

No Credit Where Credit Is Due: Attributional Rationalization of Women's Success in Male–Female Teams

Madeline E. Heilman and Michelle C. Haynes
New York University

In 3 experimental studies, the authors explored how ambiguity about the source of a successful joint performance outcome promotes attributional rationalization, negatively affecting evaluations of women. Participants read descriptions of a mixed-sex dyad's work and were asked to evaluate its male and female members. Results indicated that unless the ambiguity about individual contribution to the dyad's successful joint outcome was constrained by providing feedback about individual team member performance (Study 1) or by the way in which the task was said to have been structured (Study 2) or unless the negative expectations about women's performance were challenged by clear evidence of prior work competence (Study 3), female members were devalued as compared with their male counterparts—they were rated as being less competent, less influential, and less likely to have played a leadership role in work on the task. Implications of these results, both theoretical and practical, are discussed.

Keywords: gender stereotypes, sex stereotypes, gender bias, teams, sex discrimination

Despite the dramatic increase in the number of women in the workforce, women still are decidedly underrepresented in roles traditionally considered to be male—roles that often are the highest in authority, responsibility, and prestige in organizations (Lyness, 2002; Powell, 1999). In considering why there are so few women occupying these roles, researchers frequently designate gender stereotypes as the culprit. Indeed, gender stereotypes have repeatedly been said to play a role in biased selection (Olian, Schwab, & Haberfeld, 1988) and performance appraisal (Davison & Burke, 2000) processes. The research reported here examines an additional way in which gender stereotypes may give rise to deleterious consequences for women aspiring to traditionally male work roles. Specifically, we address the question of whether, and under what conditions, women are given disproportionately less credit for the success they achieve when they work jointly on tasks with men. This question is of particular relevance given the recent surge in the use of work teams in virtually every industry (Ilgen & Pulakos, 1999; Wisner & Feist, 2001), coupled with the substantial emphasis both academics and the popular press have placed on the importance of teams for successful organizational functioning.

There is a great deal of evidence that stereotypes about women continue to exist in work settings (Dodge, Gilroy, & Fenzel, 1995; Heilman, Block, & Martell, 1995; Heilman, Block, Martell, & Simon, 1989; Schein, 2001). However, why and how do they impede women's career progress? Empirical evidence indicates that these stereotypes depict women as caring and relationship

oriented but not forceful and achievement oriented (Bakan, 1966; Eagly & Karau, 2002; Heilman, 2001). Thus, women are perceived to be deficient in the attributes thought to be required for success at male sex-typed tasks. The "lack of fit" between perceptions of women's attributes and perceptions of male sex-typed job requirements leads to the conclusion that women are ill equipped to handle male sex-typed work and the expectation that they are unlikely to succeed in traditionally male roles (Heilman, 1983, 2001). These performance expectations are powerful in their impact. They create a predisposition to view women in a way that detrimentally affects how they are regarded and how their work is evaluated.

Moreover, stereotype-based negative expectations about women's performance in traditionally male domains are tenacious—there is a powerful tendency to support and maintain them. Ambiguity in the performance evaluation process helps fuel the dominance of these expectations when making evaluative judgments concerning women (Heilman, 1995; Heilman, Wallen, Fuchs, & Tamkins, 2004; Nieva & Gutek, 1980). Because ambiguity allows predispositions to shape perceptions, it encourages cognitive distortion in line with expectations, thereby preserving and perhaps even reinforcing them. Thus, research has shown that when there is ambiguity about performance quality, such as when there is no tangible work product or when the merit of the work product is difficult to assess, evaluations of women's competence when they perform male sex-typed tasks is negatively affected. Consequently, providing clear and irrefutable information about the quality of performance outcomes has come to be seen as essential if women are to progress in their careers; in fact, it is often thought to be an antidote to stereotype-based negative performance expectations.

However, sometimes even indisputably successful outcomes are not enough to overcome the impact of stereotypes. Ambiguity about performance quality is not the only type of ambiguity in the performance evaluation process that can help fuel the dominance

Madeline E. Heilman and Michelle C. Haynes, Department of Psychology, New York University.

We thank Amy Goodman and John Harnisher for their help in the early stages of this research.

Correspondence concerning this article should be addressed to Madeline E. Heilman, Department of Psychology, New York University, 6 Washington Place, Room 576, New York, NY 10003. E-mail: madeline.heilman@nyu.edu

of stereotype-based negative expectations in evaluations of women in traditionally male roles. There also can be ambiguity about the source of the performance outcome—that is, ambiguity about who is actually responsible for bringing it about. We propose that this type of ambiguity can act to invalidate the diagnosticity of even an undeniably excellent work product. Thus, although the success of their work product is indisputable, women may not be able to escape the reach of stereotype-based negative performance expectations in how they are evaluated.

Source ambiguity is important to consider as an impediment to women's career progress because in organizational settings work rarely is done individually; people tend to work together with others on tasks and projects. Moreover, because of the higher proportional representation of men than of women in traditionally male domains, women aspiring to nontraditional roles are frequently likely to be working together with men rather than other women. In such situations, when women work with others for whom performance expectations are more favorable, evaluators can maintain their stereotype-based negative expectations of women by calling into question not the quality of the performance itself but the extent to which the woman is actually responsible for it. The potential consequence, which we call *attributional rationalization*, is consistent with the tendency to avoid attributing positive outcomes to a woman's ability (Deaux & Emswiller, 1974; see Swim & Sanna, 1996, for a meta-analysis).

When individuals work together rather than individually, the work product is typically a collective one. Thus, teamwork, by its nature, can obscure the visibility of an individual's contribution because of the joint nature of the final product. When this happens, the nature of each individual's contribution to the completion of an excellent product is unclear—it can only be inferred. This ambiguity about whom is responsible for a successful outcome on a traditionally male task, much as does ambiguity about performance quality, creates a context in which perceptions can easily be distorted to conform to prior expectations. However, in this case the distortion takes a different form. The successful performance outcome is apt to be attributionally rationalized, with a high-quality product attributed to someone other than the woman team member. In that way, the woman's unexpected success can be explained away and stereotype-driven negative expectations can be left undisturbed and intact. No matter how extraordinary the work outcome, if the woman is not seen as responsible for it, the accuracy of sex stereotypes and the negative expectations they produce remain unchallenged.

Because working together with others allows stereotype-based expectations about women's competence to prevail through attributional rationalization, it is likely to detrimentally affect evaluations of women as compared with men who are members of a high-powered team working in a traditionally male arena, even though their joint work product is clearly excellent. We therefore expect that unless explicit information is provided elucidating the nature of individuals' unique contributions to the joint product, a view of women that is consistent with stereotypes will be maintained. We anticipate that women will get less credit for high-quality work outcomes and thus their work competence will be derogated and their task effectiveness devalued as compared with the men with whom they are working.

In the following series of studies, we explored how men and women who have worked together on a male sex-typed task, and

who have produced a joint product that is known to be of a very high quality, are evaluated in terms of competence and task effectiveness. In each of the three studies, research participants reviewed information about a task performed by two people—one male and one female. The decision to have women work with only one other person was based on our desire to test our ideas in the most controlled manner possible. Group size and composition are known to affect group functioning (see Levine & Moreland, 1998, for full review), and we felt that controlling for these in our initial studies by using dyads would not only allow a more direct test of our research question—whether, despite excellent outcomes, women are subject to biased judgments about their competence and task effectiveness when their work has been done together with men—but also allow for a more focused exploration of the role attributional rationalization and negative performance expectations play in this process.

In Study 1, we sought to test our general proposition that unless explicit information about individual performance effectiveness is provided, female members are evaluated more negatively than male members of a successful work dyad performing a male sex-typed task. We were interested in assessments of competence, influence on task outcome, and leadership behavior. After reviewing information about the work task, participants were informed that the two-member team had produced a highly successful work product and they were provided with either favorable group-level feedback regarding the specifics of the dyad's task performance or favorable individual-level feedback regarding the specifics of the individual members' task performance. Participants then provided their evaluations. Our first hypothesis was as follows:

Hypothesis 1: In a mixed-sex dyad that has successfully performed a male sex-typed task, female members will be evaluated less favorably than male members when joint performance information is provided but not when individual performance information is provided.

Furthermore, we expected that women's evaluations would be more affected by the type of performance information provided because it is only women who are burdened with negative expectations about performance that lead to attributional rationalization. Thus, our second hypothesis was as follows:

Hypothesis 2: In a successful mixed-sex dyad, the type of performance information, whether joint or individual, will affect evaluations of female members more than it will affect evaluations of male members; female members will be evaluated more negatively when joint performance information rather than individual performance information is provided.

Study 1

Method

Participants and Design

Participants were 60 undergraduates, 25 men and 35 women, who participated for partial fulfillment of an introductory psychology course requirement at New York University. The study design was a 2×2 between-groups factorial, with sex of target (male, female) and type of performance information (individual, group) as the two independent vari-

ables. Participants were run in groups of 2 to 6; each participant was randomly assigned one of the four experimental conditions.

Procedure

Participants were informed that we were interested in different modes of employee assessment used in organizations and in identifying the most efficient way to maximize assessment accuracy. A brief description of assessment centers and the rationale for their use was provided as well as an explanation for how and when they typically are used. Participants were then told that they would receive information about two people who were employees of different organizations and who were randomly assigned to work together as a team on a particular task at an assessment center. We also explained that although both employees were going to be evaluated during the course of the study, each participant would evaluate only one of them because of time constraints.

Participants received a packet of materials. The packet included a task description sheet containing a summary of a trained observer's notes describing the work on the task, background information sheets for each employee in the team (always one male and one female), and a task feedback form providing information about task performance.

The task, chosen to be male sex-typed, was to create an investment portfolio that would yield maximum return over a 20-year period. The observer's notes explained that the work on the task had proceeded in two stages. First, each team member had created his or her own portfolio; then they had come together in the second half of the task to work on the final joint product. Attached to the task description sheet was a photograph of the two employees standing together against a blank wall. Depictions of the two employees had been standardized for age, attractiveness, style of dress, and facial expression (neutral) in development of the stimulus materials.

Participants were provided with background information about the two employees through an information sheet that supposedly had been filled out by each of the employees. There were two parallel forms of the information sheet, both including information about current job title (Assistant Vice President of either Internal Finance or Financial Affairs), years at current job (2.5 or 3), specific duties and responsibilities (e.g., approval of annual budget requests, overseeing internal accounts, etc.), educational background (bachelor's of science in accounting–management or bachelor's of science in accounting–finance), and interests and hobbies (reading, travel and tennis or swimming, reading and music). The two versions were rotated such that the male target (or female target) was paired with one version of the information sheet half of the time and the other version the other half of the time. Information about the target employee, the one participants were asked to evaluate, was always presented first.

The task feedback form had an overall evaluation of the team's task outcome; it was always indicated to have been excellent, with a numeric rating of 92 out of 100. Also included were ratings of eight discrete tasks associated with successful task completion (e.g., consideration of stock histories, consideration of current and future political situations, awareness of risks, inclusion of a safety net, product presentation, and so forth) on a 5-point rating scale ranging from 1 (*poor*) to 5 (*excellent*). All participants were provided with positive feedback forms, with three "very good" ratings and five "excellent" ratings. The target of the evaluation was systematically varied (see below).

After reviewing the information provided, participants were instructed to complete the research questionnaire. They then were debriefed, the study was explained, and participants' questions were answered.

Experimental Manipulations

Sex of target. Target employee sex was manipulated by assigning either a male or a female name to each employee. The employee's sex was further reinforced by the team photograph, which clearly designated the employee the participant was to review.

Type of performance information. Participants received ratings on the eight aspects of task performance listed on the feedback form in one of two ways. They either were designated as ratings of the target's individual performance or ratings of how the two dyad members had performed jointly. The feedback factors and ratings were identical in both versions; only the heading on the feedback form was manipulated, reading either "Individual Assessment Form" or "Group Assessment Form."

Dependent Measures

There were three key dependent measures. The first, perceived competence, was an attribute rating, and the other two, perceived degree of influence on task outcome and presumed leadership behavior, were measures of task behavior. The competence measure was a composite based on three 9-point bipolar adjective scales (competent–incompetent, productive–unproductive, effective–ineffective), on which the participant was asked to describe the target. The coefficient alpha for the competence measure was .91. The influence measure was a composite based on two items: "To what extent do you think this individual was influential in determining the joint portfolio?" (1 = *not at all* to 9 = *very much*) and "To what extent do you think this individual was responsible for the final budget?" (1 = *not at all* to 9 = *very much*). The coefficient alpha for the influence measure was .85. The leadership measure was based on a single item. Participants were asked, "To what extent do you think that this individual took the leadership role?" (1 = *not at all* to 9 = *very much*). The correlations between the dependent measures of competence and influence, competence and leadership, and influence and leadership were .54, .47, and .45 (all $ps < .01$), respectively.

Results

Manipulation Checks

Responses to a series of questions, presumably collected for clerical reasons, indicated that the manipulations were successful. Participants' report of the name of the individual they were evaluating indicated that all of them were aware of the sex of the employee. Moreover, they all correctly indicated whether the feedback information they received concerned "an individual's performance" or "the joint performance of the two participants." In addition, responses indicated that all participants knew that the task outcome had been highly successful.

Dependent Measures

Initial analyses indicated no differences in male and female respondents on any of the dependent variables; thus, their data were collapsed for all subsequent analyses.¹

A multivariate analysis of variance (MANOVA) yielded significant main effects for both sex of target, $F(3, 54) = 4.96, p < .01$, and type of performance information, $F(3, 54) = 20.27, p < .001$,

¹ Sex of participant was entered as a third factor in an ANOVA for each dependent measure and was not a statistically significant predictor of responses on the competence, influence, or leadership measures, $F_s(1, 52) = 0.08, 0.71, \text{ and } 0.04$, respectively. Furthermore, neither two-way interaction involving sex of participant was significant for any of the three dependent measures: Sex of Participant \times Sex of Target, $F_s(1, 52) = 0.64, 1.68, 0.27$, respectively; and Sex of Participant \times Type of Performance Information, $F_s(1, 52) = 2.66, 2.78, \text{ and } 0.01$, respectively. Finally, the three-way interaction was not significant for competence, influence, or leadership, $F_s(1, 52) = 0.70, 0.01, \text{ and } 2.17$, respectively.

as well as a significant interaction effect between them, $F(3, 54) = 2.84, p < .001$. Subsequently, we conducted univariate analyses of variance (ANOVAs) as well as intercell contrasts, consistent with our hypotheses. All intercell contrasts were tested using the Fisher's least significant difference (LSD) test, with the significance level set at $p < .05$. Table 1 presents the relevant means and standard deviations for each of the dependent variables.

Competence. An ANOVA of the competence ratings revealed a significant main effect for both sex of target, $F(1, 56) = 4.58, p < .05, \eta^2 = .04$, and type of performance information, $F(1, 56) = 22.14, p < .001, \eta^2 = .28$, and a significant interaction between them, $F(1, 56) = 5.68, p < .05, \eta^2 = .09$. Intercell contrasts indicated that, consistent with our first hypothesis, when participants were provided with individual performance information, there was no significant difference in competence ratings of male and female employees, but when participants were provided with group performance information, the female employee was rated as significantly less competent than the male employee. In addition, consistent with our second hypothesis, women were rated as less competent when the performance information was for the group and not for the individual, whereas the type of performance information made no significant difference for men.

Influence on task outcome. ANOVA results indicated a significant main effect only for type of performance information, $F(1, 56) = 39.42, p < .001, \eta^2 = .41$. The Sex of Target \times Performance Information interaction also was significant, $F(1, 56) = 8.22, p < .01, \eta^2 = .13$. The intercell contrasts revealed that, as we had expected, although there was no significant difference between the female and male team member in perceived influence in the individual information condition, women were viewed as significantly less influential than men in the group information condition. The data also indicated, as we had expected, that women were rated as significantly more influential when individual performance information was provided than when group performance information was provided. Unexpectedly, this was also the case for men. Evidently, individual performance information enhanced perceptions of the influence of both male and female team members, although subsequent analyses indicated a strong tendency for the difference between the ratings in the group performance and the individual performance conditions to be greater for women than for men, $t(1) = 5.85, p < .06$.

Leadership. An ANOVA yielded significant main effects for both sex of target, $F(1, 56) = 9.42, p < .01, \eta^2 = .14$, and type of performance information, $F(1, 56) = 7.59, p < .01, \eta^2 = .12$; the interaction term was not significant, $F(1, 56) = 1.39, ns$. However,

inspection of the means indicated that the direction of differences were consistent with the predicted pattern: It appears women were thought to have taken the leadership role less than men in the group information condition but that this difference did not occur in the individual information condition. Furthermore, the type of performance information feedback provided seemed to affect judgments about the leadership behavior of women but not men, with women seen as having played less of a leadership role in the work on the task when group rather than individual performance information was provided. Thus, although the interaction was not significant, the data are consistent with our hypotheses.

Discussion

The results of this study demonstrate that despite outcome success, women can be disadvantaged in their evaluations when they work together with men on male sex-typed tasks. When the successful performance information was about joint work, female team members were regarded more negatively—as being less competent and as having been less influential and having taken less of a leadership role—than their male counterparts. It was only when feedback based on individual performance was provided to evaluators that there was no differential evaluation of female and male members of the successful teams. In addition, the type of performance information conveyed was shown to have a greater effect on ratings of women than on ratings of men, with female team members being more disadvantaged in evaluations when joint rather than individual performance information was provided.

These findings are not only consistent with our hypotheses but are also consistent with our ideas regarding the role of attributional rationalization in the discounting of women's contribution to a mixed-sex team that has worked successfully on a traditionally male task. With group-level performance information, there is ambiguity regarding the degree of individual contribution to the final work product and thus there is ample opportunity to maintain stereotype-based expectations and to discount the contribution of the person expected not to perform well. It is only in the face of individual performance information verifying the excellence of the contribution of the female team member that attributional rationalization is precluded and the woman is thus given her due as a contributor to the team's success.

It furthermore follows that because of attributional rationalization, the female team members in our study would be more detrimentally affected in evaluation by group-level performance information than would be the male team members. When the task

Table 1
Means (and Standard Deviations) for Study 1

Dependent variable	Individual feedback		Group feedback	
	Male target	Female target	Male target	Female target
Competence	8.13 _a (0.76)	8.22 _a (0.89)	7.29 _a (1.59)	5.64 _b (2.01)
Influence	7.33 _a (0.84)	7.57 _a (0.92)	6.50 _b (1.15)	5.33 _c (0.84)
Leadership	7.20 _a (1.08)	6.40 _a (2.26)	6.53 _a (1.73)	4.73 _b (1.22)

Note. The higher the mean, the higher the competence, influence, and leadership ratings. Ratings were done on 9-point scales, with $n = 15$ in each condition. Means in the same row that do not share subscripts are significantly different at $p < .05$.

is considered to be male in character, it is only women, not men, for whom the ambiguity regarding individual contribution inherent in group-level performance information is potentially pernicious, fostering attributional rationalization. Because men are expected to succeed at these tasks, the type of performance information provided should make little difference in how they are regarded; even when it is not individuating, performance information is likely to be interpreted in ways that are consistent with initial positive expectations, and men's contribution to the team's success should be readily acknowledged.

One potential alternative explanation for our results is that women are viewed as unable to perform effectively when paired with men not because of expectations about their lack of ability but because of assumptions about how men behave when they are in a mixed-sex context. If, for example, men are thought to take charge and dominate when working with women, preventing them from fully engaging in the task, then the evaluation of women's contribution to the joint effort would be minimal because of the assumed proclivities of men, not because of the perceived limitations of women. However, our data argue against this interpretation. Because the negativity directed at women was not exclusively focused on their engagement in performing the specific task but also focused on their general competence, it falls short of adequately explaining all of the differences found in the evaluations of our male and female team members.

In Study 1, we assumed that individual-level and group-level performance information would produce conditions differentially conducive to attributional rationalization, but we did not directly vary the ambiguity about individual contribution that is key to this process. Consequently, we designed a second study to build on the first and further examine the proposition that attributional rationalization is the process that drives the devaluation of women in mixed-sex work teams that are successful in accomplishing a male sex-typed task.

Study 2

In Study 2, we systematically manipulated whether there was source ambiguity—ambiguity concerning individual team members' contributions to a successful team product. We expected that if the possibility of attributing the success to someone other than the woman were precluded, then the differential evaluation of female and male team members would not occur despite the performance information being joint. We therefore devised a task, again male sex-typed, in which each of the two team members had either unique or overlapping task information. All teams were successful, and participants always were provided with group-level performance information. When team members had unique task information, their individual contributions to the outcome were clear and unambiguous because the task could not have been successfully accomplished without input from each of them. However, when the team members had overlapping task information, the degree to which each team member contributed to the outcome was highly ambiguous because either one of them, alone, could conceivably have been responsible for the team's successful outcome. We expected that unless it is made explicitly clear that each team member's individual contribution is essential for task accomplishment, women would be rated less favorably than men. Accordingly, we hypothesized the following:

Hypothesis 1: In a mixed-sex dyad that has successfully performed a male sex-typed task, female members will be evaluated less favorably than male members when each member of the dyad possesses overlapping task information but not when each possesses unique task information.

We further expected that women's evaluations would be more affected by the degree of ambiguity regarding work contribution because for male sex-typed tasks, it is only women who are subjected to negative expectations about performance that lead to attributional rationalization. Thus, we hypothesized the following:

Hypothesis 2: In a successful mixed-sex dyad, the distribution of task information, whether overlapping or unique, will affect evaluations of female members more than it will affect evaluations of male members; female members will be evaluated more negatively when overlapping task information rather than unique task information is provided.

We also included a control condition in which participants were not provided with any information concerning the nature of the task-relevant information made available to the two members. We expected results in this condition to parallel those in the overlapping information condition. That is, it was our expectation that without reason to think otherwise, people assume individuals who are working together on a task have access to the same information, thereby creating ambiguity about task contribution. We therefore hypothesized the following:

Hypothesis 3a: With no information about the distribution of task-relevant information, female members of a successful mixed-sex dyad will be evaluated less favorably than male members.

Hypothesis 3b: With no information about the distribution of task-relevant information, female members of a successful mixed-sex dyad will be rated less favorably than female members who possess unique task information and no differently than female members who possess overlapping task information.

Method

Participants and Design

Participants were 61 undergraduates, 43 women and 18 men, who participated for partial fulfillment of an introductory psychology course requirement at New York University. The study design was a 2×3 factorial, with sex of target (male, female) as a within-subjects variable and nature of task information (unique information, overlapping information, no information) as a between-subjects variable. Participants were run in groups of 2 to 6; each participant was randomly assigned one of the three task type conditions.

Procedure

The procedure was similar to that of Study 1, but in this study participants rated both members of the assessment center team and the performance information provided was always joint. As in Study 1, each participant received a packet of materials containing a description of the task (including a photograph of the two team members), background information sheets for each of them (always one male and one female), and a

feedback form containing performance information. As before, the two-person team was always depicted as having been successful in accomplishing its work.

A new male sex-typed task was devised for this study. The task description indicated that the two team members were together responsible for devising an appropriate budget for a computer software company that was opening production facilities in New York and California. Participants were told that to create the most accurate budget, the specific tax laws of each state had to be considered. Similar to Study 1, the trained observer's notes detailed that the work proceeded in two stages. First, both team members studied tax laws. They then came together in the second half of the task to work together on a final joint product.

The background information sheets were identical to those used in Study 1 and so was the photograph of the two team members. The layout of the feedback form was identical to that of the "Group Assessment Form" used in Study 1. The overall evaluation of the team's task outcome again was indicated as having been excellent, with the numeric rating being 92 out of 100.

After reviewing the information provided, we asked participants to complete the research questionnaire. They then were debriefed, the study was explained, and their questions were answered.

Experimental Manipulations

Sex of target. Employee sex was again manipulated by assigning either a male or a female name to each target; the manipulation was once again reinforced by the team photograph.

Task information. Task information was manipulated in the task description provided to participants. The description of the first phase of the task, during which the team members studied tax laws, differed in the three task conditions.

In the unique information condition, participants were informed that each of the two team members had obtained only part of the information necessary to accomplish the task fully and that therefore it had been necessary for both of them to pool their information for the team outcome to be successful. Specifically, participants were told that before working together, one team member had studied the tax laws of California whereas the other team member had studied the tax laws of New York.

In contrast, participants in the overlapping information condition were informed that each member of the team had obtained all the information necessary to accomplish the task and that therefore either of them could actually have successfully accomplished the task without the information possessed by the other. Specifically, participants were told that before working together each of the two members of the team had studied the tax laws of both California and New York.

Participants in the no information condition were not told anything about the information the two team members would be obtaining in the first phase of the task. They simply were informed that before working together the two team members had "studied tax laws."

Dependent Measures

The same three dependent measures used in Study 1 were used in this follow-up study: perceived competence ($\alpha = .87$), perceived degree of influence on the task outcome ($\alpha = .83$), and presumed leadership behavior of the team member being evaluated. The correlations between the dependent measures of competence and influence, competence and leadership, and influence and leadership were .18 (*ns*), .13 (*ns*), and .63 ($p < .01$), respectively.

Results

Manipulation Checks

Responses to several questions, said to be obtained for clerical reasons but in fact obtained to check the manipulations, indicated

that the manipulations were successful. All participants correctly reported the name of the individual they were evaluating, indicating that they were aware of the sex of the team member. Participants also were asked to indicate whether team members studied the tax laws of either one state or two states and whether it was necessary for both team members to exchange information to complete the task successfully (*yes-no*). All participants in the unique and the overlapping information conditions answered these two questions consistently with the condition to which they had been assigned. Finally, all participants in the study correctly indicated that the team's task outcome had been highly successful.

Dependent Measures

Once again, initial analyses indicated no differences in male and female respondents on any of the dependent variables; thus, their data were collapsed for all subsequent analyses.²

A MANOVA yielded a significant main effect only for sex of target, $F(3, 114) = 5.74, p < .001$, as well as a significant Sex of Target \times Type of Task interaction, $F(6, 230) = 2.79, p < .01$. Univariate ANOVAs and intercell contrasts to test our hypotheses were conducted. All intercell contrasts were tested using the Fisher's LSDs, with the significance level set at $p < .05$.³ Table 2 presents the relevant means and standard deviations of each of the dependent measures.

Competence. An ANOVA indicated that there was a significant main effect only for sex of target, $F(1, 58) = 13.07, p < .001$, $\eta^2 = .184$, and a significant Sex of Target \times Type of Task interaction, $F(1, 58) = 6.22, p < .01$, $\eta^2 = .177$. The intercell contrasts revealed results consistent with our first hypothesis. In the overlapping information condition, the male team member was considered significantly more competent than the female team member, but when each group member had access to unique task information, there was no significant difference in their perceived competence. However, there was less support for our second hypothesis: Although the means were in the hypothesized direction, the results of the contrasts revealed that unique, as compared with overlapping information, did not cause significantly lower evaluations of female (or male) team members (see Table 2). Our third hypothesis was strongly supported. As we had expected, the pattern of results in the no information condition were very much like that obtained in the overlapping information condition, with women rated as both less competent than men within condition and

² Sex of participant was entered as a third factor in an ANOVA for each dependent measure and was not a statistically significant predictor of responses on the competence, influence, or leadership measures, $F_s(1, 55) = 0.57, 0.00, \text{ and } 1.03$, respectively. Furthermore, neither two-way interaction involving sex of participant was significant for any of the three dependent measures: Sex of Participant \times Sex of Target, $F_s(1, 55) = 0.04, 0.05, 0.06$, respectively; and Sex of Participant \times Type of Task, $F_s(1, 55) = 0.99, 0.81, \text{ and } 0.33$, respectively. Finally, the three-way interaction was not significant for competence, influence, or leadership, $F_s(2, 55) = 2.07, 1.85, \text{ and } 0.24$, respectively.

³ Post hoc tests were conducted using the more general form of Fisher's LSD test: Fisher's protected *t* tests. Fisher's LSD test is a simplification of the protected *t* test that can only be used when sample sizes are equal across conditions. Because we had unequal sample sizes in Study 2, we needed to use the more general form of Fisher's protected *t*.

Table 2
Means (and Standard Deviations) for Study 2

Dependent variable	Unique information		Overlapping information		No information–control	
	Male	Female	Male	Female	Male	Female
Competence	7.07 _{a,b,c} (1.90)	7.40 _{a,c} (1.46)	8.13 _a (0.85)	6.68 _{b,c} (2.12)	7.63 _{a,c} (1.39)	6.15 _b (2.23)
Influence	6.98 _a (1.54)	7.40 _a (0.99)	6.98 _a (1.08)	5.81 _b (1.50)	7.10 _a (1.05)	6.15 _b (1.04)
Leadership	6.05 _{a,b,c,d} (1.39)	6.35 _{a,c} (1.31)	6.57 _a (1.43)	5.14 _b (1.98)	6.85 _a (1.31)	5.45 _{b,d} (1.28)

Note. The higher the mean, the higher the competence, influence, and leadership ratings. Ratings were done on 9-point scales, with $n = 20$ in the partial information and control conditions and $n = 21$ in the complete information. Means in the same row that do not share subscripts are significantly different at $p < .05$.

as less competent than women in the unique information condition. There was no difference in the rated competence of women in the no information and the overlapping information conditions.

Influence on task outcome. Results of the ANOVA again indicated a significant main effect only for target sex, $F(1, 58) = 15.78$, $p < .001$, $\eta^2 = .214$, and a significant Sex of Target \times Type of Task interaction, $F(2, 58) = 12.28$, $p < .001$, $\eta^2 = .297$. Intercell contrasts revealed a data pattern supportive of our three hypotheses. Women were rated as having been significantly less influential than men in the overlapping information condition; however, they were not rated differently in terms of influence in the unique information task condition. Moreover, the contrasts indicated that when the task information made available to the team members was overlapping rather than unique, there were significantly lower evaluations of women's influence but not of men's influence. Last, the pattern of results in the no information condition was found to mimic the pattern of results in the overlapping information condition, with female team members rated as less influential than male team members and less influential than female team members in the unique information condition. Female team members in the no information condition were rated no differently than female team members in the overlapping information condition.

Leadership. An ANOVA again yielded a significant main effect only for target sex, $F(1, 58) = 13.50$, $p < .001$, $\eta^2 = .189$, and the Sex of Target \times Type of Task interaction was also significant, $F(2, 58) = 6.16$, $p < .005$, $\eta^2 = .175$. Intercell contrasts once again revealed a pattern of results supporting our first hypothesis. Men were considered significantly more likely to have engaged in leadership behavior than their female counterparts when the task information provided was overlapping, but when task information was unique, men and women were rated no differently in their likelihood of having taken on the leadership role. The results of the contrasts also were consistent with the second hypothesis. Having overlapping rather than unique task information had significantly more negative consequences for evaluations of female team members but not male team members. Our third hypothesis also was supported; the ratings in the no information condition closely paralleled those in the overlapping information condition, with women rated as less likely to have taken on a leadership role than men and less likely to have taken on a leadership role than women in the unique information condition. Women in the no information condition were rated no more likely to have taken on a leadership role than women in the overlapping information condition.

Discussion

The results of Study 2 lend support to our hypotheses regarding the mediating role of attributional rationalization in the derogation of women working on male sex-typed tasks in successful mixed-sex work teams. Unless there was clarity about individual contribution to the successful group outcome (unique information condition), thereby precluding the attribution of responsibility for success to the male team member, women were once again rated as being less competent and as having been less influential and less likely to take the leadership role than were men. Moreover, as we had anticipated, the nature of the task information made a difference only for female, but not for male, team members; this was expected because women are the ones about whom there are negative performance expectations—expectations that are likely to color evaluations when conditions are less than unequivocally clear. Our third hypothesis also was supported. When no information was provided about the allocation of the task information, respondents' judgments were not significantly different than when there was overlapping task information, suggesting that short of assuring that the woman played a key role in the success of the team, the baseline assumption was that she did not. It appears that unless barred, attributional rationalization is a default cognitive process, ultimately serving to maintain the congruence between expectations and outcomes.

Results of Studies 1 and 2 lend support to the idea that women's contributions to the success of a mixed-sex work team working in a traditionally male arena is likely to be discounted unless source ambiguity is minimized and there is verification of their role in making the team successful. Putting it differently, women's contributions to the success of the team is discounted unless the negative expectations about their performance are constrained from dominating in the evaluation process. However, what if these negative expectations were themselves precluded? Then, according to our reasoning, the derogation of women's competence and contribution to the team's success should not occur because the force driving attributional rationalization—negative stereotype-based expectations—will have been eliminated.

This would be the case when a woman's track record does not allow for the negative performance expectations typically held about women—when evidence is available about past performance excellence that undercuts the stereotype-based expectations that have such deleterious effects on competence and task-related evaluations in circumstances where there is source ambiguity. If stereotype-based negative expectations about women are pre-

empted, then irrespective of whether the situation allows for attributional rationalization, there should be no impetus for making distinctions between the likely contribution of men and women to the team outcome. Study 3, in which we sought to directly influence performance expectations, was designed to test these ideas.

Study 3

In Study 3, we varied the specificity of information made available to research participants about team members' past on-the-job work competence. We expected that if there were very distinct, specific evidence that a female team member had a strong record of work effectiveness in traditionally male work settings, negative expectations about her performance in the team would be precluded, thereby averting the perceived incongruity between negative expectations and successful outcome that triggers attributional rationalization when source ambiguity exists. There would therefore be no reason to evaluate her contribution to the team's success less favorably than the contribution of her male teammate.

We used a similar procedure to that used in Studies 1 and 2. Participants read about a man and a woman who had worked together as a team on a task and learned that they had been highly successful in accomplishing their joint work objective. No information about the distribution of task information was provided and, as in Study 2, all participants were given group-level feedback about the team. In addition, participants received either specific information, vague information, or no information about the on-the-job work effectiveness of the team member whom they were assigned to evaluate. We expected that despite the high level of source ambiguity created by the group-level feedback, differential evaluation of women and men would not occur when there was little question that the female team member had been highly effective working in traditionally male domains in the past. Thus, we anticipated that providing clear and specific information of past performance excellence would buffer the female target against negative expectations and thus avert the process of attributional rationalization altogether and the devaluation of her contribution to a successful collective product.

Hypothesis 1: Female members of a mixed-sex dyad that has been successful will be evaluated less favorably than male members when information of their past performance excellence is vague but not when it is specific.

Moreover, because it is only women who are in need of competence verification to dispel negative expectations about their performance, we expected the manipulation of the specificity of the performance effectiveness information to have more of an effect on evaluations of female team members than on male team members.

Hypothesis 2: In a successful mixed-sex dyad, information of past performance excellence, whether vague or specific, will make more of a difference for women than for men; women will be evaluated more negatively when vague rather than specific information about past performance excellence is provided.

We also included a control condition in which participants received no information about past performance effectiveness. We

expected the results in this condition to be similar to those in the vague information conditions.

Hypothesis 3a: In a successful mixed-sex dyad with no information about past performance excellence, female members will be evaluated less favorably than male members.

Hypothesis 3b: In a successful mixed-sex dyad, female members for whom there is no information about past performance excellence will be rated less favorably than female members for whom specific performance excellence information has been provided and no differently than female members for whom vague performance excellence information has been provided.

Method

Participants and Design

Participants were 90 undergraduates, 35 men and 55 women, who participated for partial fulfillment of an introductory psychology course requirement at New York University. The study design was a 2×3 between-subjects factorial, with sex of target (male, female) and information of past performance excellence (specific, vague, no information) as the two independent variables. Participants were run in groups of 2 to 6; each participant was randomly assigned one of the six experimental conditions.

Procedure

The procedure was generally similar to that of Studies 1 and 2. Each participant received a packet of materials containing a description of the task (including a photograph of the two team members), background information—this time for only one team member—and a “Group Assessment Form” providing feedback about the specifics of the team's performance on the task. The team task used in this study was that used in Study 2. That is, the team was described as having worked on devising a budget for a computer software company that was opening facilities in New York and California, and the importance of having carefully considered the tax laws of each state was emphasized. As in the control condition in Study 2, no mention was made about the team members' information about the tax laws of the two states. We used the same photograph and group feedback form (the “Group Assessment Form”) used in Studies 1 and 2. The overall evaluation of the team's task outcome was once again indicated to have been excellent, with a numeric rating of 92 out of 100.

The format of the background information sheet was changed to deliver the manipulation. Participants were told that the form had been completed by both the employee and his or her supervisor as part of the required registration materials for the assessment center. The sheet was divided into two parts. Part 1 asked about biographical information and was to be completed by the target him- or herself. It also included a waiver, which the target had ostensibly signed, waiving the right to access the document once completed by the supervisor in accordance with the bogus “Employee Rights and Privacy Act of 1974.” Part 2 of the information sheet was supposedly completed by the employee's immediate supervisor and returned directly to the assessment center. Part 2 included information provided by the supervisor about the length of time the employee had worked under the supervisor's supervision, the employee's official title, the employee's job duties, and, in some instances, contained a performance evaluation. This part of the form was always signed and dated, presumably by the employee's supervisor.

Although the format of the background information was altered, its content was nonetheless identical to the information provided about the team members in Studies 1 and 2, with the exception of the past performance excellence manipulation itself. As in the earlier studies, we created

Table 3
Means (and Standard Deviations) for Study 3

Dependent variable	Specific past performance information (top 2%)		Vague past performance information (top 25%)		No past performance information (control)	
	Male	Female	Male	Female	Male	Female
Competence	8.36 _a (0.60)	8.31 _a (0.65)	8.07 _a (0.42)	7.12 _b (0.78)	8.07 _a (0.44)	6.80 _b (1.72)
Influence	7.37 _a (1.02)	7.47 _a (0.90)	7.30 _a (0.86)	6.23 _b (1.02)	7.47 _a (0.69)	6.57 _b (1.00)
Leadership	7.00 _a (1.25)	7.13 _a (1.25)	7.27 _a (1.16)	5.93 _b (1.28)	7.67 _a (0.82)	6.13 _b (1.06)

Note. The higher the mean, the higher the competence, influence, and leadership ratings. Ratings were done on 9-point scales, with $n = 15$ in all conditions. Means in the same row that do not share subscripts are significantly different at $p < .05$.

two different but similar sets of background information, both in finance, which were alternated in presentation in the stimulus materials in both the male and the female team member conditions.

After reviewing the stimulus materials, participants were instructed to complete the research questionnaire. They then were debriefed, the study was explained, and their questions were answered.

Experimental Manipulations

Sex of target. Employee sex was again manipulated by assigning either a male or a female name to each target, and the manipulation was reinforced by the team photograph.

Past performance excellence. Information of past performance excellence was manipulated on the information sheet provided to participants. For participants in the no information condition, Part 2 of the information sheet, supposedly completed by the target's supervisor, contained only the supervisor's description of the employee's job title, years supervised, and job duties. For those in the remaining conditions, there was an additional question asking the supervisor to circle the category that best described the employee's performance effectiveness in his or her job relative to that of other employees the supervisor had supervised. The employee was always rated in the highest category, but the range of possible categories shown on the form differed in the specific performance and vague performance conditions. In the specific performance condition, there were six rating categories: "Top 2%," "Top 5%," "Top 10%," "Top 25%," "Top 50%," and "Bottom 50%." In contrast, there were only three rating categories in the vague performance condition: "Top 25%," "Top 50%," and "Bottom 50%." Thus, although the employee was always given the highest rating possible, the response format with the more differentiated performance categories enabled a more well-defined and specific representation of past performance excellence than the response format with the less differentiated performance categories.

Dependent Measures

The same three dependent measures used in Studies 1 and 2 were used in Study 3: perceived competence ($\alpha = .87$), perceived influence on the task outcome ($\alpha = .78$), and presumed leadership behavior of the team member being evaluated. The correlations between the dependent measures of competence and influence, competence and leadership, and influence and leadership were .60, .52, and .78 (all $ps < .01$), respectively.

Results

Manipulation Checks

The results of the analyses of the manipulation checks suggest the manipulations were successful. Participants were asked to indicate the name of the person they had reviewed, and in all cases they gave the correct name, indicating that they were aware of the

employee's sex. They also were asked whether they had received supervisor information regarding the quality of the target's performance at his or her current job (*yes-no*), and all but 4 participants (2 in the specific performance condition and 2 in the vague performance condition) correctly indicated whether they had received such information. Finally, all participants indicated awareness that the team's task outcome had been highly successful.

Dependent Measures

Once again, initial analyses indicated no differences in male and female respondents on any of the dependent variables; thus, their data were collapsed for all subsequent analyses.⁴

A MANOVA yielded significant main effects for both sex of target, $F(3, 81) = 6.94, p < .001$, and past performance information, $F(6, 164) = 3.59, p < .002$, as well as a significant interaction between them, $F(6, 164) = 2.26, p < .05$. Univariate ANOVAs and intercell contrasts were conducted to test our hypotheses. All intercell contrasts were conducted using Fisher's LSDs, with the significance level set at $p < .05$. Table 3 presents the relevant means and standard deviations.

Competence. An ANOVA yielded a significant main effect for both sex of target, $F(1, 84) = 15.97, p < .001, \eta^2 = .161$, and past performance information, $F(2, 84) = 8.74, p < .001, \eta^2 = .174$. The Sex of Target \times Past Performance Information interaction was also significant, $F(2, 84) = 3.81, p < .05, \eta^2 = .084$. Intercell contrasts, testing our specific hypotheses, were highly supportive of our predictions. As expected, women were rated as significantly less competent than their male teammates in the vague performance information condition but not in the specific performance information condition, in which men and women were rated no differently in terms of competence. Furthermore, although the male team member was rated similarly regardless of the past performance information condition, the nature of the performance

⁴ Sex of participant was entered as a third factor in an ANOVA for each dependent measure and was not a statistically significant predictor of responses on the competence, influence, or leadership measures, $F_s(1, 89) = 0.01, 0.06, \text{ and } 0.17$, respectively. Furthermore, neither two-way interaction involving sex of participant was significant for any of the three dependent measures: Sex of Participant \times Sex of Target, $F_s(1, 89) = 0.06, 1.25, \text{ and } 1.34$, respectively; and Sex of Participant \times Performance Information, $F_s(1, 89) = 1.31, 0.19, \text{ and } 0.36$, respectively. Finally, the three-way interaction was not significant for competence, influence, or leadership, $F_s(2, 89) = 0.74, 0.86, \text{ and } 2.60$, respectively.

information made a significant difference in how the competence of female team members was evaluated, with their ratings decidedly lower in the vague performance information than in the specific performance information conditions. Finally, intercell contrasts indicated that those female team members for whom there was no performance information available were rated less favorably than male team members for whom there also was no performance information, less favorably than female team members for whom specific past performance information was provided, and no differently than female team members for whom vague performance information was provided.

Influence on task outcome. An ANOVA yielded significant main effects for both target sex, $F(1, 84) = 10.24, p < .01, \eta^2 = .102$, and past performance information, $F(2, 84) = 3.79, p < .05, \eta^2 = .084$. There was also a significant Sex of Target \times Past Performance Information interaction, $F(2, 84) = 3.51, p < .05, \eta^2 = .074$. Results of the intercell contrasts revealed a data pattern similar to that obtained in analyses of the competence ratings. Women were rated as having been significantly less influential than men in the vague performance information condition but not in the specific performance information condition. Once again, influence ratings of the team members were found to be negatively affected by the vague performance information as compared with the clear performance information only when they were women but not when they were men. In addition, ratings of the women in the no information condition were consistent with our expectations: They were rated as having been less influential than the men in the no information conditions and the women in the specific performance information conditions, but their ratings did not differ from those women in the vague performance information conditions.

Leadership. An ANOVA yielded a significant main effect only for target sex, $F(1, 84) = 14.18, p < .001, \eta^2 = .144$, and a significant Sex of Target \times Past Performance Information interaction, $F(2, 84) = 4.71, p < .05, \eta^2 = .101$. The results of the intercell contrasts again were highly supportive of our hypotheses. Men were considered likely to have taken significantly more of a leadership role than were their female counterparts in the vague performance information condition but not in the specific performance information condition. Furthermore, as with the other two measures, whether the performance information provided was vague or specific made a significant difference only for judgments about female (not male) team members, who were viewed as far less likely to have engaged in leadership behavior in the current task when the past performance information provided about them was vague. Also, as with the other two measures, the women in the no information condition were given leadership ratings lower than the men in the no performance information condition, lower than women in the specific performance information condition, and no different than women in the vague performance information condition.

Discussion

The results of this study again demonstrate that when the ambiguity about individual contribution that derives from working in a successful mixed-sex dyad on a male sex-typed task is unconstrained, negative consequences may result for female team members in terms of competence evaluations and assessments of likely influence and leadership activities. However, the results also indicate that these negative consequences do not always occur. With

explicit information attesting to past on-the-job performance excellence, the ambiguity about contribution inherent in evaluation of those who worked on a joint task was no longer harmful to female team members. These findings suggest that attributional rationalization is averted when negative stereotype-based performance expectations about women are undercut by clear indication of past performance excellence. They thus are highly supportive of our idea that negative stereotype-based expectations are at the root of attributional rationalization and that attributional rationalization serves to reconcile potential inconsistencies between performance expectations and performance outcomes.

It is important to note that when vague information about past performance effectiveness was provided to our research participants, it was virtually ignored when the individual being evaluated was a woman. That is, men who were indicated to be in the "Top 25%" of their cohort were rated no differently than men indicated to be in the "Top 2%" of their cohort, whereas women indicated to be in the "Top 25%" of their cohort were not only rated less favorably than similarly depicted men but were also rated no better than women for whom no past performance information had been provided. These results suggest that without a compelling reason for the rater to believe otherwise, negative expectations of women who work in male sex-typed tasks persist even in the face of clearly successful joint outcomes, resulting in the devaluation of women's competence and their contribution to the work product.

General Discussion

Taken together, the results of these studies indicate that working together with men in traditionally male domains can be detrimental for women—even when the work outcome is highly favorable. We found this to be the case unless (a) there was specific information about the female team member's individual performance excellence on the team task (Study 1), (b) the female team member's contribution to the successful joint outcome was irrefutable because of the structure of the task (Study 2), or (c) there was definitive information about the excellence of the team member's past performance effectiveness (Study 3). In the absence of these conditions, women were thought to be generally less competent, less influential in arriving at the successful team outcome, and less apt to have taken on a leadership role in the task than were their male counterparts.

Calculation of McGraw and Wong's (1992) common language effect size statistics underscores the practical importance of these findings. Across all three studies, for conditions in which there was ambiguity regarding the source of the success for the final joint product, the probability ranged between .74 and .87⁵ that a male

⁵ The common language (CL) effect size statistic converts the size of an effect into a probability: the probability that a score sampled at random from one distribution will be greater than a score sampled from some other distribution. In Study 1, the CL effect size for competence, influence, and leadership ratings, respectively, was .74, .80, and .80 in the joint feedback condition. In Study 2, the CL effect size for competence, influence, and leadership ratings, respectively, was .74, .74, and .72 in the overlapping information condition and .71, .74, and .78 in the control condition. In Study 3, the CL effect size for competence, influence, and leadership ratings, respectively, was .86, .79, and .78 in the vague past performance condition and .76, .77, and .87 in the control condition.

target selected at random would be evaluated more favorably (more competent, more influential, and more likely to take on the leadership role) than a female target selected at random, unless attributional rationalization was explicitly precluded. In other words, in at least 74 out of 100 (a minimum of 24% above chance) comparisons, men would be evaluated more favorably than women for their involvement in precisely the same excellent joint outcome, unless constrained otherwise. Furthermore, calculation of Cohen's standardized effect size,⁶ d , across all three studies indicates that the average man was frequently evaluated at or above the 85th percentile of the female distribution. In tandem, these results suggest that a women's performance must be at the top 20th percentile, and in many cases in the top 10th percentile, to be viewed on par with the average man's performance.

Results from Studies 1 and 2 strongly support the idea that ambiguity about the nature of individual contribution to a joint effort, which is typical of work on a collective product, promotes attributional rationalization and the crediting of team success to someone other than the female team member. Unless this source ambiguity was constrained, the contribution of women working together with men on male sex-typed tasks was devalued. The results from Study 3 support the idea that attributional rationalization about women's contributions to team effectiveness in such circumstances is a direct consequence of stereotype-based negative performance expectations; negative evaluations of female team members were shown not to occur if negative expectations about their likely performance were preempted by disconfirming information. Our findings therefore suggest that the ambiguity about individual contribution inherent in many joint endeavors produces a context in which negative stereotype-based expectations about women can persevere despite even the most successful outcomes and, unless preempted or restrained, produces attributional rationalization and negative evaluations of female team members. When the task is male sex-typed and the team is mixed-sex, success seems to do little to bolster the standing of the female team member.

The present findings therefore raise serious questions about the blanket effectiveness of providing performance information as a mechanism for dispelling negative evaluations of women in traditionally male sex-typed jobs. Much of the existing literature that has examined the consequences of negative performance expectations for women in male sex-typed occupations has touted the provision of information of performance success as a solution for mitigating biased evaluations. However, our results suggest that information of success is not always enough to preclude the derogation of women and their work. Attributional rationalization seems to be the default process, and unless the success information is unequivocally diagnostic of the woman, it takes hold.

Although these studies demonstrate the potential dangers of teamwork for women, their results also suggest that working in mixed-sex teams on male sex-typed tasks need not always result in negative evaluations of the women who work in them. They indicate that it is not working in a team per se but rather the ambiguity about individual contribution coupled with the expectations that are typically held for women's performance in male sex-typed tasks that fuels this negativity. Furthermore, when these are dispelled, whatever the means, so is attributional rationalization and the negative evaluations to which it gives rise. Unfortunately, the conditions under which most work teams function are

ones in which source ambiguity flourishes, not because of design but because of the collective nature of the work that teams do and their identity as a unit. Moreover, more often than not there is ignorance of past performance record so negative expectations of women are not challenged; few members of work organizations wear their latest performance evaluations on their lapels, and this information is usually highly confidential, whether it is positive or negative.

Another finding that has interesting implications is the consistent lack of difference found in the ratings of male and female research participants. This was so for all three studies. Although it makes sense to posit that women would be more sensitive to the general propensity to devalue other women, thereby either bending over backward to treat them equally or even better than men when evaluating them, these behaviors did not occur. This finding is not unique; it has been documented repeatedly in research investigating women in work settings. Evidently women are just as likely as are men to hold negative expectations about women and to consequently engage in attributional rationalization, derogating women's competence and task effectiveness in work settings.

The research presented here is only the first step in exploring attributional rationalization and its potentially deleterious effects on women in organizational settings. To further test our ideas, it is important to clarify the conditions under which it is, and is not, likely to occur. If we are correct in our reasoning, for example, the effects we demonstrated here should not occur when women work jointly with other women or others about whom there also are negative performance expectations because attributional rationalization presumes the availability of another who can be plausibly be credited with the joint success. Our ideas further suggest that the effects we found in our studies should occur only when the team task is male in sex-type. It is the negative expectations that derive from the perceived lack of fit between the attributes thought to characterize women and the attributes thought to be necessary to do male sex-typed tasks that we think drives attributional rationalization, and therefore when the task is neutral or female in sex-type, these expectations, and attributional rationalization, should not result.

It is important not only to gain greater conceptual clarity about attributional rationalization but also to consider its broader implications. We have proposed that attributional rationalization arises from source ambiguity—ambiguity about individual contribution

⁶ Cohen's d is the standardized effect size, calculated as the mean of the male target minus the mean of the female target divided by the sample size weighted within group standard deviation. One interpretation of Cohen's d can be thought of as the average percentile standing of the average experimental participant in one condition relative to the average participant in another condition. For example, a d of 0.0 indicates that the mean of the one group is at the 50th percentile of the other group; a d of 1.0 indicates that the mean of the treated group is at the 84th percentile of the untreated group. In Study 1, d s for competence, influence, and leadership ratings, respectively, were 0.91, 1.16, and 1.20 in the joint feedback condition. In Study 2, d s for competence, influence, and leadership ratings, respectively, were 0.90, 0.90, and 0.83 in the overlapping information condition and 0.80, 0.91, and 1.08 in the control condition. In Study 3, d s for competence, influence, and leadership ratings, respectively, were 1.52, 1.13, and 1.10 in the vague past performance condition and 1.01, 1.05, and 1.63 in the control condition.

to a work product. However, working with others is only one cause of ambiguity about individual contribution in organizational life. Other causes, such as mentoring programs, support groups, and organizational sponsors, may also pave the way for attributional rationalization and the ultimate devaluation of the competence and effectiveness of successful women by attributing their success to someone other than themselves. The underlying issue is not working in teams but the ambiguity that it fosters and the way in which such ambiguity provides a vehicle for maintaining negative performance expectations about women. Thus, any organizational procedure or practice that creates ambiguity about individual contribution might very well have similar effects.

There is little question that the methodology we used in our research limits the degree to which we can extrapolate from our findings. The studies used undergraduates as participants and, although the majority of them reported having had work experience, the type of work experience that they have had may limit the generalizations that we can make from our data. Moreover, we put participants in a passive observer role and they had no interaction with the team members they were evaluating nor did they expect any future contact with them. Although many evaluations in work settings are in fact done from this third party stance, it is important to explore in future research how women in successful mixed-sex teams are evaluated by other team members and also how they evaluate themselves. It also is important to see whether the results we found occur when there is more and richer information available about the team members and about other instances of their work together and to see how the size and composition of the team affects the processes in which we are interested. Last, in this set of studies we were concerned only with successful teams, but it is not clear what the implications of our findings are for mixed-sex teams who fail; this, too, is important to consider in the future.

These questions notwithstanding, the results reported here are not only dramatic but also are cause for concern. If women in nontraditional work domains who work in situations in which source ambiguity flourishes, such as successful mixed-sex teams, are denied credit for their part in bringing about the success, and are devalued simply because they are women, then there is a potential price to be paid. Women of talent and promise may be bypassed in career advancement or relegated to noncentral positions. In short, these women may be stopped from moving up the organizational ladder and lost as valuable resources for the organization.

References

- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Chicago: Rand McNally.
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225–248.
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology, 29*, 80–85.
- Dodge, K. A., Gilroy, F. D., & Fenzel, L. M. (1995). Requisite management characteristics revisited: Two decades later. *Journal of Social Behavior and Personality, 10*, 253–264.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573–598.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior, 5*, 269–298.
- Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality, 10*, 3–26.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*, 657–674.
- Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality, 10*, 237–252.
- Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology, 74*, 935–942.
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male tasks. *Journal of Applied Psychology, 89*, 416–427.
- Ilgen, D. R., & Pulakos, E. D. (1999). *The changing nature of performance: Implications for staffing, motivation, and development*. San Francisco: Jossey-Bass.
- Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 415–469). New York: McGraw-Hill.
- Lyness, K. S. (2002). Finding the key to the executive suite: Challenges for women and people of color. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for building leadership at the top* (pp. 229–273). San Francisco: Jossey-Bass.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361–365.
- Nieva, V. G., & Gutek, B. A. (1980). Sex effects on evaluation. *Academy of Management Review, 5*, 267–276.
- 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.
- Powell, G. N. (1999). Reflections on the glass ceiling: Recent trends and future prospects. In G. N. Powell (Ed.), *Handbook of gender and work* (pp. 325–345). Thousand Oaks, CA: Sage.
- Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues, 57*, 675–688.
- Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin, 22*, 507–519.
- Wisner, P. S., & Feist, H. A. (2001). Does teaming pay off? *Strategic Finance, 82*(8), 58–64.

Received April 26, 2003

Revision received September 21, 2004

Accepted September 29, 2004 ■

Copyright of Journal of Applied Psychology is the property of American Psychological Association. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Journal of Applied Psychology is the property of American Psychological Association. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.