

Gender and the Evaluation of Leaders: A Meta-Analysis

Alice H. Eagly and Mona G. Makhijani
Purdue University

Bruce G. Klonsky
State University of New York at Fredonia

This article reviews research on the evaluation of women and men who occupy leadership roles. In these experiments, the characteristics of leaders other than their sex were held constant, and the sex of the leader was varied. These experiments thus investigated whether people are biased against female leaders and managers. Although this research showed only a small overall tendency for subjects to evaluate female leaders less favorably than male leaders, this tendency was more pronounced under certain circumstances. Specifically, women in leadership positions were devalued relative to their male counterparts when leadership was carried out in stereotypically masculine styles, particularly when this style was autocratic or directive. In addition, the devaluation of women was greater when leaders occupied male-dominated roles and when the evaluators were men. These and other findings are interpreted from a perspective that emphasizes the influence of gender roles within organizational settings.

The sparse representation of women as leaders and managers, especially at the higher levels of organizations (Bergman, 1986; Nieva & Gutek, 1981; Powell, 1988, 1990), raises the question of why women have limited access to leadership roles. Although there may be many reasons why relatively few women attain positions of leadership, one of the most widely discussed causes is discrimination against women. To the extent that discrimination accounts for women's underrepresentation in the ranks of leaders and managers, women's credentials and performance are not fairly evaluated. The same leadership behaviors, when performed by a woman, may be viewed less favorably than they are when performed by a man. This issue has become critical in legal cases that focus on gender discrimination. For example, discrimination was contended in the well-known case of Ann Hopkins, who was denied partnership in the accounting firm of Price Waterhouse, despite her outstanding record by objective criteria (e.g., the monetary value of the accounts she obtained for the firm; Glaberson, 1988; McCarthy, 1988). According to the discrimination interpretation of the Hopkins case, her apparently assertive and forceful behavior in relation to her staff and colleagues was negatively evaluated, merely

because she is female. As Fiske, Bersoff, Borgida, Deaux, and Heilman (1991) argued, gender stereotypes may have caused her behavior to be interpreted differently than that of male colleagues. The very same behavior would have been viewed as acceptable and perhaps evaluated quite favorably had she been male.

The gender discrimination issue that such cases raise is extremely important: Are women evaluated less favorably than men when performing leadership and managerial behaviors, even though in some objective sense the women's and men's behaviors are equivalent? If people are biased to evaluate female leaders' efforts less favorably than those of their male counterparts, women who aspire to leadership roles would encounter very serious barriers to entering these roles and advancing to higher levels within organizations.

We evaluate this gender discrimination issue by integrating research that has addressed the issue of whether women are devalued in leadership roles. This review is confined to experiments whose designs held constant the characteristics of leaders other than their sex and varied the sex of the leaders. In this article, we first place our project in the context of other discussions of research on gender and leadership. We then describe the experimental paradigms within which evaluations of female and male leaders have been compared and provide a theoretical analysis that yields predictions for our meta-analysis. Finally, after presenting the method and results of our review, we discuss its findings and explore their implications for women's functioning in leadership roles.

An Analysis of Gender and the Evaluation of Leaders *Other Discussions of Research on Gender and Leadership*

Fortunately, the empirical literature addressing the issue of whether women are devalued in leadership roles is substantial. Although a number of social scientists have attempted to sum-

This research was supported by National Science Foundation Grants BNS-8605256 and BNS-8807495 to Alice H. Eagly as well as a Research Opportunity Award from the National Science Foundation that sponsored Bruce G. Klonsky's participation in the research. A table showing the effect sizes and study characteristics for each study included in the meta-analysis is available from Alice H. Eagly.

We thank Wendy Wood, Janet Swim, Rebecca Henry, and three anonymous reviewers for their comments on a draft of this article and Katherine Markee for help in conducting computer-based information searches.

Mona G. Makhijani is now at Leo Burnett U.S.A., Evanston, Illinois.

Correspondence concerning this article should be addressed to Alice H. Eagly, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907-1364.

marize this research, these reviews have not been based on a systematic sample of the available studies, and most have been somewhat equivocal in their conclusions. For example, Bartol and Martin (1986) suggested that any tendency to evaluate women unfavorably in leadership roles depends on a number of interacting variables (e.g., sex of evaluators). Several reviewers have suggested that leadership style may be a moderating variable (e.g., Betz & Fitzgerald, 1987; Friesen, 1983; Nieva & Gutek, 1980, 1981; Terborg, 1977). In general, most reviewers have been quite cautious and have emphasized that findings have been inconsistent across studies. Yet Powell and Butterfield (1982) favored the null hypothesis that "female leaders are not evaluated or perceived differently from male leaders when engaging in the same behavior" (p. 1172). In contrast, Van Fleet and Saurage (1984) maintained that "there is . . . considerable research showing that [leadership] performance by females is frequently subjectively evaluated less favorably than identical performance by males" (p. 20).

Given this considerable divergence of opinions about the findings of empirical research on the evaluation of male and female leaders, a thorough review of the relevant findings is worthwhile. Moreover, methodological advances allow contemporary reviewers to use meta-analytic techniques, which apply statistically justified methods to the task of integrating research findings (see Cooper, 1989; Hedges & Olkin, 1985; Rosenthal, 1984). Therefore, our review is designed to provide a systematic, quantitative integration of the available research on the evaluation of male and female leaders. Our review can be contrasted with prior efforts to summarize this research, which are quite vulnerable to error because of their relatively informal, qualitative methods and their incomplete and unsystematic sampling of the available studies.

The only prior meta-analysis containing studies on the evaluation of male and female leaders is a considerably more general review carried out by Swim, Borgida, Maruyama, and Myers (1989). The domain of this review was all studies in which the sex of target persons was varied and subjects evaluated "a behavior produced by the target person, such as a job performance or articles written by the target person" (p. 412). Although the sample of leadership studies that we located for our meta-analysis was suitable for Swim and colleagues' review, they included only 14 of the 56 documents that we included in our review.¹ This fact suggests that Swim and associates achieved only very partial coverage of research in which the behavior that subjects evaluated was leadership. Moreover, because of the very general focus of the analysis provided by Swim and her colleagues, they did not address the issues that have been considered important in discussions of gender and leadership—for example, whether women are devalued when manifesting some but not other leadership styles or in some but not other organizational contexts.

To place our meta-analysis of research on the evaluation of male and female leaders in the context of other research on gender and leadership, we note that investigators have examined several issues in addition to potentially prejudicial evaluations of female leaders. A large number of studies explored the leadership styles of women and men to determine whether they carry out leadership roles differently. Eagly and Johnson (1990) have meta-analyzed this domain of research. Their review produced a number of findings, most notably a tendency for

women to lead in a more democratic and participative style than men (see discussion in Predictions About Moderating Variables subsection). Other studies of gender and leadership examined the extent to which women or men emerge as leaders in initially leaderless groups. Eagly and Karau's (1991) quantitative review of this literature produced an overall tendency for men to emerge as leaders as well as a number of situational variables that moderated this tendency. Finally, another focus of research on gender and leadership is the relative effectiveness of men and women who occupy leadership roles in groups or organizations. Eagly and her collaborators are currently working on a quantitative review of this domain.

Research Paradigms for Examining the Evaluation of Leaders

In interpreting our meta-analysis, readers should keep in mind the research paradigms that investigators have used to study the evaluation of female and male leaders. The most important attribute of the design of these studies is that they are true experiments in the sense that characteristics of leaders other than their sex are held constant and the sex of the leader is a manipulated variable. With such a design, any differences in evaluations of women and men can be ascribed to subjects' biased perceptions, at least insofar as researchers have successfully equated female and male leadership behavior. Because of the requirement that leaders' sex be a manipulated variable, these studies do not examine subjects' evaluations of men and women who actually occupy leadership or managerial roles in natural settings. Because there would be no way to ensure that men's and women's behavior is equivalent in natural settings, differences in the evaluation of male and female leaders could be due to genuine differences in their behavior as well as to perceivers' gender bias.

Within this general paradigm of presenting respondents with equivalent male and female leaders, two types of studies have been popular, one using written vignettes and the other using confederates trained to lead in a particular style. Although the mode of presenting information about leadership is quite different in these two types of studies, we did not predict that findings would differ. Thus, the most common type, featuring written vignettes, is modeled after the seminal Rosen and Jerdee (1973) study, in which written descriptions of managerial behavior were presented to subjects who evaluated these managers' effectiveness. Each subject read one version of a vignette in which a female or male manager responded to a supervisory problem using one of four managerial styles in dealing with subordinates who were female, male, or of both sexes. Subjects then rated the manager on several evaluative scales. In another early study of this general type, Bartol and Butterfield (1976) had subjects read descriptions of (and subsequently evaluate) male and female managers whose characteristic leadership style took four different forms.

These studies in which subjects evaluated descriptions of

¹ A portion of this discrepancy in the number of documents is accounted for by Swim, Borgida, Maruyama, and Myer's (1989) exclusion of unpublished studies, including dissertations.

managerial behavior can be contrasted with experiments in which subjects evaluated male and female confederates who had been trained to lead in particular styles. In the earliest of these studies, Lee and Alvares (1977) trained confederates to supervise a simulated industrial task in three different leadership styles. The subjects carried out the task as the confederate's subordinates and subsequently evaluated him or her on a questionnaire. In another example of a study in which confederates served as leaders, Kushell and Newton (1986) trained female and male confederates to lead in two differing styles. While behaving in one of the two styles, each confederate led a group of subjects in a decision-making task. The subjects then evaluated the confederate's leadership.

Overall Prediction for the Evaluation of Male and Female Leaders

Our predictions about the evaluation of male and female leaders within these research paradigms stemmed from Eagly's (1987) gender-role theory as well as the findings produced by Eagly and Johnson's (1990) meta-analysis of sex differences in leadership style. Gender-role theory maintains that people develop expectations for their own and others' behavior based on their beliefs about the behavior that is appropriate for men and women. However, in an organizational setting, people develop expectations about the behavior that is appropriate for a leader or manager (Phillips & Lord, 1982), and these more specific expectations should be a more important determinant of people's reactions to one another than more diffuse, gender-based expectations. Nonetheless, gender roles may continue to have some importance in organizational settings, with the consequence that women are regarded not as generic managers, but as female managers and men are regarded as male managers. Indeed, research supports the idea that people's expectations about managerial behavior depend to some extent on managers' gender (e.g., Heilman, Block, Martell, & Simon, 1989; Russell, Rush, & Herd, 1988). The idea that leaders are perceived simultaneously in terms of their gender and their organizational role is consistent with the more general concept of *gender-role spillover*, which is "a carryover into the workplace of gender-based expectations for behavior" (Guttek & Morasch, 1982, p. 58; see also Nieva & Guttek, 1981).

Gender-role spillover would have different consequences for women than for men. Because people's expectations about the behaviors appropriate for leaders and managers match their expectations about men more closely than their expectations about women (Heilman et al., 1989; Schein, 1973), women are to some extent subjected to incompatible expectations from leadership roles and the female gender role, as numerous social scientists have maintained (e.g., Bass, 1981; Bayes & Newton, 1978; Kruse & Wintermantel, 1986; O'Leary, 1974; Ragins & Sundstrom, 1989). According to this argument, female leaders and managers face a dilemma: By fulfilling people's expectations concerning leadership, they violate conventions concerning appropriate female behavior. As a consequence, women in leadership roles may be devalued relative to their male counterparts—that is, perceived as behaving less competently and as having less ability and effectiveness as a leader. Thus, our overall prediction for the meta-analysis is that female leaders are

perceived somewhat less favorably than equivalent male leaders.

In addition to a tendency for women in leadership roles to be evaluated negatively, their behavior may be regarded as more extreme than that of their male counterparts—that is, as more dominant and controlling, and, in general, as embodying a higher level of prototypical leadership qualities. This perception of female leaders as more extreme than their male counterparts would be likely to occur to the extent that female leadership behaviors are quite discrepant from people's stereotypes about women and are therefore perceptually contrasted from these stereotypes (see Manis, Nelson, & Shedler, 1988). This possibility is also consistent with the variant of attribution theory known as *correspondent inference theory* (Jones & Davis, 1965; Jones & McGillis, 1976), if women's lower status in society is seen as a situational pressure that makes leadership behavior less likely. Correspondent inference theory maintains (and empirical studies have shown; e.g., Ajzen, 1971) that the less likely an act, given the actor's situation, the stronger are perceivers' inferences that the actor's underlying disposition corresponds to the actor's behavior. Perceivers might thus believe that leadership behaviors indicate a stronger underlying disposition to lead when they observe women in a leadership role, because women's lower status militates against leadership.

Predictions About Moderating Variables

Leadership style. In addition to our general prediction that female leaders are evaluated somewhat more negatively than their male counterparts, we derived a more detailed set of predictions that take leadership style into account. These predictions stem from a construct that might be dubbed *gender-role congruency*, which we defined as the extent to which leaders behave in a manner that is congruent with gender-role expectations. Specifically, women can adopt a leadership style that is relatively feminine and therefore congruent with their gender role, or they can adopt a leadership style that is relatively masculine and therefore incongruent with their gender role. The gender-role congruency of female leaders' behavior should influence the degree to which they experience role conflict and violate other people's expectations about their behavior. To the extent that women lead in a feminine style, they may largely escape the role conflict that they would otherwise experience in leadership and managerial roles and may not be subjected to the negative evaluations that they would otherwise receive. In contrast, to the extent that women lead in a masculine style, they may exacerbate their role conflict and increase the likelihood of receiving unfairly negative evaluations of their performance. Consistent with this reasoning, when reviewers of research on gender and leadership have acknowledged a specific set of positive findings, some have maintained that research has tended to support these predictions about gender-typed leadership styles (e.g., Betz & Fitzgerald, 1987; Friesen, 1983; Nieva & Guttek, 1980, 1981; Powell, 1990; Terborg, 1977).

Whether the gender-role congruency of men's leadership styles would have a similar impact on how they are evaluated is a more complex question. It might seem that a straightforward application of the congruency principle would suggest that men would elicit negative evaluations when leading in a femi-

nine style, just as women would elicit negative evaluations when leading in a masculine style. However, another important aspect of male leaders' situation needs to be taken into account—namely, that they do not face a basic role conflict parallel to the conflict that women face in their dual status as women and leaders. Whereas the details of female leaders' style may be scrutinized because of this role conflict and may ameliorate the conflict, men's styles may be less consequential for how they are evaluated because their leadership is not viewed as problematic. Given that leadership by men is ordinarily perceived as legitimate, the details of their performance are less likely to be questioned, given a generally satisfactory level of competence. Therefore, as a consequence of the consensual belief that men have a right to lead, they may enjoy greater latitude to carry out leadership in a variety of masculine or feminine styles.

Masculine and feminine leadership styles can be understood in terms of the content of people's stereotypes about women and men. Thus, factor-analytic studies of gender stereotypes (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Eagly & Steffen, 1984) have shown that the majority of people's beliefs about male and female behavior can be summarized in a general way in terms of differences on two dimensions, the *communal* and the *agentic* (Bakan, 1966). Women are expected to possess high levels of communal attributes, including being friendly, unselfish, concerned with others, and emotionally expressive. Men are expected to possess high levels of agentic qualities, including being independent, masterful, assertive, and instrumentally competent. When applied to leadership, these communal and agentic stereotypes suggest that female-stereotypic forms of leadership are interpersonally oriented and collaborative, whereas male-stereotypic forms of leadership are task-oriented and dominating (see Cann & Siegfried, 1990).

These gender-stereotypic aspects of leadership style mirror Bales's (1950) distinction between socioemotional leaders and task leaders as well as the distinction between the interpersonally oriented and task-oriented aspects of leadership that was emphasized in the Ohio State studies on leadership (e.g., Halpin, 1957; Halpin & Winer, 1957; Hemphill & Coons, 1957; Stogdill, 1963). In the Ohio State research, interpersonal orientation, labeled *consideration*, included leadership behaviors such as helping and doing favors for subordinates, looking out for their welfare, explaining procedures, and being friendly and available. Task orientation, labeled *initiation of structure*, included leadership behaviors such as having subordinates follow rules and procedures, maintaining high standards for performance, and making leader and subordinate roles explicit.

Although the gender-role congruency argument thus suggests that women may be evaluated relatively favorably when leading in an interpersonally oriented style and unfavorably when leading in a task-oriented style, the findings of Eagly and Johnson's (1990) meta-analysis suggest that a more narrowly defined aspect of leadership style may be especially relevant to prejudicial evaluations of female leaders. This stylistic dimension is the tendency to (a) behave *democratically* and allow subordinates to participate in decision making or (b) behave *autocratically* and discourage subordinates from participating in decision making (Lewin & Lippitt, 1938; Likert, 1961; Vroom & Yetton, 1973).² This democratic-autocratic dimension of style

relates to gender stereotypes because one component of the agentic or instrumental aspect of these stereotypes is that men are relatively dominant and controlling—in other words, more autocratic and directive than women are.

Our prediction that women would be more negatively evaluated than men when behaving autocratically or directive stems from Eagly and Johnson's (1990) meta-analytic finding that this aspect of leadership style produced the largest sex difference in leadership style. Basing their meta-analysis on studies that examined the most frequently researched types of leadership style, Eagly and Johnson found that in studies conducted in organizations, female and male leaders did not differ in interpersonally oriented style or in task-oriented style, but they did differ in the tendency to lead democratically or autocratically. Women tended to adopt a more democratic or participative style and a less autocratic or directive style than men did.³

Eagly and Johnson (1990) speculated that the sex difference in democratic-autocratic style among organizational managers might be due to any of several factors, including underlying differences in female and male personality and skills. Yet they emphasized that this sex difference might be due to attitudinal bias against female leaders. Thus, the skepticism that people commonly express about women's capabilities in managerial and leadership roles (see reviews by O'Leary, 1974; Riger & Galligan, 1980; Terborg, 1977) may be intensified when women in these roles attempt to take charge in an especially authoritative manner. Autocratic, directive leadership on the part of women would be especially disruptive to traditional patterns of deference between women and men. Proceeding in a participative and collaborative mode in accomplishing managerial tasks may enable many female managers to win acceptance from initially skeptical subordinates and thereby remove one barrier to effectiveness. Because male leaders are not ordinarily constrained by subordinates' and colleagues' negative attitudinal bias, they are freer to lead in an autocratic and nonparticipative

² Although Bass (1981) distinguished between (a) democratic versus autocratic leadership and (b) participative versus directive leadership, Eagly and Johnson (1990) treated these measures as a single class because they found this distinction difficult to maintain when categorizing leadership styles and measures. Researchers typically have regarded styles of this general type as a single, bipolar dimension, because democratic and autocratic styles presumably are incompatible. In contrast, interpersonal and task styles apparently are not incompatible, as suggested by the preference of most researchers for treating these styles as separate, relatively orthogonal dimensions.

³ However, both interpersonally oriented and task-oriented styles were somewhat gender-stereotypic in the two other classes of leadership studies that Eagly and Johnson (1990) investigated—namely, (a) *laboratory experiments* and (b) *assessment studies*, which were defined as studies that assessed the leadership styles of people not selected for occupancy of leadership roles. These findings are consistent with the predictions of gender-role theory (Eagly, 1987), which suggests that behavior is less gender-stereotypic in organizations because organizational roles tend to be more salient and have more of an impact than gender roles in these settings. In contrast, the tendency for women to lead more democratically and less autocratically than men was found in all three classes of leadership studies (laboratory experiments, assessment studies, and organizational studies).

manner, should they so desire. The findings of Eagly and Johnson's meta-analysis and their argument about attitudinal bias thus suggest that the tendency to evaluate female leaders more negatively than male leaders should be especially strong when women lead in an autocratic, directive style.

Sex distribution in leadership role. A related prediction concerns the impact of the distribution of women and men into leadership roles. Although leadership roles are regarded as generally more suitable for men than women, certain of these roles are female-dominated (e.g., nursing supervisor) and are presumably perceived as congenial for women. In contrast, other leadership roles (e.g., military officer) are extremely male-dominated and are presumably perceived as especially congenial for men. The gender-role congruency principle suggests that an overall tendency for women to be devalued relative to men in leadership roles would be accentuated to the extent that such roles are male-dominated (or on other bases are perceived as especially masculine) and lessened to the extent that such roles are female-dominated (or on other bases are perceived as especially feminine). Consistent with this prediction, narrative reviewers of research on evaluations of work performance have stated that people derogate women's performance more strongly in male-typed jobs than in female-typed jobs (Landy & Farr, 1980; Nieva & Gutek, 1980, 1981; Ragins & Sundstrom, 1989).

Any tendency for women to be devalued more strongly in male-dominated roles than in female-dominated roles may also be related to the greater salience of numerically rare (i.e., token) members of social categories as well as to their greater causal prominence as shown by observers' ascription of greater assertiveness, confidence, and strength of personality to tokens (e.g., Mullen, 1991; Taylor, Fiske, Etcoff, & Ruderman, 1978). In addition, some studies have suggested that numerically rare group members are perceived more stereotypically (Crocker & McGraw, 1984; Kanter, 1977; Taylor et al., 1978). Although the tendency to be perceived as stronger and more assertive and the tendency to be perceived gender-stereotypically are contradictory for women, should either one of these tendencies emerge, it could be problematic for female leaders. Thus, given some attitudinal bias against women's leadership (see Riger & Galligan, 1980; Terborg, 1977), increased salience and causal prominence could well be disadvantageous for female leaders, who might then be perceived as overly strong and assertive. In contrast, enhanced strength and assertiveness should be acceptable for male leaders, because their leadership is regarded as legitimate. Also, given the disparity we have noted between the stereotypes of women and manager, it would no doubt be disadvantageous for female managers to be perceived gender-stereotypically and therefore to be regarded, for example, as relatively dependent and unaggressive. As a consequence, numerical rarity in a role or occupation may well result in devaluation and general disadvantage for token women but not token men, a generalization that has received some empirical support (Johnson & Schulman, 1989; Ott, 1989; Yoder & Sinnett, 1985).

Sex of subjects and sex of subordinates. Additional predictions concern the impact of the sex of the subjects and the sex of the leader's subordinates on perceptions of male and female leaders. Because placing women in leadership positions upsets the traditional societal gender hierarchy, male subjects might, in a sense, have more to lose by approving female leadership

because their status vis-à-vis women would decline. Thus, male subjects may be more prone than female subjects to reject female leaders. Such a tendency is potentially consistent with Swim and co-workers' (1989) report that even though male subjects were in general no more likely than female subjects to derogate women's behavior, male subjects' findings were particularly inconsistent across studies. Perhaps men are especially likely to devalue women's behavior in the situation we are examining—namely, when women occupy leadership roles. The gender hierarchy argument that female leadership upsets traditional relations between the sexes also suggests that female leaders may be especially devalued when they direct male subordinates. Women's leadership of women, although often somewhat nontraditional, would not be as thorough a reversal of the traditional pattern as women's leadership of men.

Other predictions. The quality of women's leadership might affect the amount of bias against them. According to some reviewers (e.g., Riger & Galligan, 1980; Wallston & O'Leary, 1981), when the quality of women's leadership is ambiguous, or at least not known to be outstanding, it may be devalued relative to men's leadership. In contrast, when women's behavior is identified as outstanding, it may be evaluated without negative bias or even be overevaluated relative to men's behavior (see Abramson, Goldberg, Greenberg, & Abramson, 1977). However, other reviewers (e.g., Nieva & Gutek, 1980) have argued that successful or competent women would be especially devalued relative to equivalent men because successful performance is inconsistent with women's generally lower status and presumed lesser competence. Following this logic, the level of a leader's position within an organization (i.e., as line manager, middle manager, or higher manager) might also affect evaluations: Women who attain relatively high managerial positions may receive especially prejudicial evaluations because their high status is inconsistent with their female gender.

Finally, the effect of leaders' sex should become smaller as the amount of other information that subjects possess about the leaders increases. Such a prediction is consistent with more general evidence suggesting that stereotypic judgments are weakened in the presence of individuating information (Deaux & Lewis, 1984; Eagly, Ashmore, Makhijani, & Longo, 1991; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982).

Method

Sample of Studies

Computer-based information searches were first conducted using the keywords *leader* and *leadership* when paired with terms such as *gender*, *sex*, *sex differences*, *women*, and *female*. These keywords were searched using the following data bases: *Psychological Abstracts* (PsycINFO, 1966 to 1988), *Dissertation Abstracts International* (DISS, 1961 to 1988), Educational Resources Information Center (ERIC, 1966 to 1988), and a worldwide business and management data base (ABI/INFORM, 1971 to 1988). Subsequent searches in PsycINFO, DISS, and ABI/INFORM used keywords related to leadership (e.g., *manager*, *executive*, *supervisor*, *administrator*, *coach*, and *military officer*) and paired them with terms such as (a) *evaluation*, *perception*, and *rating*, and (b) *gender*, *sex*, *sex differences*, *women*, and *female*. The Social SciSearch data base (1973 to 1988) was also searched to locate articles

that cited Rosen and Jerdee (1973), the seminal study dealing with standardized presentation of male and female leaders. All searches ended with the latest information available in October or November 1988. Additionally, we searched through numerous review articles, chapters, and books as well as the reference lists of all located studies.

Criteria for including studies in the sample were that (a) leadership, management, or supervision was portrayed or enacted; (b) subjects rated or otherwise reacted to one or more leaders; (c) at least one of subjects' ratings or other reactions was evaluative; (d) one independent variable was the sex of the target person in a leadership position; (e) any characteristics of the leader other than sex were either held constant or varied but represented identically for both sexes; and (f) research participants were adolescents and adults from the United States or Canada who were not sampled from populations with clinically diagnosed disorders.

Studies were omitted from the sample if there was insufficient documentation of the stimulus persons' status as leaders, managers, or supervisors. For example, studies excluded on this basis identified target individuals as occupants of seating positions at the head of the table (e.g., Porter & Geis, 1981; Porter, Geis, & Jennings, 1983) or as job candidates whose resumes were presented to subjects (e.g., Dipboye, Arvey, & Terpstra, 1977).

Studies were also excluded from the sample if there was insufficient evidence that the male and female leaders were equivalent except for their sex. This outcome occurred when the past performance and style of the male and female leaders were not identical (e.g., April, 1976) and when role-playing studies failed to provide detailed standardized scripts for the actors playing the leaders (Stitt, Schmidt, Price, & Kipnis, 1983). Studies involving only portrayals of extremely deviant leadership were also eliminated—for example, a history of making inappropriate decisions and creating problems in the organization (Giacalone, 1988).

In summary, our selection criteria narrowed our sample to studies depicting male and female leadership in a standardized fashion. The resulting sample (see Appendix) consisted of 56 documents reporting 61 studies.

To enable us to test our hypotheses about potential moderating variables such as leadership style, we divided each study that manipulated such a variable. For such studies, subjects within each level of the potential moderating variable were thus separately represented in our data analysis. Although this partitioning of some studies created non-independence in our data, the strategy allowed us to take studies' theory-relevant interactions into account and to test our predictions about moderating variables.

This partitioning was undertaken when (a) different leadership styles were portrayed (25 studies; e.g., Bartol & Butterfield, 1976), (b) the masculinity–femininity of the leader's role or task was varied (3 studies; e.g., Knight & Saal, 1984), (c) different levels of the quality of the leader's performance were portrayed (16 studies; e.g., Brief & Wallace, 1976), or (d) the sex of the leader's subordinates was varied (10 studies; e.g., Pence, 1980).⁴ A study that manipulated one of these variables was partitioned only if its reported findings were sufficient to allow the computation of separate effect sizes within the levels of the variable or to determine the direction or significance of its findings within the levels. By this strategy, 15 studies yielded only 1 unit, 29 studies yielded 2, 4 yielded 3, 11 yielded 4, 1 yielded 6, and 1 yielded 12. Our original sample of 61 studies thus produced 147 units, of which 15 were intact studies and 132 were subdivided parts of studies. For simplicity of exposition, we refer to these units as *studies* in the remainder of the manuscript. For 33 of these 147 units, the report lacked enough statistical detail to allow an effect size to be computed (but did at least yield a report that the sex comparison was nonsignificant or that the comparison went in the male or female direction). For some analyses, the 147 units were further partitioned by sex of subject when the data

were reported separately for male and female subjects. When thus partitioned, 221 units were available for analysis, 76 of which lacked an effect size.

Variables Coded From Each Study

The following general information was coded from each report: (a) publication year, (b) publication form (journal article, book chapter, dissertation or master's thesis, or unpublished document), (c) proportion of male authors,⁵ (e) number of observations,⁶ and (f) number of judgments aggregated into each observation underlying the study-level effect sizes.⁷

The following aspects of the dependent variables were also coded: (a) type of dependent variable (competence [e.g., expertise, effort, productivity, and general evaluation], satisfaction with leader [e.g., likability, desire to work with leader, and group cohesiveness], or leadership style),⁸ (b) specific leadership style used as dependent variable (interpersonal orientation [e.g., consideration, maintenance, and socioemotional tendency], task orientation [e.g., initiation of structure and focus on task], potency [e.g., power, authority, and influence], or not a style measure),⁹ and (c) method of assessing dependent variable (subjects' rating of leader's characteristics or researchers' coding of subjects' behavior).

The following characteristics of the portrayal of leadership were coded: (a) leadership style portrayed (interpersonally oriented, task oriented, autocratic, democratic, other, or mixed or no information), (b) masculinity–femininity of leadership style portrayed (masculine, feminine, or mixed or no information),¹⁰ (c) quality of leader's performance (high, medium, low, or mixed or no information), (d) sex distri-

⁴ Among the 61 studies, 8 were partitioned on two bases (e.g., leadership style and quality of leader's performance; Butterfield & Powell, 1981).

⁵ The sex of the authors of the research reports was included because this variable has predicted effect sizes in several studies of sex differences in social behavior (e.g., Eagly & Carli, 1981; Eagly & Johnson, 1990; Wood, 1987).

⁶ The number of observations is the n taken into account by the statistical analysis for the effects of leader sex.

⁷ For example, each study-level observation might represent a subject's ratings of a leader on four items, for a total of four judgments. Study-level effect sizes were aggregated across all of the dependent variables (see Computation and Analysis of Effect Sizes subsection). To the extent that measures were based on multiple observations, they should yield more reliable estimates of evaluations of female and male leaders, in the manner that the number of items in a test relates to the reliability of the total test (e.g., Ghiselli, 1964).

⁸ Because we included only evaluative dependent variables, certain classes of variables encountered in the studies were omitted (e.g., attributional ratings).

⁹ The interpersonal-orientation category included all measures assessing communal tendencies, and the task-orientation category included all measures assessing agentic tendencies (see discussion of communal and agentic dimensions of gender stereotypes early in this article).

¹⁰ High autocratic, high task, and low interpersonal styles were considered masculine; high interpersonal, low autocratic, and low task styles were considered feminine. Combinations of styles (e.g., high task and low interpersonal) were similarly classified if both components fit the masculine or feminine class. Because authors provided clear labels for the styles that were portrayed (e.g., autocratic style and consideration style), classification of styles into masculine and feminine categories was easily accomplished.

bution in leadership role (mainly men, mainly women, balanced, or unclear),¹¹ (e) sex of subordinates (men, women, both, or unknown or unclear), (f) level of leadership portrayed (first or line, second or middle, third or higher, or ambiguous, mixed, or unknown), (g) stimulus modality (written, videotaped, live, or more than one), and (h) amount of individuating information presented along with sex information (considerable [e.g., face-to-face interaction longer than a few minutes] or moderate [e.g., short face-to-face interaction and informative written description]).

The following aspects of the group or organizational context of leadership were coded: (a) social context of leadership (small group with leader-subordinate interaction observed by subjects, small group with subjects serving as subordinates, organizational with leadership described, organizational with leader-subordinate interaction observed by subjects, organizational with subjects serving as subordinates, or organizational in more than one context), (b) type of organizational context (business or manufacturing; educational, excluding athletics; educational, athletics; other or more than one; or not organizational context [i.e., small group]), and (c) group size when context was a small group.

In addition, the following characteristics of the subjects were coded: (a) sex of subjects (men, women, or both), (b) nationality of subjects (American or Canadian), and (c) source of subject population (high school students, college undergraduates, business or management graduate students, other graduate students, managers, nonmanagerial or mixed employee sample, or other or mixed).

The following characteristics of the research designs were coded: (a) method of ensuring equivalence between male and female leadership (same written description, same script, or both methods), (b) multiple leaders for each sex (yes or no),¹² (c) within- versus between-subjects variation of leader sex (within, between, or other, mixed, or unclear), (d) independent variables other than leader's sex in design (number recorded), and (e) unusual features of statistical analysis (yes or no).

These variables were independently coded by Mona G. Makhijani and Bruce G. Klonsky, with a median agreement of 99% (estimated kappa = .94); the variable *multiple leaders for each sex* yielded the lowest agreement, 85% (kappa = .69). Disagreements were resolved by discussion.

Computation and Analysis of Effect Sizes

The effect size calculated is g , the difference between the means of the female and male groups, divided by the pooled standard deviation.¹³ The sign of the difference was positive when male leaders were rated more favorably than female leaders and negative when female leaders were rated more favorably. For the 510 effect sizes that were computed, the computation was based on (a) means and standard deviations or error terms for 483 effect sizes, (b) F and t for 15 effect sizes, and (c) level p values (e.g., $p < .05$) for 12.

Because an effect size was computed whenever authors had provided sufficient statistical detail for any dependent variable, the 510 effect sizes were derived from data on different individual rating scales, as well as on different composite measures that authors of the studies had computed by aggregating two or more ratings. To facilitate the analysis of these diverse measures, we aggregated them in various ways prior to analyzing them. When two or more measures assessed the same type of dependent variable (e.g., competence) or the same specific leadership style (e.g., interpersonal orientation), we computed a single effect size by combining the separately computed effect sizes. We then computed an effect size that aggregated all style measures, as well as an effect size that aggregated all measures that did not assess style. Finally, to provide only one effect size for each study to satisfy the independence assumption of meta-analytic statistics (Hedges & Olkin, 1985), we combined all separately computed effect sizes to create a

study-level effect size for each study (i.e., for each intact study or subdivided part of a study; see prior discussion of partitioning). All combined effect sizes were calculated using Rosenthal and Rubin's (1986) suggested formula and assuming that the correlation between measures was .42. This correlation was estimated by averaging the correlations between dependent variables either reported in the studies or derived from coefficient alphas reported for multiple-item dependent variables used in the meta-analysis.

To reduce computational error, these effect size calculations were performed independently by Mona G. Makhijani and Bruce G. Klonsky, who then resolved any discrepancies. The statistical significance and direction of the comparisons between male and female leaders were also recorded; this information provided the only record of the comparison for studies that provided insufficient information to calculate an effect size.

The g s were converted to d s by correcting them for bias (i.e., g 's overestimate of the population effect size, which occurs especially for small samples; see Hedges & Olkin, 1985). To obtain an overall estimate of the difference in the evaluation of male and female leaders reported in the available research, we then combined the relevant study outcomes by averaging the d s. All such means were computed with each effect size weighted by the reciprocal of its variance, a procedure that gives more weight to effect sizes that are more reliably estimated. To determine whether each set of d s shared a common effect size (i.e., was consistent across the studies), we calculated a homogeneity statistic Q , which has an approximate chi-square distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes (Hedges & Olkin, 1985).

In the absence of homogeneity, we accounted for variability in heterogeneous effect sizes by relating them to the attributes of the studies. To determine the relation between these study characteristics and the magnitude of the effect sizes, both categorical and continuous models were tested (Hedges & Olkin, 1985). Categorical models, which are analogous to analyses of variance (ANOVAs), may show that heterogeneous effect sizes are homogeneous within the subgroups established by dividing studies into classes based on study characteristics (e.g., female vs. male subordinates). The techniques for calculating categorical models provide a between-classes effect (analogous to a main effect in an ANOVA) and a test of the homogeneity of the effect sizes within each class. The between-classes effect is estimated by Q_b , which has an approximate chi-square distribution with $p - 1$ degrees of freedom, where p is the number of classes. The homogeneity of the effect sizes within each class is estimated by Q_{wt} , which has an approximate chi-

¹¹ Census data were typically used for estimating the sex distribution of leaders portrayed in organizational contexts (e.g., business executive was classified as *mainly men*). Actual sex distribution of leaders was used in small group contexts. Global descriptions of leadership roles in many studies (e.g., "manager") precluded more exact estimates of sex distributions.

¹² Studies having more than one leader for each sex incorporated an internal replication on the sex of leader variable by representing each level of this variable by more than one scenario or confederate enacting leadership. For example, with two leaders of each sex, both John and George portray male leadership and both Susan and Jean portray female leadership. Studies with multiple leaders for each sex should provide more generalizable conclusions about the impact of leaders' sex.

¹³ When leader sex was a between-subjects variable, this standard deviation was computed separately within the male and female leadership conditions and pooled. When leader sex was a within-subjects variable, the standard deviation was the standard deviation of the differences between the paired observations. These standard deviations were estimated, whenever possible, only from the portion of each study's data entering into the effect size.

square distribution with $m - 1$ degrees of freedom, where m is the number of effect sizes in the class. The tables reporting tests of categorical models also include the mean weighted effect size for each class, calculated with each effect size weighted by the reciprocal of its variance, and an indication of whether this mean differed significantly from the value of 0.00, which indicates exactly no difference in the evaluation of female and male leaders.

The continuous models are least squares simple linear and multiple regressions, calculated with each effect size weighted by the reciprocal of its variance. Each such model yields a test of the significance of each predictor as well as a test of model specification, which evaluates whether significant systematic variation remains unexplained in the regression model. The sum of squares error statistic, Q_E , which provides this test of model specification, has an approximate chi-square distribution with $k - p - 1$ degrees of freedom, where k is the number of effect sizes and p is the number of predictors (not including the intercept). If correctly specified models are not achieved when implementing continuous models (or homogeneity is not achieved within the classes when implementing categorical models), the results of these analyses cannot be interpreted as confidently as they would otherwise be.

As a supplementary analysis, we attained homogeneity by identifying outliers among the effect sizes and sequentially removing those that reduced the homogeneity statistic by the largest amount (see Hedges & Olkin, 1985). Using such a procedure, Hedges (1987) found for several meta-analyses on psychological topics that the removal of up to 20% of the outliers in a group of heterogeneous effect sizes usually resulted in a high degree of homogeneity. Studies yielding effect sizes identified as outliers can then be examined after the fact to determine whether they appear to differ methodologically from the other studies. In addition, inspection of the percentage of effect sizes removed to attain homogeneity allows one to determine whether the effect sizes are homogeneous aside from the presence of relatively few aberrant values. Under such circumstances, the mean attained after removal of such outliers may better represent the distribution of effect sizes than the mean based on all of the effect sizes.

Results

Characteristics of the Studies

Before considering the findings reported in research on the evaluation of leaders, we examined the characteristics of the studies (or subdivided study units) from which conclusions about this research will be drawn. Table 1 shows many of these study characteristics.

As shown by the central tendencies of the characteristics listed in Table 1, studies generally (a) were published relatively recently, (b) were published as journal articles, (c) had a majority of male authors, (d) involved a moderate number of observations, and (e) aggregated a small number of judgments into each observation. Examination of the dependent variables showed that the studies typically (a) used the leader's competence as a dependent variable, (b) did not include a specific leadership style as a dependent variable but used one or more of several styles when style was included, and (c) assessed these dependent variables by having subjects rate the leader's characteristics.

The portrayal of leadership varied in several respects in the studies in our sample. Although sometimes subjects received no information about the style of the leader, a variety of different styles were portrayed when this information was given. Although often the masculinity or femininity of the leader's style

could not be discerned, the majority of leader portrayals could be classified as either masculine or feminine. Typically no information was given about the quality of the leader's performance, but high quality performance was most common when this information was given. In addition, the studies generally (a) portrayed leadership roles occupied mainly by men, (b) gave no information about the sex of subordinates, (c) portrayed first-level or line leadership, (d) presented written descriptions of leaders, and (e) presented a moderate amount of individuating information.

Examining the group or organizational context of leadership in the research, our analysis showed that the studies generally employed an organizational context for leadership with the leader's interaction with subordinates described to the subjects. When the context was organizational, the type of organization was most often business or manufacturing; when the context was a group, the median group size was 5.

The studies typically used male and female, American, college undergraduate subjects. Examination of the characteristics of the research design showed that the designs (a) used the same written description to ensure equivalence between male and female leadership, (b) did not have multiple leaders for each sex, and (c) varied leader sex on a between-subjects basis.

Overall Summary of the Relative Evaluation of Female and Male Leaders

In presenting the findings of our meta-analysis, we first consider the overall difference in the evaluation of female and male leaders and then report a number of models showing that several characteristics of the studies moderated the small overall tendency for female leaders to be devalued. The overall analysis, which is the summary of the study-level effect sizes given in Table 2, allows one to determine whether female leaders were evaluated less favorably than male leaders on the basis of all of the reports.¹⁴ A difference prejudicial to women is shown by a positive mean effect size that differed significantly from the value of 0.00, which indicates exactly no difference (i.e., the confidence interval did not include 0.00). As expected, evaluations were less favorable for female leaders than for male leaders, as indicated by the weighted mean of the effect sizes. This weighted mean was small but significantly different from 0.00; however, the unweighted mean did not differ from 0.00.

As also shown in Table 2, the effect sizes were not homogeneous, and homogeneity was attained by removing 15% of the effect sizes, a fairly typical meta-analytic outcome. The removal of outliers produced a somewhat smaller weighted mean that did not differ significantly from 0.00.¹⁵

¹⁴ As explained in the Method section, each study-level effect size summarized all of the effect sizes comparing evaluations of male and female leaders for a given intact study or subdivided study unit.

¹⁵ Inspection of these outliers suggested that they did not differ in obvious ways from other studies in the sample, except for an unremarkable tendency for them to reflect the stronger predictors of the effect sizes (e.g., educational, athletic organizational context; autocratic leadership style portrayed; see next subsection for discussion of moderating variables).

Table 1
Summary of Study Characteristics

Variable and class	Value	Variable and class	Value
Median publication year	1980	Stimulus modality	88
Publication form	94	Written	6
Journal article	12	Videotaped	32
Book chapter	39	Live	21
Dissertation or master's thesis	2	More than one	
Unpublished document		Amount of individuating information	37
Mean proportion of male authors	.64	Considerable	110
Median no. observations	80	Moderate	
Median no. judgments aggregated into each observation	3	Social context of leadership	3
Type of dependent variable ^a	85	Small group with leader-subordinate interaction observed by subjects	
Competence	58	Small group with subjects serving as subordinates	28
Satisfaction with leader	74	Organizational with leadership described	93
Leadership style		Organizational with leader-subordinate interaction observed by subjects	16
Specific leadership style used as dependent variable ^a	31	Organizational with subjects serving as subordinates	4
Interpersonal orientation	23	Organizational in more than one context	3
Task orientation	20	Type of organizational context	101
Potency	143	Business or manufacturing	
Not a style measure		Educational, excluding athletics	5
Method of assessing dependent variable ^a	214	Educational, athletics	6
Rating of leader's characteristics	3	Other or more than one	4
Coding of subjects' behavior		Not organizational context (i.e., small group)	31
Leadership style portrayed	16	Median group size when context was group	5
Interpersonally oriented	16	Sex of subjects ^b	12
Task oriented	23	Men	4
Autocratic	23	Women	131
Democratic	24	Both	
Other or mixed	45	Nationality of subjects	141
No information		American	6
Masculinity-femininity of leadership style portrayed	44	Canadian	
Masculine	45	Source of subject population	5
Feminine	58	High school students	
Mixed or no information		College undergraduates	80
Quality of leader's performance	27	Business or management graduate students	12
High	7	Other graduate students	2
Medium	14	Managers	13
Low	99	Nonmanagerial or mixed employee sample	9
Mixed or no information		Other or mixed	26
Sex distribution in leadership role	98	Method of ensuring equivalence between male and female leadership	94
Mainly men	2	Same written description	
Mainly women	40	Same script	48
Balanced	7	Both methods	5
Unclear		Multiple leaders for each sex	55
Sex of subordinates	23	Yes	92
Men	17	No	
Women	30	Within- versus between-subjects variation of leader sex	
Both	77	Within	11
Unknown or unclear		Between	131
Level of leadership portrayed	80	Other, mixed, or unclear	5
First or line	46		
Second or middle	5		
Third or higher	16		
Ambiguous, mixed, or unknown			

Note. For categorical variables, numbers in tables represent frequencies of reports in each class.

^a The frequencies do not add to 147 because some studies included measures of more than one of our five types of dependent variable (i.e., competence, satisfaction with leader, interpersonal orientation, task orientation, and potency). ^b When studies were subdivided by sex of subjects, the frequencies were 49 for men, 41 for women, and 131 for both.

There is no completely satisfactory method to compute a mean effect size that takes into account the nonsignificant comparisons that could not be represented as effect sizes because of a lack of sufficient information. Nevertheless, one possible solution is to give these studies the value of 0.00 (indicating exactly no difference between female and male leaders).

When this step was taken, the mean unweighted effect size (see "All reports" in Table 2) became slightly smaller than the unweighted mean (before outlier removal) that omitted these 0.00 values.

Table 2 also reports that slightly more than half (.56) of the comparisons between male and female leaders favored men.

Table 2
Summary of Study-Level Effect Sizes

Criterion	Value
Known effect sizes	
<i>n</i>	114
Mean weighted $d(d_+)$ ^a	0.05
95% CI for d_+	0.02/0.08
Homogeneity (<i>Q</i>) of d s comprising d_+ ^b	551.41*
Mean unweighted d	0.07
95% CI for mean unweighted d	-0.02/0.16
Median effect size	0.02
Known effect sizes excluding outliers	
<i>n</i>	97
<i>n</i> removed outliers	17 (.15)
Mean weighted $d(d_+)$	0.01
95% CI for d_+	-0.02/0.05
Homogeneity (<i>Q</i>) of d s comprising d_+	123.15
All reports	
<i>n</i>	147
Mean unweighted d	0.06
95% CI for mean unweighted d	-0.02/0.14
Differences favoring male leaders ^c	75/133 (.56)

Note. Effect sizes are positive for more favorable evaluation of male than female leaders and negative for more favorable evaluation of female leaders. *n* = sample size; CI = confidence interval; d = effect size; d_+ = mean weighted effect size; *Q* = homogeneity of d s.

^a Effect sizes were weighted by the reciprocal of the variance. ^b Significance indicates rejection of the hypothesis of homogeneity. ^c Frequencies are the number of study-level differences favoring the male leaders divided by the total number of differences. The proportion appears in parentheses.

* $p < .001$.

This proportion did not differ significantly from .50, the proportion expected under the null hypothesis.

Impact of Moderating Variables on the Relative Evaluation of Female and Male Leaders

Classifications of dependent variables. The type of dependent variable produced a significant between-classes effect (see Table 3). Contrasts showed that the tendency to favor men over women was larger when the dependent variable was the leader's competence or subjects' satisfaction with the leader rather than perceptions of leadership style, $ps < .001$.¹⁶ Thus, the measures that were more purely evaluative yielded stronger evidence of the devaluation of women's leadership. Moreover, the model for the specific leadership style used as a dependent variable indicated that measures that did not assess leadership style produced a stronger bias in favor of male leaders than did those style measures that assessed interpersonal orientation, $p < .05$, or task orientation, $p < .01$. Women were perceived as significantly more task-oriented than men.

Portrayal of leadership. The relations between various study attributes and the evaluation of leaders are given in Table 4.¹⁷ As predicted, the model for the leadership style portrayed was significant. A priori contrasts indicated that the autocratic style

produced significantly more favorable evaluations of male leaders than of female leaders than did any of the other four styles or the portrayals that did not include style information, $ps < .001$. Furthermore, the model for masculinity-femininity of the leadership style portrayed was significant. A priori contrasts showed that masculine leadership styles produced a larger difference in favor of male leaders than did feminine styles or styles that could not be classified as either masculine or feminine, $ps < .01$.

As also shown in Table 4, the sex distribution in the leadership role produced a significant model. Consistent with our hypothesis, a priori contrasts showed that the tendency for men to be more favorably evaluated than women was more pronounced for roles occupied mainly by men than for roles occupied equally by men and women (i.e., balanced), $p < .001$, or for roles for which the sex distribution was unclear, $p < .001$. Because only two studies presented roles occupied mainly by women, our estimate of evaluations of leaders in female-dominated roles may be unreliable. Yet the studies portraying leadership in roles occupied mainly by men differed significantly from all other studies combined, $p < .001$. The model for the sex of the subordinates did not conform to our prediction that female leaders would be especially devalued when they direct male subordinates. Instead, devaluation of female leaders occurred with female subordinates and with subordinates whose sex was unknown. These evaluations differed from the devaluation of male leaders that was obtained with male subordinates, $ps < .001$.

Organizational context of leadership. As shown by the significant model for type of organizational context, the preference for male leaders was larger in the educational, athletic context than in each of the other three types of organizational contexts, $ps < .001$. All of these studies on athletics presented basketball coaches in the leadership role.¹⁸ In addition, business or manufacturing contexts produced a larger difference favoring male leaders than did other organizational contexts (i.e., nonbusiness and noneducational organizations or multiple contexts), which favored female leaders, $p < .05$.

Characteristics of subjects. The model for sex of subjects, which was calculated on the study-level effect sizes with units partitioned by sex of subject (see Method), proved to be significant. A priori contrasts indicated that studies using male subjects produced a larger preference for male leaders than studies using female subjects, $p < .05$, or studies using both male and female subjects, $p < .001$. The model for the source of the sub-

¹⁶ Contrasts were post hoc unless a priori is indicated. Contrasts that are not described were nonsignificant.

¹⁷ A few models with weak, yet significant, between-classes effects were omitted because (a) only one class with an extremely small sample size differed from other classes or (b) the comparisons relevant to our hypotheses were nonsignificant.

¹⁸ These six studies derived from four documents (Cottle, 1982; Parkhouse & Williams, 1986; Weinberg, Reveles, & Jackson, 1984; Williams & Parkhouse, 1988). Two studies used female players as subjects, and four studies used female and male players. In the studies with players of both sexes, male players rated the female coach considerably less favorably than the male coach, whereas female players differed less in their evaluation of male and female coaches.

Table 3
Categorical Models for Classification of Dependent Variables

Variable and class	Between-classes effect (Q_B)	n	Mean weighted effect size (d_{i+})	95% CI for d_{i+}		Homogeneity within each class (Q_{wi}) ^a
				Lower	Upper	
Type of dependent variable ^b	21.11*					
Competence		79	0.09	0.05	0.12	473.73*
Satisfaction with leader		57	0.10	0.05	0.16	313.72*
Leadership style		56	-0.03	-0.07	0.01	104.30*
Specific leadership style used as dependent variable ^c	18.05*					
Interpersonal orientation		31	-0.02	-0.07	0.03	72.63*
Task orientation		23	-0.09	-0.16	-0.01	32.32
Potency		20	0.02	-0.07	0.12	28.24
Not a style measure		101	0.07	0.04	0.10	522.36*

Note. Effect sizes are positive for more favorable evaluation of male than female leaders and negative for more favorable evaluation of female leaders. CI = confidence interval.

^a Significance indicates rejection of the hypothesis of homogeneity. ^b Calculated on effect sizes for competence and satisfaction with leader as well as effect sizes aggregated across all leadership style dependent variables. ^c Calculated on effect sizes for interpersonal orientation, task orientation, and potency as well as effect sizes aggregated across all dependent variables not assessing leadership style.

* $p < .001$.

ject population also was significant. Although studies with high school subjects produced a stronger tendency to favor male leaders than did studies with any other type of subjects, $ps < .01$ or smaller, this finding may well reflect the organizational context of the high school studies (namely, basketball teams; see prior model and Footnote 18). In addition, nonmanagerial employees produced a stronger tendency to favor male leaders than did (a) managers, $p < .01$, (b) business or management graduate students, $p < .05$, or (c) college undergraduates, $p < .01$.

Other characteristics. The model for the method of ensuring equivalence between male and female leadership also was significant: Studies presenting the same written description favored men more strongly than studies presenting the same script, $p < .05$. Finally, two continuous models were significant (see Table 5). Year of publication related positively to the preference for male leaders: The more favorable evaluation of men than women increased over time. The number of independent variables in the design (other than sex of the leaders) related negatively to the preference for male leaders: The more positive evaluation of men was stronger in studies with few other independent variables.

Multiple regression model. Although most of our theory-relevant predictors were relatively independent of one another, concerns about their possible dependence as well as our desire to determine how much variability the set of predictors accounted for led us to estimate various multiple regression models. For purposes of such analyses, categorical variables were dummy-coded. The relatively large number of categorical and continuous variables that produced significant one-way models (and the presence of three or more classes for some of the categorical variables) restricted our efforts to test multiple regression models because the number of potential predictors was quite large in relation to the number of effect sizes. Nonetheless, in Table 5 we present one of the several models we com-

puted. This model entered most of the theory-relevant predictors that produced significant univariate models, namely, leadership style portrayed, sex distribution in leadership role, sex of subordinates, type of organizational context, year of publication, and number of independent variables. The significant predictors in this model were leadership style portrayed, sex distribution in role, sex of subordinates, and type of organizational context. As reflected in the multiple R of .59, this model was moderately successful in accounting for variability in the magnitude of the effect sizes, although the test of model specification (Q_E , see Table 5) showed that it cannot be regarded as correctly specified.

Discussion

Our review suggests that people do evaluate female leaders slightly more negatively than equivalent male leaders. Although this overall trend was small, we found that this bias against women in leadership roles was considerably larger under specific circumstances. Indeed, a substantial proportion of the variability in the findings available in this research area can be accounted for by a few characteristics of the portrayal of leadership and its social and organizational context as well as by attributes of the respondents. We consider first the magnitude of the overall bias against women in leadership roles and then discuss the circumstances under which the bias becomes particularly serious.

Strength of the Tendency for Female Leaders to Be Less Favorably Evaluated Than Male Leaders

If we consider the entire research literature on the evaluation of leaders, the tendency for men to be more favorably evaluated than women was weak, as shown by the several measures of central tendency reported in Table 2. These values varied from

Table 4
Categorical Models on Study-Level Effect Sizes

Variable and class	Between-classes effect (Q_B)	n	Mean weighted effect size (d_{i+})	95% CI for d_{i+}		Homogeneity within each class (Q_{wi}) ^a
				Lower	Upper	
Leadership style portrayed	26.59***					
Interpersonally oriented		9	0.00	-0.10	0.11	25.78***
Task oriented		9	0.08	-0.01	0.18	13.68
Autocratic		18	0.30	0.19	0.41	62.44***
Democratic		16	-0.02	-0.12	0.08	24.71
Other or mixed		21	-0.03	-0.13	0.06	45.90**
No information		41	0.04	0.00	0.08	351.46***
Masculinity-femininity of leadership style portrayed	12.06**					
Masculine		32	0.15	0.09	0.22	91.93***
Feminine		31	0.00	-0.07	0.07	59.94**
Mixed or no information		51	0.03	-0.01	0.07	386.64***
Sex distribution in leadership role	25.45***					
Mainly men		68	0.09	0.06	0.13	433.94***
Mainly women		2	-0.02	-0.31	0.27	1.50
Balanced		37	-0.06	-0.13	0.01	84.85***
Unclear		7	-0.19	-0.34	-0.04	4.83
Sex of subordinates	27.12***					
Men		19	-0.16	-0.26	-0.07	26.72
Women		14	0.16	0.04	0.28	16.46
Both		26	-0.01	-0.10	0.08	71.89***
Unknown or unclear		55	0.08	0.04	0.12	410.21***
Type of organizational context	148.45***					
Business or manufacturing		70	0.04	0.01	0.08	187.95***
Educational, excluding athletics		5	-0.02	-0.16	0.12	9.12
Educational, athletics		6	1.03	0.87	1.20	11.78
Other or more than one		4	-0.15	-0.26	-0.04	10.16*
Not organizational (i.e., small group)		29	-0.04	-0.14	0.06	64.91***
Sex of subjects ^b	14.05***					
Men		44	0.15	0.09	0.21	325.00***
Women		41	0.04	-0.04	0.12	129.34***
Both		66	0.01	-0.03	0.05	163.37***
Source of subject population	178.71***					
High school students		5	1.19	1.01	1.37	113.96***
College undergraduates		63	-0.01	-0.05	0.04	154.69***
Business or management graduate students		8	-0.05	-0.20	0.11	5.87
Other graduate students		2	0.35	0.01	0.69	1.40
Managers		9	0.01	-0.05	0.07	24.61**
Nonmanagerial or mixed employee sample		8	0.43	0.25	0.62	24.54***
Other or mixed		19	0.01	-0.06	0.08	48.59***
Method of ensuring equivalence between male and female leadership	7.57*					
Same written description		66	0.07	0.04	0.11	418.99***
Same script		43	-0.04	-0.12	0.03	101.17***
Both methods		5	0.02	-0.17	0.22	23.65***

Note. Effect sizes are positive for more favorable evaluation of male than female leaders and negative for more favorable evaluation of female leaders. CI = confidence interval.

^a Significance indicates rejection of the hypothesis of homogeneity. ^b Calculated on study-level effect sizes that were, whenever possible, partitioned by sex of subject.

* $p < .05$. ** $p < .01$. *** $p < .001$.

a high of 0.07 for the unweighted mean to a low of 0.01 for the mean with outliers excluded. Although the weighted mean indicated a significant sex difference, clearly the overall trend favored men only very slightly. In fact, our overall estimates of central tendency are very close to those that Swim and her colleagues (1989) estimated for their more general sample of studies in which subjects evaluated male and female behavior, including leader behavior in some studies. Yet it is worth noting that the tendency for subjects to favor male leaders was stronger

on the more general evaluative measures (i.e., perception of leader's competence and satisfaction with leader; see Table 3).

Methodological features of the studies we reviewed should be taken into account in interpreting the magnitude of the effect sizes we obtained. Especially important in considering magnitude is the degree of experimental control, because such control reduces the standard deviations that serve as the denominator of the effect sizes. Relevant to this control issue is the fact that the great majority of the studies in our sample presented sub-

Table 5
Continuous Models on Study-Level Effect Sizes

Predictor or outcome	Simple linear regressions		Multiple regression	
	<i>b</i>	<i>b*</i>	<i>b</i>	<i>b*</i>
Continuous variables				
Year of publication	0.01**	0.13	0.00	0.00
No. independent variables	-0.06**	-0.12	-0.01	-0.01
Categorical variables				
Leadership style portrayed ^a			0.06***	0.23
Sex distribution in leadership role ^b			0.17***	0.21
Sex of subordinates ^c			0.13*	0.09
Type of organizational context ^d			0.35***	0.51
Additive constant			-1.47	
Multiple <i>R</i>			.59	
<i>SE</i> of estimate			1.83	
<i>Q_E</i> ^e			357.52***	

Note. Models are weighted least squares simple linear and multiple regressions calculated with weights equal to the reciprocal of the variance for each effect size. In the multiple regression model, the predictors were entered simultaneously. *b* = unstandardized regression coefficient; *b** = standardized regression coefficient. Effect sizes are positive for more favorable evaluation of male than female leaders and negative for more favorable evaluation of female leaders; *n* = 114.

^a 1 = autocratic style, 0 = other styles. ^b 1 = mainly men, 0 = other distributions. ^c 1 = female subordinates, 0 = other subordinates. ^d 1 = educational, athletics, 0 = other contexts. ^e Significance indicates model not correctly specified.

* *p* < .05. ** *p* < .01. *** *p* < .001.

jects with written descriptions of leaders' behavior. Because the only features of these portrayals that investigators changed to vary the leader's sex were generally the leader's name and the personal pronouns that referred to him or her in a vignette, most studies reflected a high degree of experimental control. Although control was no doubt less in the studies in which leadership was enacted by male and female confederates, even these studies were relatively controlled because of the training procedures investigators implemented to ensure equivalence of male and female leaders' behaviors. Therefore, the studies in this research literature have a level of experimental control comparable to many studies in other domains of experimental social and organizational psychology and offer considerably more control than studies in the other areas of leadership research for which we have performed meta-analyses (leadership style, Eagly & Johnson, 1990; emergent leadership, Eagly & Karau, 1991).

Given this relatively high degree of experimental control in studies on the evaluation of leaders, the mean effect sizes we obtained can be compared with those produced by meta-analyses on related topics in social and organizational psychology that feature relatively controlled research. In addition to the comparison with the Swim and associates (1989) meta-analyses that we already noted, one relevant comparison is with Olian, Schwab, and Haberfeld's (1988) review of experiments on the effects of job applicants' sex on hiring recommendations in an employment-seeking context. In these studies, which typically presented subjects with resumes to evaluate, the tendency for male applicants to fare better than female applicants yielded a mean effect size of 0.41. Another interesting comparison is with Eagly and colleagues' (1991) review of the physical attractiveness stereotype, which surveyed experimental studies as-

sessing the hypothesis that attractive people are evaluated more favorably than unattractive people (i.e., the "beauty-is-good" hypothesis). The mean weighted effect size for the physical attractiveness stereotype was 0.58. Thus, the tendency for experimental subjects to devalue women in leadership roles, which produced a mean weighted effect size of 0.07 in our meta-analysis, is very substantially weaker than the tendencies for subjects to discriminate against women in a hiring context and to devalue physically unattractive people. Moreover, Eagly's (1987) overview of various mean effect sizes produced by meta-analyses in social and personality psychology suggested that they ranged from roughly 0.00 to 1.20, and mean effect sizes for sex-of-subject differences ranged from 0.13 to 1.19. According to these comparisons, as well as Cohen's (1977) suggestion that effect sizes of 0.20 should be labeled *small*, this meta-analysis produced a very weak overall effect.

Confirming our judgment that the overall preference for male leaders was small, our analysis of the direction of the effects proved that direction was quite inconsistent: 56% of the reports favored men and 44% favored women. Yet the effect sizes, which centered around a slight devaluation of women, were relatively consistent in their magnitude, as shown by the fact that a homogeneous set of effect sizes was achieved by removing 15% of the values, a fairly typical meta-analytic finding (see Hedges, 1987).

Impact of Type of Dependent Variable on the Evaluations of Male and Female Leaders

In addition to general evaluative measures, the studies in our meta-analysis included more specific evaluative measures assessing perceptions of leadership style. These measures as-

sessed relatively circumscribed aspects of subjects' perceptions of good performance. For example, for ratings of interpersonal orientation, typical measures assessed perceptions of the friendliness and accessibility of leaders and their concern for subordinates' welfare. For ratings of task orientation, typical measures assessed perceptions of tendencies to have subordinates follow rules and procedures, maintain high standards, and make leader and subordinate roles explicit. For two of our three categories of these specific stylistic attributes (i.e., interpersonal orientation and potency; see Table 3), the sexes were not perceived to differ. However, for the remaining category, task orientation, our findings countered the general prediction that men would be evaluated more favorably: Women were perceived as significantly more task-oriented than men.

The tendency for women's leadership behaviors to be perceived as somewhat more task-oriented than men's equivalent behaviors may reflect a tendency to contrast women's leadership behavior from the female stereotype and thereby to view this behavior as more extreme (see discussion early in this article). Subjects thus exaggerated the extent to which these women "took charge" in an agentic and task-effective manner. If it is task-oriented behavior that people identify most closely with leadership (Stein & Heller, 1979), this slight exaggeration of women's task orientation would be consistent with the attributional principle that the less expected an act, given the constraints of the actor's role or situation, the stronger are perceivers' inferences that the actor's dispositions correspond to the actor's actions (Jones & Davis, 1965). Thus, because of the presumed constraints of female gender, prototypical leadership qualities may be ascribed more strongly to female occupants of these roles than to male occupants.

Impact of Other Study Attributes on Evaluations of Male and Female Leaders

Leadership style portrayed. Our findings provided support for the gender-role congruency hypothesis that women are negatively evaluated when they exhibit masculine leadership styles. When we classified leadership styles as masculine or feminine by gender-stereotypic criteria (see Footnote 10), our findings showed that women were devalued more (relative to men) when exhibiting masculine styles than they were when exhibiting other leadership styles, including feminine ones. Yet we presented a more specific prediction about the effects of leadership style portrayals based on Eagly and Johnson's (1990) finding that on the average, women lead in a more democratic and less autocratic style than men do and their suggestion that an attitudinal bias might underlie this finding. Such an attitudinal bias would take the form of disapproval directed specifically toward women who lead in an autocratic and directive manner, which would be particularly disruptive to traditional patterns of gender deference. Indeed, our findings substantiated this logic: The tendency to devalue female leaders was larger when leaders behaved in an autocratic manner than it was when leaders behaved in accord with any other style.

As we expected (see discussion early in this article), leading in a feminine manner did not create a disadvantage for men relative to women: Our data indicated that subjects evaluated women and men equivalently when they carried out leadership

in more stereotypically feminine styles (i.e., democratic and interpersonally oriented leadership). These findings are thus consistent with our logic that whereas feminine styles ameliorate female leaders' role conflict, they do not compromise male leaders' success. It appears that all other factors being equal, men may have greater freedom than women to lead in a range of styles without encountering negative reactions.

This asymmetry in reactions to women's and men's leadership styles resembles Ridgeway's (1982) findings on conformity and status attainment in small groups. Ridgeway found that friendly, cooperative, interpersonally oriented behavior enhanced women's status and their influence over other group members but had little or no impact on men's status and influence. According to Ridgeway, this stereotypically feminine behavior served to demonstrate women's group-oriented motivation and their lack of self-oriented motivation in a potential leadership situation. This proof of acceptable motivation was evidently a prerequisite for effective leadership by women. Men, in contrast, were not suspected of having self-aggrandizing motives and were not required to prove their group-oriented motivation, because group members perceived them as having an inherent right to lead. From Ridgeway's perspective, the ability of feminine leadership styles to ameliorate female leaders' potential role conflict may stem from the specific meaning these styles convey—namely, that a leader is concerned about the success of the group and not about enhancing personal power and status.

Sex distribution in leadership role and type of organizational context. Consistent with our gender-role congruency hypothesis, the tendency for men to be more favorably evaluated than women was greater for roles occupied mainly by men than for roles occupied equally by both sexes, or for roles where the sex distribution was unclear. These findings are consistent with the conclusions of earlier reviews highlighting the less favorable evaluations that women receive when they violate gender-role expectations by being interviewed for or working in a male-typed job (e.g., Landy & Farr, 1980; Nieva & Gutek, 1981; Ragins & Sundstrom, 1989; but see Olian et al., 1988).

To gain some insight into the specific type of male-dominated roles that produced this bias against women, it is helpful to examine in addition the model for type of organizational context of the studies. These findings indicated that the more favorable evaluation of men than women was more pronounced in the school and college athletic context than in any of the other organizational contexts that were examined. Because only six athletic studies were available, all of which examined evaluations of basketball coaches (e.g., Parkhouse & Williams, 1986; Weinberg, Reveles, & Jackson, 1984), the generalizability of the finding to all athletics is hazardous. Nonetheless, this devaluation seems consistent with the substantial decrease in the proportion of women in coaching positions during the period when the studies were conducted (Acosta & Carpenter, 1985; Hart, Hasbrook, & Mathes, 1986; Holmen & Parkhouse, 1981). Men now constitute a solid majority of coaches in women's interscholastic and intercollegiate sport. Given this decreasing proportion of women who coach women's and girls' athletic teams and society's ambivalence about women's participation in sports emphasizing power, strength, and speed, such as basketball (Coakley, 1982; King & Chi, 1979; LeUnes & Na-

tion, 1989; Ostrow, Jones, & Spiker, 1981), this tendency to derogate female coaches should not be surprising. Perhaps athletes' biases against female basketball coaches are due in part to stereotypes perpetuated by the lack of exposure to successful female coaches (Heilman & Martell, 1986; Williams & Parkhouse, 1988).

Another male bastion, business and manufacturing, generated a greater tendency to favor male leaders than did organizational contexts not involving business and manufacturing or education. Indeed, managerial roles in business and manufacturing were the male-dominated leadership positions used most frequently in the studies included in our review. Viewing the business management position as male-dominated, especially at the upper levels, still is accurate given continuing evidence of the underrepresentation of women in these roles (e.g., Morrison & Von Glinow, 1990). The slightly more favorable evaluation of male business managers than female business managers shown by our review parallels recent findings indicating that the stereotype of managers remains masculine rather than androgynous or feminine in business organizations (Heilman et al., 1989; Powell & Butterfield, 1989).

Sex of subjects and subordinates. Consistent with our reasoning that men, as the sex accorded higher societal status, have "more to lose" by approving of women in leadership roles, they showed a stronger tendency to devalue female leaders than women did. This finding is thus consistent with the general in-group bias that social psychologists have demonstrated in several domains (Brewer, 1979; Brewer & Kramer, 1985). It should be noted, however, that women merely showed no gender bias and did not favor female over male leaders.

Somewhat puzzling are the findings we obtained for the sex of leaders' subordinates. Contrary to our hypothesis that subjects would react especially negatively to women who were given authority over men, subjects favored female leaders over male leaders with male subordinates, but favored male leaders over female leaders with female subordinates. We investigated these unexpected trends separately within the levels of our sex-of-subject classification and found no evidence that they interacted with subjects' sex. Perhaps subjects viewed the pairings of women with men as potentially more interesting and provocative, perhaps for themselves as perceivers or for subordinates. Another possibility is that subjects engaged in relatively subtle attributional reasoning that ascribed special competence to women who are in charge of men because such women have to be competent enough to withstand the countervailing pressures from the traditional gender hierarchy. Yet a parallel attributional logic would suggest devaluing men who lead women because such men have the gender hierarchy operating in their favor. Contrary to this logic, we obtained the opposite trend when men had authority over women.

Quality of leader's performance, level of leadership, and amount of individuating information. No evidence was obtained for the hypothesis that the devaluation of female leaders is stronger when the quality of their performance is not known to be outstanding, or for the contrasting hypothesis that competent, successful women are evaluated especially harshly. Yet the presence of a number of studies in which performance quality was varied allowed a reasonably good test of the impact of this variable. Because Swim and associates (1989) also failed to find

an effect of this variable in their heterogeneous sample of studies examining the evaluation of male and female behavior, the idea that bias in evaluations depends on the quality of role performance should be regarded as empirically unsubstantiated. Yet quality might predict evaluations if interacting conditions were taken into account (see Heilman, Martell, & Simon, 1988). We also failed to obtain increased bias against women as the level of a leader's position became higher within the organization.

On the basis of our coding of the amount of individuating information subjects had available about the male or female leaders, no support was obtained for the hypothesis that the bias against women weakens with increases in the amount of individuating information that perceivers have available. We had reasoned that in an informationally impoverished situation, gender should be a more important cue for perceivers, and whatever stereotypic perceptions this cue might elicit should operate more strongly. As we indicated early in this article, this hypothesis has been supported in a variety of domains. Our failure to confirm the hypothesis about individuating information is probably due to the fact that the studies in our sample were in no case informationally impoverished to an extreme degree (see Table 1). Because the leader's behavior was described or enacted in all experiments, subjects possessed at least a moderate amount of individuating information. The exaggeration of stereotypic perceptions may occur only when perceivers possess very little information.¹⁹

Other findings. Consistent with the findings we have already presented for the organizational context of the studies, the class of subjects exhibiting the greatest tendency to favor male leaders was high school varsity basketball players who evaluated male and female coaches. Also, the method of ensuring equivalence between male and female leadership produced a weak, yet significant, model suggesting more bias against women with written vignettes than scripts enacted by confederates. Because these two types of experiments differed in several ways, we do not attempt to interpret this trend.

Finally, the tendency for the more favorable evaluation of male leaders than of female leaders to be larger in more recently published studies is provocative because it violates the common assumption that equality of opportunity for women has increased as prejudice and stereotyping have declined. How-

¹⁹ A tendency for individuating information to lessen bias against women was found in Tosi and Einbender's (1985) meta-analysis of studies in which subjects evaluated resumes of female and male job candidates. Tosi and Einbender's analysis was based on the number of cues available to subjects, a count that was derived in part from the number of independent variables other than sex that were varied in each experiment's design. Each such manipulation should make an additional cue available to subjects and thereby decrease the impact of sex. Although we also found that the number of independent variables in the design related negatively to the bias against women, this predictor was not significant in our multiple regression model (see Table 5). Moreover, we do not believe that the number of independent variables yields a good test of the hypothesis about individuating information in our meta-analysis because, as we just noted, subjects in the studies we reviewed generally possessed a substantial amount of information that was held constant across the experimental conditions.

ever, caution is warranted in interpreting this trend because publication year did not remain a significant predictor in the multiple regression model in which the variable was entered along with other predictors. This fact suggests that the relation between publication year and the devaluation of female leaders could have been an artifact of correlations of year with methodological features of the studies. To the extent that the secular trend might nonetheless be interpreted in terms of social change, we suggest that as more women enter managerial roles, they may be less likely to be regarded as extremely rare exceptions who have proven themselves outstanding by having overcome very serious discrimination. With a larger proportion of female leaders and managers, these women may instead be devalued according to the logic we presented early in this article—that is, they may be somewhat disadvantaged by the incompatible expectations they face from the leadership role and the female gender role.

Conclusion

Although our meta-analysis produced only a slight tendency for female leaders to be devalued relative to male leaders, our findings gave considerable evidence of selective devaluation. Women in leadership roles were devalued relative to their male counterparts when leadership or management was carried out in stereotypically masculine styles, particularly when this style was autocratic and nonparticipative. In addition, the devaluation of women was stronger when leaders occupied male-dominated roles and when the evaluators were men. These findings suggest that gender can influence evaluations of managers, even though this impact does not typically take the form of a general tendency to devalue women's managerial contributions. Rather, many of these findings are interpretable in terms of the gender-role congruency perspective that we discussed early in this article. Thus, gender roles appear to restrict the options of female managers in the sense that they "pay a price" in terms of relatively negative evaluation if they intrude on traditionally male domains by adopting male-stereotypic leadership styles or occupying male-dominated leadership positions. Of course, women may be able to compensate for this unfavorable evaluation by being more competent than their male counterparts, but it is in this sense that the adage about having to be "twice as good as a man" to obtain advantages such as promotion may have a degree of accuracy. Indeed, as the case of Ann Hopkins suggests, even a high level of success and competence may not protect women from the negative evaluations that follow from gender-incongruent behaviors.

To the extent that the findings of this review can be generalized to natural settings, they suggest that female managers may indeed be victims of unfair evaluations. As women enter male-dominated leadership roles, in organizations in which autocratic styles are common and evaluations of performance are typically conducted by men, the bias these women encounter may be decidedly nontrivial, given the evidence this review provides of selective devaluation in experimental studies.

In thinking about the potential generalizability of these findings to natural settings, readers should consider whether the demands of the experimental settings in which the studies were conducted may have restrained subjects from showing as much

bias as they would in organizational settings. To the extent that experimental settings have a normative environment relevant to evaluating leadership behavior, it would seem to be an environment that fosters fairness toward persons of differing sex (as well as race and ethnicity). Given that subjects know that their evaluative behavior is under scrutiny in these experiments, ordinarily by a research project sponsored by one or more professors and graduate students, subjects are no doubt reluctant to appear prejudiced. Indeed, a desire of experimental subjects to "look good" (e.g., emotionally adequate, moral, honest, and unprejudiced) has been suggested, especially in terms of Rosenberg's (1969) concept of *evaluation apprehension*. Of course, subjects' awareness that the research project pertains to gender discrimination would be lessened by the nearly universal use of between-subjects designs in this research literature (see Table 1). Because subjects did not evaluate a woman and an equivalent man, they had no chance to match their response across targets of different sex. Nonetheless, subjects receiving a description of a woman, especially in a male-dominated managerial role, might well have inferred that their fairness in relation to women was under scrutiny. In organizations, which in some cases may have a normative environment that tolerates or even encourages a certain skepticism about women in managerial roles, the tendency to evaluate women's managerial performance unfavorably may be more pronounced. It is thus not unreasonable to argue that experimental studies may produce an underestimate of perceivers' tendency to evaluate women's leadership behavior more negatively than the equivalent behavior carried out by men.

Another concern in generalizing our findings to organizational settings is that organizational evaluators often have more information available to them about managers' performance than did the subjects in the experiments that were included in our sample. The majority of these studies used a hypothetical leader paradigm that presented each subject with a fairly brief scenario that summarized a manager's behavior and performance. In the studies in which confederates played the role of a leader, subjects gleaned information about their leader from relatively brief face-to-face interaction. Although the information available to the subjects in these two paradigms was not extremely impoverished, it was fairly limited. These experimental settings thus differed from organizational settings, where evaluators presumably benefit from more extensive interaction with the employees they evaluate. Consistent with claims about the effect of individuating information (e.g., Eagly et al., 1991; Locksley et al., 1980), the more extensive information available to organizational evaluators could dampen the effect of employees' sex. Nonetheless, biased expectations are often powerful influences in social interaction (e.g., Snyder, 1984). A prejudiced evaluation based on limited information available in an early encounter in an organizational setting could set up unfavorable expectations that would prejudice later evaluations even though they are based on more information.

Finally, we note that this meta-analysis is not informative about the consequences of selective devaluation of female leaders. Although these consequences would seem to be negative for women, it is important for future research to determine how women in leadership and managerial roles react when they are evaluated less favorably than equivalent men. Women, wish-

ing to advance in organizations and be treated fairly, may tend to avoid the behaviors and situations that elicit prejudicial evaluations. Avoidance of an autocratic and directive style, which elicits devaluation of female leaders, is consistent with Eagly and Johnson's (1990) finding that women in leadership roles in fact adopt a more democratic and participative style than men in these roles. To the extent that women in addition avoid male-dominated leadership roles and situations in which men serve as evaluators, the selective devaluation phenomena that we have documented in this meta-analysis would serve to preserve the traditional division of labor and discourage women from seeking positions that offer higher income and status.

References

- Abramson, P. E., 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, 114-124.
- Acosta, R., & Carpenter, L. (1985). Women in athletics: A status report. *Journal of Physical Education, Recreation and Dance*, 56(7), 30-34.
- Ajzen, I. (1971). Attribution of dispositions to an actor: Effects of perceived decision freedom and behavioral utilities. *Journal of Personality and Social Psychology*, 18, 144-156.
- April, E. A. A. (1976). The use of judgment analysis in determining perceived differences in judging the competence of male and female administrators. (Doctoral dissertation, University of Colorado, 1975). *Dissertation Abstracts International*, 36, 4883A-4884A.
- Bakan, D. (1966). *The duality of human existence*. Chicago: Rand McNally.
- Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Reading, MA: Addison-Wesley.
- Bartol, K. M., & Butterfield, D. A. (1976). Sex effects in evaluating leaders. *Journal of Applied Psychology*, 61, 446-454.
- Bartol, K. M., & Martin, D. C. (1986). Women and men in task groups. In R. D. Ashmore & F. K. Del Boca (Eds.), *The social psychology of female-male relations* (pp. 259-310). San Diego, CA: Academic Press.
- Bass, B. M. (1981). *Stogdill's handbook of leadership: A survey of theory and research* (Rev. ed.). New York: Free Press.
- Bayes, M., & Newton, P. M. (1978). Women in authority: A sociopsychological analysis. *Journal of Applied Behavioral Science*, 14, 7-20.
- Bergman, B. R. (1986). *The economic emergence of women*. New York: Basic Books.
- Betz, N. E., & Fitzgerald, L. F. (1987). *The career psychology of women*. San Diego, CA: Academic Press.
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86, 307-324.
- Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual Review of Psychology*, 36, 219-243.
- Brief, A. P., & Wallace, M. J., Jr. (1976). The impact of employee sex and performance on the allocation of organizational rewards. *Journal of Psychology*, 92, 25-34.
- Browerman, I. K., Vogel, S. R., Browerman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues*, 28(2), 59-78.
- Butterfield, D. A., & Powell, G. N. (1981). Effect of group performance, leader sex, and rater sex on ratings of leader behavior. *Organizational Behavior and Human Performance*, 28, 129-141.
- Cann, A., & Siegfried, W. D. (1990). Gender stereotypes and dimensions of effective leader behavior. *Sex Roles*, 23, 413-419.
- Coakley, J. J. (1982). *Sport in society: Issues and controversies* (2nd ed.). St. Louis, MO: Mosby.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). San Diego, CA: Academic Press.
- Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.
- Cottle, S. A. (1982). Sex bias and professional status level effects in the evaluation of coaching ability (Master's thesis, California State University, Fullerton). *Masters Abstracts*, 20, 340.
- Crocker, J., & McGraw, K. M. (1984). What's good for the goose is not good for the gander: Solo status as an obstacle to occupational achievement for males and females. *American Behavioral Scientist*, 27, 357-369.
- Deaux, K., & Lewis, L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, 46, 991-1004.
- Dipboye, R. L., Arvey, R. D., & Terpstra, D. E. (1977). Sex and physical attractiveness of raters and applicants as determinants of resume evaluations. *Journal of Applied Psychology*, 62, 288-294.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.
- Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, 90, 1-20.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 108, 233-256.
- 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., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46, 735-754.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1991). Social science research on trial: The use of sex stereotyping research in *Price Waterhouse v. Hopkins*. *American Psychologist*, 46, 1049-1060.
- Friesen, L. (1983). Women and leadership. *Contemporary Education*, 54, 223-230.
- Ghiselli, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.
- Giacalone, R. A. (1988). The effect of administrative accounts and gender on the perception of leadership. *Group and Organization Studies*, 13, 195-207.
- Glaberson, W. (1988, October 2). Determined to be heard: Four Americans and their journeys to the Supreme Court. *New York Times Magazine*, pp. 32-40.
- Gutek, B. A., & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harassment of women at work. *Journal of Social Issues*, 38(4), 55-74.
- Halpin, A. W. (1957). *Manual for the Leader Behavior Description Questionnaire*. Columbus: Bureau of Business Research, Ohio State University.
- Halpin, A. W., & Winer, B. J. (1957). A factorial study of the leader behavior descriptions. In R. M. Stogdill & A. E. Coons (Eds.), *Leader behavior: Its description and measurement* (pp. 39-51). Columbus: Bureau of Business Research, Ohio State University.
- Hart, B., Hasbrook, C., & Mathes, S. (1986). An examination of the

- reduction in the number of female interscholastic coaches. *Research Quarterly for Exercise and Sport*, 57, 68–77.
- Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, 42, 443–455.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- 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. (1986). Exposure to successful women: Antidote to sex discrimination in applicant screening decisions? *Organizational Behavior and Human Decision Processes*, 37, 376–390.
- 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.
- Hemphill, J. K., & Coons, A. E. (1957). Development of the Leader Behavior Description Questionnaire. In R. M. Stogdill & A. E. Coons (Eds.), *Leader behavior: Its description and measurement* (pp. 6–38). Columbus: Bureau of Business Research, Ohio State University.
- Holmen, M., & Parkhouse, B. (1981). Trends in the selection of coaches for female athletes: A demographic inquiry. *Research Quarterly for Exercise and Sport*, 52, 9–18.
- Johnson, R. A., & Schulman, G. I. (1989). Gender-role composition and role entrapment in decision-making groups. *Gender & Society*, 3, 355–372.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). San Diego, CA: Academic Press.
- Jones, E. E., & McGillis, D. (1976). Correspondent inferences and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 389–420). Hillsdale, NJ: Erlbaum.
- Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82, 965–990.
- King, J. P., & Chi, P. S. (1979). Social structure, sex roles, and personality: Comparison of male/female athletes/nonathletes. In J. H. Goldstein (Ed.), *Sports, games, and play: Social and psychological viewpoints* (pp. 115–148). Hillsdale, NJ: Erlbaum.
- Knight, P. A., & Saal, F. E. (1984). Effects of gender differences and selection agent expertise on leader influence and performance evaluations. *Organizational Behavior and Human Performance*, 34, 225–243.
- Kruse, L., & Wintermantel, M. (1986). Leadership Ms.-qualified: I. The gender bias in everyday and scientific thinking. In C. F. Graumann & S. Moscovici (Eds.), *Changing conceptions of leadership* (pp. 171–197). New York: Springer-Verlag.
- Kushell, E., & Newton, R. (1986). Gender, leadership style, and subordinate satisfaction: An experiment. *Sex Roles*, 14, 203–209.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 70–107.
- Lee, D. M., & Alvares, K. M. (1977). Effects of sex on descriptions and evaluations of supervisory behavior in a simulated industrial setting. *Journal of Applied Psychology*, 62, 405–410.
- LeUnes, A. D., & Nation, J. R. (1989). *Sport psychology: An introduction*. Chicago: Nelson-Hall.
- Lewin, K., & Lippitt, R. (1938). An experimental approach to the study of autocracy and democracy: A preliminary note. *Sociometry*, 1, 292–300.
- Likert, R. (1961). *New patterns of management*. New York: McGraw-Hill.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39, 821–831.
- Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, 18, 23–42.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology*, 55, 28–36.
- McCarthy, M. J. (1988, June 14). Supreme Court to rule on sex-bias case: Is stereotyping of employees discriminatory? *Wall Street Journal*, p. 37.
- Morrison, A. M., & Von Glinow, M. A. (1990). Women and minorities in management. *American Psychologist*, 45, 200–208.
- Mullen, B. (1991). The phenomenology of being in a group: Integrations of social cognition and group processes. *Journal of Experimental Social Psychology*, 27, 297–323.
- Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. *Academy of Management Review*, 5, 267–276.
- Nieva, V. F., & Gutek, B. A. (1981). *Women and work: A psychological perspective*. New York: Praeger.
- O'Leary, V. E. (1974). Some attitudinal barriers to occupational aspirations in women. *Psychological Bulletin*, 81, 809–826.
- Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes*, 41, 180–195.
- Ostrow, A. G., Jones, D. C., & Spiker, D. D. (1981). Age role expectations and sex role expectations for selected sport activities. *Research Quarterly for Exercise and Sport*, 52, 216–227.
- Ott, E. M. (1989). Effects of the male-female ratio at work: Police-women and male nurses. *Psychology of Women Quarterly*, 13, 41–57.
- Parkhouse, B. L., & Williams, J. M. (1986). Differential effects of sex and status on evaluation of coaching ability. *Research Quarterly for Exercise and Sport*, 57, 53–59.
- Pence, E. C. (1980). The perception-attribution process in leadership (Doctoral dissertation, Virginia Polytechnic Institute and State University). *Dissertation Abstracts International*, 41, 2371B.
- Phillips, J. S., & Lord, R. G. (1982). Schematic information processing and perceptions of leadership in problem-solving groups. *Journal of Applied Psychology*, 67, 486–492.
- Porter, N., & Geis, F. L. (1981). Women and nonverbal leadership cues: When seeing is not believing. In C. Mayo & N. M. Henley (Eds.), *Gender and nonverbal behavior* (pp. 39–61). New York: Springer-Verlag.
- Porter, N., Geis, F. L., & Jennings, J. (1983). Are women invisible as leaders? *Sex Roles*, 9, 1035–1049.
- Powell, G. N. (1988). *Women & men in management*. Newbury Park, CA: Sage.
- Powell, G. N. (1990). One more time: Do female and male managers differ? *Academy of Management Executive*, 4, 68–75.
- Powell, G. N., & Butterfield, D. A. (1982). Sex, attributions, and leadership: A brief review. *Psychological Reports*, 51, 1171–1174.
- Powell, G. N., & Butterfield, D. A. (1989). The "good manager": Did androgyny fare better in the 1980's? *Group and Organization Studies*, 14, 216–233.
- Ragins, B. R., & Sundstrom, E. (1989). Gender and power in organizations: A longitudinal perspective. *Psychological Bulletin*, 105, 51–88.
- Ridgeway, C. L. (1982). Status in groups: The importance of motivation. *American Sociological Review*, 47, 76–88.
- Riger, S., & Galligan, P. (1980). Women in management: An exploration of competing paradigms. *American Psychologist*, 35, 902–910.

- Rosen, B., & Jerdee, T. H. (1973). The influence of sex-role stereotypes on evaluations of male and female supervisor behavior. *Journal of Applied Psychology*, 57, 44–48.
- Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). San Diego, CA: Academic Press.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400–406.
- Russell, J. E. A., Rush, M. C., & Herd, A. M. (1988). An exploration of women's expectations of effective male and female leadership. *Sex Roles*, 18, 279–287.
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, 57, 95–100.
- Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 247–305). San Diego, CA: Academic Press.
- Stein, R. T., & Heller, T. (1979). An empirical analysis of the correlations between leadership status and participation rates reported in the literature. *Journal of Personality and Social Psychology*, 37, 1993–2002.
- Stitt, C., Schmidt, S., Price, K., & Kipnis, D. (1983). Sex of leader, leader behavior, and subordinate satisfaction. *Sex Roles*, 9, 31–42.
- Stogdill, R. M. (1963). *Manual for the Leader Behavior Description Questionnaire—Form XII*. Columbus: Bureau of Business Research, Ohio State University.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105, 409–429.
- Taylor, S. E., Fiske, S. T., Etoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36, 778–793.
- Terborg, J. R. (1977). Women in management: A research review. *Journal of Applied Psychology*, 62, 647–664.
- Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *Academy of Management Journal*, 28, 712–723.
- Van Fleet, D. D., & Saurage, J. G. (1984). Recent research on women in management. *Akron Business and Economic Review*, 15, 15–24.
- Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision-making*. Pittsburgh, PA: University of Pittsburgh Press.
- Wallston, B. S., & O'Leary, V. E. (1981). Sex makes a difference: Differential perceptions of women and men. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 9–41). Beverly Hills, CA: Sage.
- Weinberg, R., Reveles, M., & Jackson, A. (1984). Attitudes of male and female athletes toward male and female coaches. *Journal of Sport Psychology*, 6, 448–453.
- Williams, J. M., & Parkhouse, B. L. (1988). Social learning theory as a foundation for examining sex bias in evaluation of coaches. *Journal of Sport and Exercise Psychology*, 10, 322–333.
- Wood, W. (1987). Meta-analytic review of sex differences in group performance. *Psychological Bulletin*, 102, 53–71.
- Yoder, J. D., & Sinnett, L. M. (1985). Is it all in the numbers? A case study of tokenism. *Psychology of Women Quarterly*, 9, 413–418.

Appendix

Studies Included in the Meta-Analysis

- Anderson, L. R., Finn, M., & Leider, S. (1981). Leadership style and leader title. *Psychology of Women Quarterly*, 5, 661–669.
- Andruskiw, O. (1978). *Women and men higher education administrators: Sex role attitudes and evaluations* (Report No. HE-011900). Troy, NY: Russell Sage College. (ERIC Document Reproduction Service No. ED 177 996).
- Bartol, K. M., & Butterfield, D. A. (1976). Sex effects in evaluating leaders. *Journal of Applied Psychology*, 61, 446–454.
- Brief, A. P., & Wallace, M. J., Jr. (1976). The impact of employee sex and performance on the allocation of organizational rewards. *Journal of Psychology*, 92, 25–34.
- Brown, V., & Geis, F. L. (1984). Turning lead into gold: Evaluations of men and women leaders and the alchemy of social consensus. *Journal of Personality and Social Psychology*, 46, 811–824.
- Buchheister, M. S. (1983). The effects of the sex of the supervisor and the supervisor's leadership style on subjects' cooperative behavioral responses and leader behavior descriptions (Doctoral dissertation, Ball State University, 1982). *Dissertation Abstracts International*, 43, 3397B.
- Butler, D., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: Implications for leadership evaluations. *Journal of Personality and Social Psychology*, 58, 48–59.
- Butterfield, D. A., & Bartol, K. M. (1977). Evaluators of leader behavior: A missing element in leadership theory. In J. G. Hunt & L. L. Larson (Eds.), *Leadership: The cutting edge* (pp. 67–188). Carbonale: Southern Illinois University Press.
- Butterfield, D. A., & Powell, G. N. (1981). Effect of group performance, leader sex, and rater sex on ratings of leader behavior. *Organizational Behavior and Human Performance*, 28, 129–141.
- Cage, J. H. (1983). Subordinate perception and evaluation of leaders who differ on a personal characteristic: Are leaders interchangeable? *Dissertation Abstracts International*, 43, 1657B. (University Microfilms No. 82-22356).
- Cohen, S. L., Bunker, K. A., Burton, A. L., & McManus, P. D. (1978). Reactions of male subordinates to the sex-role congruency of immediate supervision. *Sex Roles*, 4, 297–311.
- Cottle, S. A. (1982). Sex bias and professional status level effects in the evaluation of coaching ability (Master's thesis, California State University, Fullerton). *Masters Abstracts*, 20, 340.
- Dansby, M. R. (1979). Leader effectiveness as a function of leader's sex and leadership style (Doctoral dissertation, University of Florida). *Dissertation Abstracts International*, 40, 1951B.
- Decker, W. H. (1987). Attributions based on managers' self-presentation, sex, and weight. *Psychological Reports*, 61, 175–181.
- Dobbins, G. H., Stuart, C., Pence, E. C., & Sgro, J. A. (1985). Cognitive mechanisms mediating the biasing effects of leader sex on ratings of leader behavior. *Sex Roles*, 12, 549–560.
- Faranda, J. A. (1981). The influence of sex-role stereotypes on the evaluations of male and female leaders (Doctoral dissertation, University of Delaware). *Dissertation Abstracts International*, 42, 2128B.
- Frank, F. D., & Drucker, J. (1977). The influence of evaluatee's sex on evaluations of a response on a managerial selection instrument. *Sex Roles*, 3, 59–64.

- Frasher, J. M., & Frasher, R. S. (1980). Sex bias in the evaluation of administrators. *Journal of Educational Administration, 18*, 245-253.
- Futrell, C. M. (1984). Salespeople's perceptions of sex differences in sales managers. *Journal of Personal Selling & Sales Management, 4*, 19-23.
- Garland, H. (1977). Sometimes nothing succeeds like success: Reactions to success and failure in sex-linked occupations. *Psychology of Women Quarterly, 2*, 50-61.
- Giacalone, R. A., & Riordan, C. A. (1990). Effect of self-presentation on perceptions and recognition in an organization. *Journal of Psychology, 124*, 25-38.
- Haccoun, D. M., Haccoun, R. R., & Sallay, G. (1978). Sex differences in the appropriateness of supervisory styles: A nonmanagement view. *Journal of Applied Psychology, 63*, 124-127.
- Hall, F. S., & Hall, D. T. (1976). Effects of job incumbents' race and sex on evaluations of managerial performance. *Academy of Management Journal, 19*, 476-481.
- Heilman, M. E., & Guzzo, R. A. (1978). The perceived cause of work success as a mediator of sex discrimination in organizations. *Organizational Behavior and Human Performance, 21*, 346-357.
- Heilman, M. E., & Stopeck, M. H. (1985). Attractiveness and corporate success: Different causal attributions for males and females. *Journal of Applied Psychology, 70*, 379-388.
- Hicks, R. B. (1982). The effects of management styles assimilating three degrees of client-centered counseling on the description and evaluation of male versus female managers (Doctoral dissertation, Kansas State University, 1981). *Dissertation Abstracts International, 42*, 3073A.
- Jabes, J. (1980). Causal attributions and sex-role stereotypes in the perceptions of women managers. *Canadian Journal of Behavioural Science, 12*, 52-63.
- Jacobson, M. B., Antonelli, J., Winning, P. U., & Opeil, D. (1977). Women as authority figures: The use and nonuse of authority. *Sex Roles, 3*, 365-375.
- Karol, B. L., & Klimoski, R. J. (1976). Worker performance and evaluation of a supervisor as a function of subordinate sex, supervisor sex, and leadership style. *Academy of Management Proceedings, 36*, 99-103.
- Knight, P. A., & Saal, F. E. (1984). Effects of gender differences and selection agent expertise on leader influence and performance evaluations. *Organizational Behavior and Human Performance, 34*, 225-243.
- Kushell, E., & Newton, R. (1986). Gender, leadership style, and subordinate satisfaction: An experiment. *Sex Roles, 14*, 203-209.
- Lee, D. M., & Alvares, K. M. (1977). Effects of sex on descriptions and evaluations of supervisory behavior in a simulated industrial setting. *Journal of Applied Psychology, 62*, 405-410.
- Linsenmeier, J. A. W., & Wortman, C. B. (1979). Attitudes toward workers and toward their work: More evidence that sex makes a difference. *Journal of Applied Social Psychology, 4*, 326-334.
- McGillivray, C. W. (1977). *The effects of styles of leadership and sex of leader upon three areas of satisfaction of male and female subordinates*. Unpublished master's thesis, California State University, Fullerton.
- Norton, S. D., Gustafson, D. P., & Foster, C. E. (1977). Assessment for management potential: Scale design and development, training effects and rater/ratee sex effects. *Academy of Management Journal, 20*, 117-131.
- Offermann, L. R. (1982). Visibility and perception of male and female leaders as a function of follower sex, level of performance, and rater sex (Doctoral dissertation, Syracuse University, 1981). *Dissertation Abstracts International, 42*, 4245B.
- Parkhouse, B. L., & Williams, J. M. (1986). Differential effects of sex and status on evaluation of coaching ability. *Research Quarterly for Exercise and Sport, 57*, 53-59.
- Patterson, R. A. (1975). An experimental study of the effects of sex and marital status on job performance ratings, promotability ratings, and promotion decisions (Doctoral dissertation, University of Minnesota). *Dissertation Abstracts International, 36*, 3108B-3109B.
- Pence, E. C. (1980). The perception-attribution process in leadership (Doctoral dissertation, Virginia Polytechnic Institute and State University). *Dissertation Abstracts International, 41*, 2371B.
- Powell, G. N., & Butterfield, D. A. (1984). The "high-high" leader rides again! *Group and Organization Studies, 9*, 437-450.
- Remland, M., Jacobson, C., & Jones, T. (1983). Effects of psychological gender and sex-incongruent behavior on evaluations of leadership. *Perceptual and Motor Skills, 57*, 783-789.
- Rodriguez, D. A. (1988). The effects of situational factors and leader sex on the perceived appropriateness of leader behaviors (Doctoral dissertation, New York University, 1987). *Dissertation Abstracts International, 48*, 3441B.
- Rose, G. L. (1978). Sex effects on effort attributions in managerial performance evaluation. *Organizational Behavior and Human Performance, 21*, 367-378.
- Rosen, B., & Jerdee, T. H. (1973). The influence of sex-role stereotypes on evaluations of male and female supervisor behavior. *Journal of Applied Psychology, 57*, 44-48.
- Sheppeck, M. A. (1978). The effects of sex-role stereotypes, situation specific factors, and psychological processes upon reactions to male and female managers (Doctoral dissertation, University of South Florida, 1977). *Dissertation Abstracts International, 38*, 3451B.
- Smith, T. O. (1984). Leader evaluation as a function of gender, performance outcome, method of leader selection, and task type (Doctoral dissertation, Northern Illinois University). *Dissertation Abstracts International, 45*, 1325B.
- Snodgrass, S. E., Krasner, S., & Rosenthal, R. (1988). *Is the executive woman an oxymoron? Tone of voice and the evaluation of executive competence*. Unpublished manuscript, Skidmore College, Saratoga Springs, NY.
- Spencer, B. A., & Taylor, G. S. (1988). Effects of facial attractiveness and gender on causal attributions of managerial performance. *Sex Roles, 19*, 273-285.
- Trempe, J., Rigny, A., & Haccoun, R. R. (1985). Subordinate satisfaction with male and female managers: Role of perceived supervisory influence. *Journal of Applied Psychology, 70*, 44-47.
- Weinberg, R., Reveles, M., & Jackson, A. (1984). Attitudes of male and female athletes toward male and female coaches. *Journal of Sport Psychology, 6*, 448-453.
- Welsh, M. C. (1979). Attitudinal measures and evaluation of males and females in leadership roles. *Psychological Reports, 45*, 19-22.
- Wenze, L. (1983). Conceptualizations of sex-typed roles of effective and ineffective managers (Doctoral dissertation, Hofstra University). *Dissertation Abstracts International, 44*, 2004B.
- Wiley, M. G., & Eskilson, A. (1982). The interaction of sex and power base on perceptions of managerial effectiveness. *Academy of Management Journal, 25*, 671-677.
- Williams, J. M., & Parkhouse, B. L. (1988). Social learning theory as a foundation for examining sex bias in evaluation of coaches. *Journal of Sport and Exercise Psychology, 10*, 322-333.
- Wilson, M., Crocker, J., Brown, C. E., Johnson, D., Liotta, R., & Konat, J. (1985). The attractive executive: Effects of sex of business associates on attributions of competence and social skills. *Basic and Applied Social Psychology, 6*, 13-23.
- Wright, F. (1976). The effects of style and sex of consultants and sex of members in self-study groups. *Small Group Behavior, 7*, 433-456.

Received October 16, 1990

Revision received April 22, 1991

Accepted April 22, 1991 ■