**	.79¹									
3. IM	.56**	.61**	.79								
4. IS	.69**	.57**	.44**	.79							
5. IC	.71**	.57**	.45**	.69**	.75¹						
6. CR	.64**	.63**	.57**	.59**	.57**	.69¹					
7. MBEP	-.23**	-.18**	-.25**	-.25**	-.22**	-.22	.75¹				
8. LF	-.20**	-.21**	-.31**	-.17**	-.13*	-.23**	.46**	.68¹			
9. EFF	.66**	.56**	.59**	.59**	.62**	.60**	-.34**	-.28**	.77		
10. EE	.71**	.56**	.55**	.66**	.71**	.62**	-.28**	-.23**	.77**	.82	
11. SAT	.68**	.52**	.55**	.66**	.65**	.62**	-.28**	-.20**	.75**	.77**	.77

Note. Alpha reliabilities are reported on the diagonal; ¹for these scales, one to two items exhibited low item-total correlations and were thus deleted from the scale (the resulting alpha for the scale was corrected with the Spearman-Brown prophecy formula); IIA = attributed idealized influence; IIB = behavioral idealized influence; IM = inspirational motivation; IS = intellectual stimulation; IC = individualized consideration; CR = contingent reward; MBEA = active management-by-exception; MBEP = passive management-by-exception; LF = laissez-faire leadership; EFF = effectiveness; EE = extra effort; SAT = satisfaction of followers. $n = 349$. * $p < .05$; ** $p < .01$.

TABLE 2: MEANS AND STANDARD DEVIATIONS OF LEADER SCALES, OUTCOMES, AND PERFORMANCE ATTRIBUTIONS

Context Factor	<u>Bank</u>								<u>Store</u>							
	<u>Man</u>				<u>Woman</u>				<u>Man</u>				<u>Woman</u>			
	<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>	
Cue Factor	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. TF	1.86	.78	1.88	.72	2.05	.87	1.65	.78	2.08	.80	1.72	.80	1.95	.72	1.94	.76
2. IIA	1.95	.83	1.95	.66	2.06	1.11	1.67	.90	2.18	.83	1.71	.94	1.90	.83	1.97	.80
3. IIB	2.12	.96	1.99	.80	2.40	.93	1.86	1.03	2.31	1.00	1.85	.83	2.42	.90	2.15	.96
4. IS	1.67	.94	1.83	.94	2.03	.96	1.57	.94	1.95	.97	1.64	.88	1.96	.92	1.78	.87
5. IC	1.83	.94	1.81	.89	1.86	.99	1.60	.79	2.01	.91	1.77	.90	1.67	.84	2.00	.89
6. CR	2.20	.94	2.04	.86	2.26	.91	1.96	.86	2.39	.76	1.90	.83	2.29	.65	2.04	.95
7. PA	1.51	.83	1.73	.80	1.45	.64	1.99	.74	1.38	.69	2.11	.65	1.84	.75	1.89	.70
8. MBEP	1.91	.96	2.39	1.02	2.08	.75	2.62	.81	1.89	.82	2.74	.81	2.47	.84	2.55	.89
9. LF	1.23	1.05	1.22	.92	.80	.93	1.64	1.13	.99	.94	1.60	.94	1.27	1.08	1.26	.93
10. EE	2.05	1.06	1.99	.86	2.33	.98	1.74	.92	2.34	1.02	1.70	1.13	2.03	.75	1.87	1.09
11. SAT	2.08	1.11	1.97	1.02	1.99	1.18	1.81	.90	2.19	.98	1.60	1.13	1.96	.98	1.71	1.17
12. ATTR	-.80	2.07	-.06	2.18	-.38	1.73	.54	2.13	-.51	1.99	.76	2.16	-.63	1.89	-.30	2.22

Note: IIA = attributed idealized influence; IIB = behavioral idealized influence; IS = intellectual stimulation; IC = individualized consideration; TF=transformational leadership (mean of transformational leadership items); CR = contingent reward; MBEP = passive management-by-exception; LF = laissez-faire leadership; PA=passive avoidant leadership (mean of passive management-by-exception and laissez-faire leadership items); EE = extra effort; SAT = satisfaction of followers; OUT = Outcomes (mean of the extra effort and satisfaction items); ATTR=attribution of organizational to leader (positive values) or team (negative values) .

TABLE 3: MEANS AND STANDARD DEVIATIONS OF PERSONALITY SCALES

Context Factor	<u>Bank</u>								<u>Store</u>							
	<u>Man</u>				<u>Woman</u>				<u>Man</u>				<u>Woman</u>			
	<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>		<u>Good</u>		<u>Bad</u>	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
N	1.27	1.16	1.22	1.10	1.38	1.21	1.73	1.30	1.08	1.07	1.50	1.29	1.47	1.31	1.73	1.09
O	1.93	1.23	1.98	1.24	2.10	1.45	1.53	1.15	2.10	1.31	1.57	1.31	1.79	1.38	1.75	1.32
A	1.78	1.08	2.02	1.07	1.73	1.30	1.78	1.19	2.00	1.20	1.98	1.30	1.68	1.14	2.08	1.35
C	2.64	1.19	2.47	1.10	2.70	1.30	2.33	1.12	3.08	.92	2.14	1.14	2.92	1.12	2.60	1.08

Table 4: Predicting Leadership with the five factor model of personality

Factor	Adjusted R-Square	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
		β	β	β	β	β
IIA	.48***	-.07	.16***	.28***	.24***	.19***
IIB	.32***	.03	.30***	.23***	.04	.20***
IM	.37***	-.07	.34***	.11**	-.01	.33***
IC	.47***	-.03	.14***	.43***	.20***	.04
IS	.45***	.02	.12***	.39***	.23***	.11*
CR	.38***	-.00	.33***	.21***	.09	.20***
MBP	.11***	.18***	-.09	-.16*	.03	-.09
LF	.11***	.16***	-.09	-.09	.11	-.22***
EFF	.44***	.01	.23***	.26***	.12*	.30***
EE	.46***	.01	.21***	.33***	.17**	.19***
SAT	.51***	.01	.15**	.36***	.23***	.22***
TF	.58***	-.01	.19***	.40***	.22***	.15***
PA	.17***	.23***	-.11	-.14*	.08	-.17***

Note. β = standardized beta coefficient; IIA = attributed idealized influence; IIB = behavioral idealized influence; IM = inspirational motivation; IC = individualized consideration; IS = intellectual stimulation; CR = contingent rewards; MBP = passive management-by-exception; LF = laissez-faire leadership; EFF = leader general effectiveness; EE = subordinate extra effort; SAT = subordinate effectiveness; TF = transformational leadership style; PA = passive avoidant leadership style. *** $p < .01$; ** $p < .10$; * $p < .05$.

Figure 1: Hypothesized three-way interaction for ratings of prototypically good leadership or effective leader outcomes

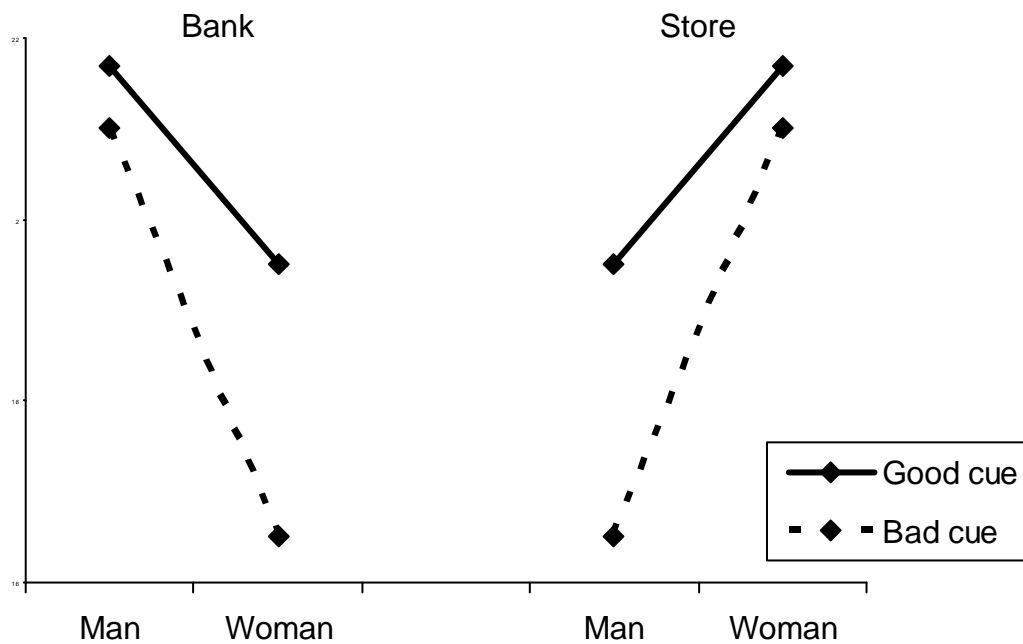


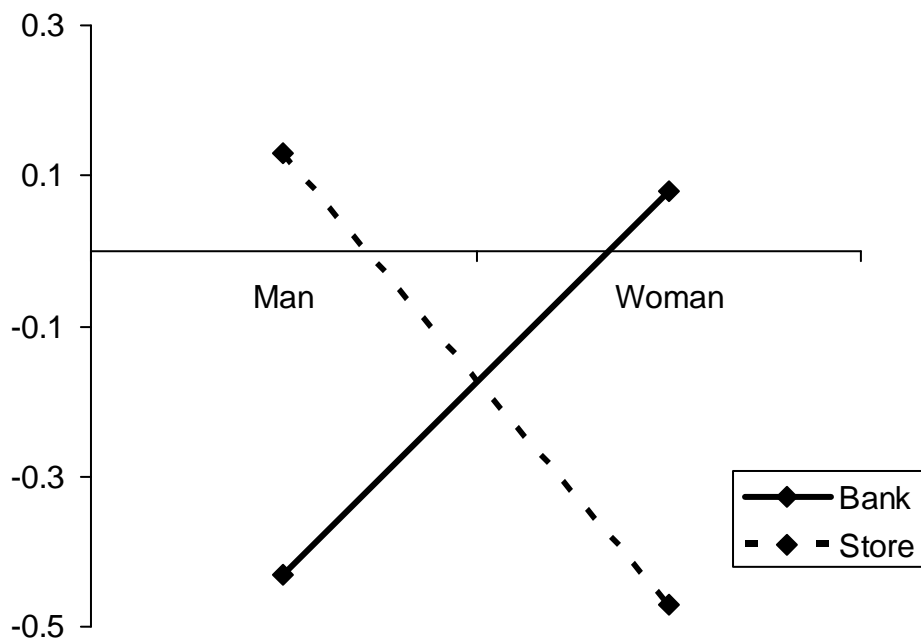
Figure 2: Three-way interaction on ratings of transformational leadership



Figure 3: Three-way interaction on ratings of passive-avoidant leadership



Figure 4: Two-way interaction on attributions of organizational outcomes



REFERENCES

- Abramson, P.R., Goldberg, P.A., Greenberg, J.H., & Abramson, L.M. (1977). The talking platypus phenomenon: Competency ratings as a function of sex and professional status. *Psychology of Women Quarterly*, 2, 114-125.
- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, 2, (pp. 267-299). NY: Academic Press.
- Anderson, C. A., Lindsay, J. L., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8, 3-9.
- Antonakis, J., Avolio, B.J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine- factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly*, 14, 261-295.
- Antonakis, J., Cianciolo, A. T., & Sternberg, R. J. 2004. Leadership: Past, present, and future. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.). *The nature of leadership* (pp. 3-15). Thousand Oaks, CA: Sage Publications.
- Antonakis, J., & House, R.J. (2002). The Full-Range Leadership Theory: the way forward. In: B.J. Avolio & F.J. Yammarino (Eds.). *Transformational and Charismatic Leadership: The Road Ahead* (pp. 3-34). Amsterdam: JAI Press.
- Bass, B. M., & Avolio, B. J. (1995). *MLQ Multifactor leadership questionnaire for research: Permission set*. Redwood City, CA: Mindgarden
- Bass, B. M. (1980). *Bass and Stogdill's handbook of leadership* (2rd ed.). New York: Free Press.
- Bass, B. M. (1998). *Transformational leadership: Industrial, military, and educational impact*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, 59, 313-320.

- Bennis, W. (2004). The crucibles of authentic leadership. In: J. Antonakis, A.T. Cianciolo, & R.J. Sternberg (Eds.). *The Nature of Leadership* (pp. 331-342). Thousand Oaks, CA: Sage Publications.
- Bernard, L. C., Walsh, R. P., & Mills, M. (2005). Ask once, may tell: Comparative validity of single and multiple item measurement of the big-five personality factors. *Counseling and Clinical Psychology Journal*, 2, 40-57.
- Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, 58, 1019-1027.
- Bigness, W. J., (1976). Effect of applicant's sex, race, and performance on employer's performance ratings: Some additional findings. *Journal of Applied Psychology*, 61, 80-84.
- Binning, J. F., Zaba, A. J., & Whattam, J. C. (1986). Explaining the biasing effects of performance cues in terms of cognitive categorization. *Academy of Management Journal*, 28, 521-535.
- Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, 89, 901-910.
- Brown, D. J., & Lord, R. G. (1999). The utility of experimental research in the study of transformational/charismatic leadership. *Leadership Quarterly*, 10, 531-539.
- Brown, D.J., Scott, K.A., & Lewis, H. (2004). Information processing and leadership. In: J. Antonakis, A.T. Cianciolo, & R.J. Sternberg (Eds.). *The Nature of Leadership* (pp 125-147). Thousand Oaks, CA: Sage Publications.
- Bryman, A. (1987). The generalizability of implicit leadership theory. *Journal of Social Psychology*, 127, 129-141.
- Burns, J. M. (1978). *Leadership*. NY: Harper & Row.

- Butler, D., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: Implications for leader evaluations. *Journal of Personality and Social Psychology, 58*, 48-59.
- Butterfield D. A., & Grinnell, J. P. (1999). "Re-viewing" gender, leadership, and managerial behavior. In: G. N. Powell (Ed.) *Handbook of Gender & Work* (pp. 223-238). Thousand Oaks: Sage Publications.
- Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35*, 38-48.
- Cantor, N., & Mischel., W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, (Vol. 12, pp. 3-52). New York: Academic Press.
- Catalyst (2004). *The bottom line: Connecting performance and gender diversity*. New York: Catalyst.
- Cejka, M.A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*, 413-423.
- Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. *Journal of Applied Psychology, 60*, 566-572.
- Costa, P. T., & McCrae, R. R. (1992). *NEO-PI professional manual*. Lutz, FL: Psychological Assessment Resources.
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225-248.
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology, 29*, 80-85.

- Eagly, A. H., & Carli, L. L. (2004). Women and men as leaders. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.). *The Nature of Leadership* (pp. 279-301). Thousand Oaks: Sage Publications.
- Eagly, A. H. & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Eagly, A. H. , & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology, 60*, 685-710.
- Eagly, A. H. , & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573-598.
- Eagly, A. H. , & Johannesen-Schmidt, M. C. (2001). The leadership styles of women and men. *Journal of Social Issues, 57*, 781-797.
- Eagly, A. H. , Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez- faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin, 129*, 569-591.
- Eagly, A. H. , Karau, S. J. , & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin, 117*, 125-145.
- Eagly, A. H. , Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*, 3-22.
- Erickson, D. J., & Krull, D. S. (1999). Distinguishing judgments about *what* from judgments about *why*: Effects of behavior extremity on correspondent inferences and causal attributions. *Basic and Applied Psychology, 21*, 1-11.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*, 127-148.
- Fiske, S. T. (1995). Social cognition. In A. Tesser (Ed.). *Advanced social psychology* (pp. 149-193). Boston, MA: McGraw Hill.

- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 26, 21-42.
- Gardner, J. (1990). *On leadership*. New York: Free Press.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-38.
- 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, 619-632.
- Gioia, D. A., & Sims, H. P. (1985). On avoiding the influence of implicit leadership theories in leader behavior descriptions. *Educational and Psychological Measurement*, 45, 217-232.
- Global Database of Quotas for Women (2004). <http://www.quotaproject.org/index.cfm>.
[retrieved March 8, 2005]
- Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. (1974). Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. *Journal of Applied Psychology*, 59, 705-711.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, 5, 269-298.
- Heilman, M. E. (1984). Information as a deterrent against sex discrimination: The effects of applicant sex and information type on preliminary employment decisions. *Organizational Behavior and Human Performance*, 33, 174-186.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57, 657-674.

- Heilman, M. E. , Block, C. J., & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality, 10*, 237-252.
- Heilman, M. E. , Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology, 74*, 935-942.
- Heilman, M. E. , Martell, R. F., & Simon, M. C. (1988). The vagaries of sex bias: conditions regulating the undervaluation, equivaluation, and overvaluation of female job applicants. *Organizational Behavior and Human Decision Processes, 41*, 98-110.
- Heilman, M. E. , Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male tasks. *Journal of Applied Psychology, 89*, 416-427.
- Hofstede, G. (1991). *Cultures and organizations: Software of the mind*. Berkshire, England: McGraw-Hill Book Company.
- Hunt, J. G. (1999). Transformational/charismatic leadership's transformation of the field: An historical essay. *The Leadership Quarterly, 10*, 129-144.
- Hyde, J. S., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly, 25*, 364-378.
- Jacobson M. B., & Effertz, J. (1974). Sex roles and leadership: Perceptions of the leaders and the led. *Organizational behavior and human performance, 12*, 383-396.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and Leadership: A Qualitative and Quantitative Review. *Journal of Applied Psychology, 87*, 765-780.
- Katz, I., & Glass, D.C. (1979). An ambivalence-amplification theory of behavior toward the stigmatized. In: W. Austin & S. Worchel (Eds.). *The social psychology of intergroup relations* (pp. 55-70). Monterey, CA: Brooks/Cole.

- Katz, I., Glass, D.C., & Cohen, S. (1973). Ambivalence, guilt, and the scapegoating of minority group victims. *Journal of Experimental Social Psychology, 9*, 423-436.
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 6*, 893-905.
- Katz, I., Hass, R. G., & Wackenhut J. (1986). Racial ambivalence, value duality, and behavior. In: J. F. Dovidio & S. L. Gaertner (Eds.) (pp. 35-59). *Prejudice, discrimination, and racism*. New York: Academic Press.
- Kellerman, B. (2004). *Bad leadership: What it is, how it happens, why it matters*. Boston: Harvard Business School Press.
- Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology, 31*, 457-501.
- Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999a) Big Five markers. *Personality and Individual Differences, 35*, 1127-1140.
- Larson, J. R. (1982). Cognitive mechanisms mediating the impact of implicit theories of leader behavior on leader behavior ratings. *Organizational Behavior and Human Performance, 29*, 129-140.
- Lord, R. G., Binning, J. F., Rush, M. C., & Thomas, J. C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational Behavior and Human Performance, 21*, 27-39.
- 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. *The Leadership Quarterly, 12*, 311-338.

- 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, 343-378.
- Lowe, K. B., & Gardner, W. L. (2000). Ten years of *The Leadership Quarterly*: Contributions and challenges for the future. *Leadership Quarterly*, 11, 459-514.
- Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the literature. *The Leadership Quarterly*, 7, 385-425.
- Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist*, 157-158.
- Nelson, T. E., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology*, 32, 13-38.
- Plato (1991). *The republic of Plato* (A. D. Bloom, Trans., 2nd ed.). New York: Basic Books. (Original work published circa 360BCE).
- Powell, G. N., Butterfield, D. A., Alves, J. C., & Bartol, K. M. (2004). Sex effects in evaluations of transformational and transactional leaders. *Academy of Management Best Paper Proceedings*, E1-E6.
- Powell, G. N., Butterfield, D. A., & Parent J. D. (2002). Gender and managerial stereotypes: Have the times changed? *Journal of Management*, 28, 177-193.
- Ritter, B. A., & Yoder J. D. (2004). Gender differences in leader emergence persist even for dominant women: An updated confirmation of role congruity theory. *Psychology of Women Quarterly*, 28, 187-193.
- Rush, M. C., Thomas, J. C., & Lord, R. G. (1977). Implicit leadership theory: A potential threat to the internal validity of leader behavior questionnaires. *Organizational Behavior and Human Performance*, 20, 93-110.

- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.
- Swim J., Borgida E., & Maruyama G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*, 409-429.
- Swim J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin, 22*, 507-519.
- Taynor, J., & Deaux, K. (1973). When women are more deserving than men: Equity, attribution, and perceived sex differences. *Journal of Personality and Social Psychology, 28*, 360-367.
- Taynor, J., & Deaux, K. (1975). Equity and perceived sex differences: Role behavior as defined by the task, the mode, and the actor. *Journal of Personality and Social Psychology, 32*, 381-390.
- US Senate (2004). Women in the United States congress: 1917-2004. CRS Report for Congress. <http://www.senate.gov/reference/resources/pdf/RL30261.pdf> [Retrieved March 8, 2005].
- Vroom, V. H. (1976). Leadership. In M. D. Dunnette (Ed.). *Handbook of industrial and organizational psychology* (pp. 1527-1551). Chicago: Rand McNally.
- Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin, 128*, 699-727.
- Yukl, G. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *Leadership Quarterly, 10*, 285-305.

¹ The interaction of leader sex*performance cue*participant sex had a significant effect on the three following measures: inspirational motivation, leader general effectiveness, and leader extraversion. The general pattern of results indicated that, in the presence of good performance signals, males favored the female leader rating her as being more inspirational/effective/extraverted than the male leader. Contrarily, and in the presence of bad performance signals, males favored the male leader, rating him as more inspirational/effective/extraverted than the female leader. Interestingly, female raters favored the male leader in good performance signal situations and the female leader in bad performance situations. These three measures were excluded in the relevant factorial ANOVA analyses.

² For the two leader scales that were not influenced by the three-way interaction, there were only significant main effects for the performance cue factor, that is, for (a) behavioral idealized influence, $F(1, 341) = 12.43$, $p < .001$, partial $\eta^2 = .04$, and (b) contingent rewards, $F(1, 341) = 10.82$, $p < .001$, partial $\eta^2 = .03$. Examination of the means for these two factors indicated that participants rated leaders higher on both factors when exposed to a good cue.

³ The only significant effect on satisfaction was a main effect for cue, $F(1, 340) = 6.17$, $p < .05$, partial $\eta^2 = .02$. An examination of these means indicated that participants rated leaders higher on subordinate satisfaction when exposed to a good cue.

⁴ http://womenshistory.about.com/od/morepoliticalfigures/a/maureen_reagan.htm. [Retrieved August 24, 2004].