

Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks

Madeline E. Heilman and Aaron S. Wallen
New York University

Daniella Fuchs and Melinda M. Tamkins
Teachers College, Columbia University

A total of 242 subjects participated in 3 experimental studies investigating reactions to a woman's success in a male gender-typed job. Results strongly supported the authors' hypotheses, indicating that (a) when women are acknowledged to have been successful, they are less liked and more personally derogated than equivalently successful men (Studies 1 and 2); (b) these negative reactions occur only when the success is in an arena that is distinctly male in character (Study 2); and (c) being disliked can have career-affecting outcomes, both for overall evaluation and for recommendations concerning organizational reward allocation (Study 3). These results were taken to support the idea that gender stereotypes can prompt bias in evaluative judgments of women even when these women have proved themselves to be successful and demonstrated their competence. The distinction between prescriptive and descriptive aspects of gender stereotypes is considered, as well as the implications of prescriptive gender norms for women in work settings.

There is ample evidence in the literature that women are subjected to gender-biased evaluations, with their performance on male gender-typed tasks often devalued and their competence denied (see Swim, Borgida, Maruyama, & Myers, 1989, for a review). These biased evaluations have been attributed to cognitive distortion in the service of perpetuating negative expectations of women. The negative expectations are thought to result from the inconsistency between stereotypic perceptions of what women are like and the qualities thought necessary to perform a typically male job (Dipboye, 1985; Heilman, 1983, 1995). Indeed, bias seems to flourish in situations in which there is ambiguity about performance quality and cognitive distortion can easily occur (Heilman, 1995, 2001; Nieva & Gutek, 1980; Tosi & Einbinder, 1985). However, sometimes a woman's accomplishments are clearly and irrefutably excellent. In this case, the tendency to distort is constrained; consequently, expectations are overridden, success is acknowledged, and she is accepted as competent.

With her ability to handle "man's work" demonstrated, it would seem that a woman would no longer be susceptible to gender-biased evaluation, but this conclusion may be unwarranted. Recent data tracking the advancement of 30,000 corporate managers indicated that women at upper levels of the management hierarchy received decidedly fewer promotions than comparable men (Lyness & Judiesch, 1999). These data suggest that even when they

are successful, women may be vulnerable to biased judgments. The research presented here addresses this issue. Specifically, we posited that, although it is likely to dispel perceptions of her incompetence a woman's success can create new problems for her by instigating her social rejection. It was our aim to demonstrate this phenomenon, to provide insight into why and under what conditions it is likely to occur, and to examine its consequences for how women are evaluated and rewarded in work settings.

Key to our argument is the dual nature of gender stereotypes. Gender stereotypes not only are descriptive but also are prescriptive. That is, gender stereotypes not only denote differences in how women and men actually are but also denote norms about behaviors that are suitable for each—about how women and men "should be" (Burgess & Borgida, 1999; Eagly & Karau, 2002; Heilman, 2001; Rudman & Glick, 2001). There is substantial overlap between the content of the prescriptive and descriptive elements of gender stereotypes, with the behavior that is prescribed being directly related to the attributes that are positively valued for each sex. Thus, the socially sensitive and service-oriented communal traits for which women are so positively valued (Eagly & Mladinic, 1989; Eagly, Mladinic, & Otto, 1991) are a central part of their "shoulds." But gender prescriptions also include "should nots." For women, these typically include behaviors associated with men that are believed to be incompatible with the behaviors deemed desirable for women. Thus, the self-assertive and tough, achievement-oriented, agentic behaviors for which men are so positively valued are typically prohibited for women.

Although most efforts to explain gender bias have focused attention on the descriptive content of gender stereotypes and the self-fulfilling expectations that they produce, gender-stereotypic prescriptions also are likely to prompt bias in evaluations of women. When there is no opportunity to distort the nature or the source of stereotype-inconsistent behavior and it therefore is acknowledged to have occurred, its inconsistency with normative

Madeline E. Heilman and Aaron S. Wallen, Department of Psychology, New York University; Daniella Fuchs and Melinda M. Tamkins, Department of Organization and Leadership, Teachers College, Columbia University.

We thank Bill Battle and Evan Sung 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: mehl@nyu.edu

requirements provokes reactions that can negatively influence evaluative judgments.

When a woman is acknowledged to have been successful at performing male gender-typed work, she is, by definition, thought to have the attributes necessary to effectively execute the tasks and responsibilities required. But it is these same attributes that are in violation of gender-prescriptive norms. So, although there is a good fit between what the woman is perceived to be like and what the job is thought to entail, there is a bad fit between what the woman is perceived to be like and the conception of what she should be like. Also, like other counternormative behavior, this perceived violation of the stereotypic prescription is likely to arouse disapproval and subsequent penalties (Cialdini & Trost, 1998). Indeed, the construct of *fear of success*, so popular in the 1970s (Horner, 1970, 1972), was based on the idea that women's motivation in achievement situations was inhibited by their fear of disapproval for not being feminine.

Penalties for women who violate gender-stereotypic prescriptions by being successful are apt to take the form of social censure and personally directed negativity. Research findings have indicated that women who behave in ways typically reserved for men are found to be less socially appealing than men who behave similarly or women who behave more in line with normative prescriptions (Bartol & Butterfield, 1976; Carli, 1990; Carli, LaFleur, & Loeber, 1995; Jago & Vroom, 1982; Rudman, 1998). These studies indicated that social penalties result when women have actually engaged in behavior that is distinctly counter to gender-normative prescriptions. We are suggesting that the mere recognition that a woman has achieved success on a traditionally male task produces inferences that she has engaged in counternormative behavior and therefore causes similarly negative consequences.

There is some indication in the literature that success can be costly for women in terms of social approval. Competent women as compared with competent men have been depicted as cold (Porter & Geis, 1981; Wiley & Eskilson, 1985) and undesirable as fellow group members (Hagen & Kahn, 1975) and have been found to elicit visible cues of negative affect (Butler & Geis, 1990). Also, successful female managers have been described as severely wanting interpersonally (e.g., bitter, quarrelsome, selfish, deceitful, and devious) as compared with similarly successful male managers (Heilman, Block, & Martell, 1995; Heilman, Block, Martell, & Simon, 1989).

The results of these studies lend support to the idea that a woman's success in areas traditionally reserved for men can give rise to social penalties, causing them to be disliked and negatively viewed. These results also are suggestive about the nature of the negative characterizations likely to result. When women violate gender prescriptions by being successful in areas that are not traditionally part of their domain, they seem to be cast in a light that not only is negative but also is antithetical to the traditional stereotype of women and conceptions of how they should be. That is, the achievement of success appears to provoke a boomerang reaction, with successful women seen not just as noncommunal but as countercommunal—as hostile in their dealings with others. The objective of our first study was to systematically investigate these ideas.

In this study we sought to demonstrate the reactions to women and men working on a male gender-typed job when their perfor-

mance on that job was clearly successful rather than ambiguous with regard to performance outcome. Our focus was on competence and achievement-related attribute ratings as well as on liking and interpersonal hostility ratings. We expected that whether or not performance information was clear would have major consequences for how women were viewed. Consistent with earlier research, we predicted the following:

Hypothesis 1. In a male gender-typed job, women will be rated as less competent and less achievement oriented than men when information about performance outcome is ambiguous but not when success is clear.

However, we expected a different pattern of social reactions:

Hypothesis 2. In a male gender-typed job, women will be rated as less likable and more interpersonally hostile than men when information about their success is clear but not when the performance outcome is ambiguous.

Study 1

Method

Subjects and Design

Subjects were 48 undergraduates from a large northeastern university (20 men, 27 women, and 1 who did not indicate sex) recruited from an introductory psychology course in which over 90% of the students enrolled typically reported having had work experience. Their mean age was 20.5 years. Subjects came to the psychology lab and participated in groups of 2–8, and they received credit toward experiment requirements in return for their participation in the study. During the experiment each subject was exposed to manipulations of two different levels of two independent variables: sex of stimulus person (male or female) and clarity of performance outcome (unclear or clearly successful). Subjects in each condition reviewed both a male and a female target in the 2 × 2 factorial design with repeated measures on the sex of stimulus person variable. Twenty-four subjects were randomly assigned to each of the two clarity of performance outcome conditions.

Procedure

The experimenter informed the subjects that the study concerned personnel decision making in work settings and, in particular, how people combine different sources and types of information when evaluating others. The experimenters explained that they would be reading about and reacting to several employees holding the same position in a large organization. The subjects were given packets containing information about the job and about three employees and were asked to give their reactions to each employee on a brief questionnaire.

The stimulus packet began with a job description summary indicating the incumbent's position, which was an Assistant Vice President (AVP) for sales in an aircraft company, and the responsibilities of the job (e.g., training and supervising junior executives, breaking into new markets, keeping abreast of industry trends, generating new clients). The gender-typed nature of the job was communicated via the products involved, including engine assemblies, fuel tanks, and other aircraft equipment and parts. Also bolstering the male gender type of the job was a sheet listing the names of 10 employees to potentially be evaluated, 8 of whom were men and only 2 of whom were women; the employees the subject was to review were designated with a check mark. Next, there was a background information sheet containing both a concise description of each target stimulus person's background and a statement about his or her current status in the

company. The background information included birthplace, college attended, grade point average, tenure within the company, management training history, present number of employees supervised, and a listing of personal interests. The current status information concerned the stimulus person's annual performance review.

After reading the information about the first stimulus person, subjects rated him or her on a two-page research questionnaire. Subjects then read about and rated the other stimulus person. Materials for a third stimulus person, who was always male and therefore consistent with the gender type of the job, were also included in the research packet to make our focus on employee gender less salient; however, ratings of this third stimulus person were not of interest. Accordingly, the order of presentation of the two target stimulus people was systematically rotated so that the woman was first half of the time, and the man was first the other half of the time; the dummy stimulus person was always presented last. Also, the specific information provided about the male and female targets was alternated so that each description was given for the male stimulus half of the time and for the female stimulus the other half of the time. The information provided about the dummy stimulus person was always the same. After the questionnaires had been completed and collected, the experimenter gave both a written and a verbal debriefing that revealed the purpose of the experiment and the manipulations used in the study.

Experimental Manipulations

Sex of stimulus person. Each subject was exposed to both a male and a female target (as well as a third dummy target who was male). Information about sex was manipulated by the names on the background information sheets.

Clarity of performance outcome. The clarity of performance outcome was manipulated by whether or not the target was said to have already gone through the company's annual performance review. This was done in the Current Status section of the background information sheet. In the unclear performance outcome condition, a standard statement appeared on both the male and female stimulus persons' information sheets indicating that the stimulus individual was "about to undergo her (his) annual performance review" and that her or his evaluation would be "based on sales volume, number of new client accounts, and actual dollars earned." In the clear success condition, however, the target individual was reported to have already undergone the annual review and to have been designated as a top performer by the organization. There were two alternate forms of the clear success manipulation; the male and female stimulus persons in the clear success condition had one of the following statements on their background sheets:

Andrea (James) has recently undergone the company-wide annual performance review and she (he) received consistently high evaluations. She (he) has been designated as a "stellar performer" based on sales volume, number of new client accounts, and actual dollars earned. Her (his) performance is in the top 5% of all employees at her (his) level.

or

Andrea (James) has undergone the company-wide performance review and she (he) was evaluated very highly by all reviewers. She (he) was highly praised for her (his) sales volume figures, number of new client accounts, and actual dollars earned. She (he) has been identified as one of a small group of rising stars. Her (his) performance is in the top 5% of all company AVPs.

Dependent Measures

There were two different measures of competence and two different measures of liking. First, there was a composite competence score (coef-

ficient $\alpha = .81$) based on three 9-point bipolar adjective scale ratings describing the stimulus individual (*competent–incompetent*, *productive–unproductive*, and *effective–ineffective*) and a composite liking score (coefficient $\alpha = .74$) based on one bipolar adjective scale rating (*likable–not likable*) and responses to the question, "How much do you think you would like this individual?" (*very much–not at all*). Second, subjects provided comparative judgments of the employees they reviewed. After reviewing both the female employee and the male employee, they indicated (a) whom they thought to be more competent and (b) whom they thought they would like better.

In addition, we obtained characterizations of the target individuals in terms of achievement-related attributes and in terms of interpersonal hostility by creating composites of ratings on 9-point bipolar adjective scales. The scale items were culled from those used in previous research (Heilman et al., 1995; Heilman et al., 1989). The items included in the achievement-related measure were *unambitious–ambitious*; *passive–active*; *indecisive–decisive*; *weak–strong*; *gentle–tough*; *timid–bold*; and *unassertive–assertive* (coefficient $\alpha = .83$). The items included in the interpersonal hostility measure were *abrasive–not abrasive*; *conniving–not conniving*; *manipulative–not manipulative*; *not trustworthy–trustworthy*; *selfish–not selfish*; and *pushy–accommodating* (coefficient $\alpha = .84$).

Results

Manipulation Checks

To determine whether the clarity of performance outcome manipulation was effective, we asked subjects, "How successful has this individual been in the current job?" (with response choices of "very successful," "not very successful," and "I have no information about this"). Responses indicated that our manipulation was highly effective: All but 1 of the subjects in the clear success conditions viewed the stimulus persons they evaluated as "very successful," and all but 3 of the subjects in the unclear performance outcome conditions indicated that they did not have the information necessary to make this judgment.

To make certain that the job was seen as male gender typed, we asked subjects to indicate whether most people holding the job they reviewed were either men or women. To avoid being too conspicuous, we buried this question among other questions in the final questionnaire asking for assessments of job composition, for example, whether most people were older or younger than 35 years of age or were or were not college graduates. Responses indicated that 100% of the subjects believed the job holders to be predominantly men. Furthermore, our data indicated that subjects experienced a gender-role incongruity when confronted with a female employee working at the AVP job. An analysis of variance (ANOVA) of responses to the question, "How surprised were you to find this individual in this job?" (*very surprised–not at all surprised*) revealed a significant main effect, $F(1, 46) = 69.02$, $p < .01$, indicating that subjects were more surprised to find a female employee ($M = 5.64$) than a male employee ($M = 2.52$) in the job.

Data Analysis

There were two types of measures in this study—composite 9-point scales and dichotomous, forced-choice questions. A multivariate analysis of variance (MANOVA) was conducted on the four composite scales (competence evaluations, liking, achievement-related attributes, and interpersonal hostility). Results indi-

cated significant main effects for both sex of stimulus person, $F(4, 43) = 9.06, p < .01$, and clarity of performance outcome, $F(4, 43) = 40.57, p < .01$, and a significant Sex of Stimulus Person \times Clarity of Performance Outcome interaction, $F(4, 43) = 11.53, p < .01$. Univariate ANOVAs were therefore conducted for each of the four dependent variable scales, and intercell comparisons using Fisher's least significant difference (LSD) tests, with the significance level set at $p < .05$, were conducted both for clarification and to directly examine differences of pertinence to the hypotheses. The intercorrelations between the four dependent measures are presented in Table 1,¹ and Table 2 presents the means and standard deviations for each of the dependent measure scales.

Responses to the two forced-choice questions, one involving a choice between the two target employees on the basis of competence evaluations and the other involving a choice on the basis of liking, were analyzed with chi-square tests. Table 3 presents the frequencies of the competence choices and the liking choices in each experimental condition.

Initial ANOVAs with subject sex included as a factor indicated no significant main effects or interactions involving subject sex on any of the dependent measure scales. Moreover, the results of chi-square analyses of the liking and competence choices were no different when the data of male and female subjects were analyzed separately. Consequently, the data of male and female respondents were combined for all subsequent analyses.

Competence

Scale ratings. An ANOVA of subjects' ratings on the competence scale revealed a significant main effect for clarity of performance outcome, $F(1, 46) = 116.61, p < .001, \eta^2 = .72$, indicating that those in clear success conditions were seen as generally more competent than those in unclear performance outcome conditions. Also evident was a significant main effect for sex of stimulus person, $F(1, 46) = 36.07, p < .001, \eta^2 = .44$, as well as a significant Sex of Target \times Clarity of Performance Outcome interaction, $F(1, 46) = 19.88, p < .001, \eta^2 = .30$. Intercell contrasts were conducted to further clarify these effects and to test the specific hypotheses of the study. The results of these tests supported our predictions. There was no significant difference between the male and female targets on ratings of competence when the individual's prior success was made explicit; however, when information about performance outcome was left ambiguous, the female target was rated as significantly less competent than the male target. In fact, the Fisher's LSD tests revealed the female target in the unclear performance outcome condition to be rated as less competent than targets in all other conditions.

Comparative judgments. The results of the chi-square analyses of the competence choices closely paralleled those of the rating scale. When subjects were asked whom they thought was the more competent—the male or the female employee—their responses also supported our predictions. Whereas there was no significant difference in the frequency of choosing men or women when their success was clear, $\chi^2(1, N = 21) = 2.38, ns$, women were almost never indicated as the more competent when information about performance outcome was not provided, $\chi^2(1, N = 22) = 11.72, p < .01$ (see Table 3).

Liking

Scale ratings. An ANOVA of the ratings of liking of the target individual revealed a significant main effect for sex of target, $F(1, 46) = 4.54, p < .05, \eta^2 = .09$, and a Clarity of Performance Outcome \times Sex of Target interaction, $F(1, 46) = 7.26, p < .01, \eta^2 = .14$. Intercell comparisons were conducted to help interpret these results, and they provided support for our predictions. As expected, responses to male and female targets differed depending on the clarity of the performance outcome information provided about them, and these responses contrasted sharply with those evidenced in the competence ratings. When there was ambiguity about the target person's performance outcome, there was no significant difference between the liking ratings of male and female targets, but when there was clear evidence of success, the female target was liked significantly less than the male target. In fact, the Fisher's LSD tests indicated that the clearly successful female was liked significantly less than the targets in each of the other conditions.

Comparative judgments. The analysis of the liking choices yielded similar results to those of the scale ratings. When subjects were asked whom they thought they would like better—the male or the female employee—the pattern of their choices was consistent with our hypotheses. As can be seen in Table 3, whereas subjects in unclear performance outcome conditions were no more likely to choose the male than the female employee, $\chi^2(1, N = 22) = 0.36, ns$, subjects in clear success conditions overwhelmingly demonstrated a preference for the male rather than the female employee, $\chi^2(1, N = 23) = 5.60, p < .05$.

Attribute Ratings

Achievement-related attributes. An ANOVA of the achievement-related attributes scale revealed a main effect for clarity of performance outcome, $F(1, 46) = 18.18, p < .001, \eta^2 = .28$, indicating that those who were clearly successful were rated more

Table 1
Study 1: Intercorrelations Between Dependent Measures

Dependent variable	1	2	3	4
1. Competence	—	.07	.76**	.35**
2. Liking		—	.10	-.36**
3. Achievement-related attributes			—	.30*
4. Interpersonal hostility				—

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

¹ Correlations between the liking measure and the competence ratings and achievement-related attribute ratings were positive and statistically significant at the $p < .05$ level for male targets ($r = .31$) and not significant for female targets ($r_s = -.22$ and $-.15$, respectively). Furthermore, the correlations between the interpersonal hostility ratings and the competence and achievement-related attribute ratings were negative and statistically significant at the $p < .01$ level for female targets ($r_s = .54$ and $.45$, respectively) but not for male targets ($r_s = .18$ and $.15$, respectively). For both male and female targets, there was a significant correlation between competence ratings and achievement-related attribute ratings and also between liking and interpersonal hostility ratings.

Table 2
Study 1: Means and Standard Deviations for Dependent Measures

Condition and target	Competence		Liking		Achievement attributes		Interpersonal hostility	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Clear success								
Male target	8.21	0.69	7.13	1.01	7.32	1.02	5.29	1.03
Female target	8.03	0.92	5.81	1.35	7.16	1.19	3.99	1.16
Ambiguous success								
Male target	7.11	0.88	6.79	1.44	6.78	0.86	5.47	1.12
Female target	5.51	0.80	6.94	0.88	5.70	0.91	6.01	1.06

Note. $n = 24$ in each condition. All ratings were done on 9-point scales. The higher the mean, the more favorable the ratings, that is, the more competent, likable, achievement oriented, and less interpersonally hostile the rating.

favorably in terms of achievement-related attributes than those whose performance outcome was unclear. A significant main effect for sex of target person, $F(1, 46) = 9.63, p < .01, \eta^2 = .17$, together with a significant two-way interaction, $F(1, 46) = 6.82, p < .05, \eta^2 = .13$, prompted further analyses. Intercell comparisons on the achievement-related attributes scale produced results similar to those obtained on the competence scale. Whereas the female target was rated significantly more negatively than the male target in the unclear performance outcome condition, women were seen no differently than men in the clear success condition. Additional comparisons also made clear that the female target in the unclear success condition was rated least favorably in terms of achievement-related attributes than those in all other conditions.

Interpersonal hostility. An ANOVA of the interpersonal hostility scale revealed main effects for clarity of performance outcome, $F(1, 46) = 17.51, p < .001, \eta^2 = .28$. There also was a significant Clarity of Performance Outcome \times Sex of Target Person interaction, $F(1, 46) = 22.25, p < .001, \eta^2 = .33$. Intercell comparisons indicated that the female target was rated as significantly less hostile than the male target in the unclear performance outcome condition but was rated as significantly more hostile than the male target when she had been clearly successful. Furthermore, the Fisher's LSD tests indicated that the female target who was clearly successful was rated significantly more negatively in terms of interpersonal hostility than all other targets in the study.

Discussion

The results of this study supported our hypotheses. Women were viewed as less competent and characterized as less achievement

oriented than men only when there was ambiguity about how successful they had been; when the women's success was made explicit, there were no discernible differences in these characterizations. However, when success was made explicit, there was differentiation between women and men in how they were viewed interpersonally, with women deemed to be far less likable and more interpersonally hostile. Thus, our findings indicate that being successful does not necessarily put an end to problems for women holding traditionally male jobs. They also lend support to the idea that gender bias can derive from the prescriptive as well as the descriptive properties of gender stereotypes.

The results not only are suggestive about the effects of the two aspects of gender stereotypes but also suggest that in performance-based evaluation, the bias deriving from descriptive elements of the stereotype takes precedence over that deriving from the prescriptive element. It seems that people prefer to maintain gender stereotypes of women as lacking in achievement-related skills and attributes and will engage in the cognitive distortion necessary to accomplish this unless they are constrained from doing so. Thus, it is only when a woman's success is a fact, and the inconsistency with the descriptive stereotype is therefore undeniable, that the disapproval arising from the violation of prescriptive norms comes into play.

Finally, it is interesting to note just how strong the reaction is to women who have proved themselves to be successful at a male gender-typed task. Women, although rated less competent and achievement-oriented than men when information about success was ambiguous, were also rated as less hostile interpersonally. This is not a surprise given the "goodness" of women portrayed in the descriptive stereotype (Eagly & Mladinic, 1989; Eagly et al., 1991). But the turnaround with clear indication of success is dramatic—women who are acknowledged as successful are viewed not merely as indifferent to others but as downright uncivil.

Although these data are suggestive about the mediating role of the perceived violation of stereotyped gender prescriptions in bringing about social penalties for successful women, we have not yet provided a direct test of this idea. To do so, we would have to demonstrate that success, in and of itself, is not the precipitant for these penalties, but only success that is a violation of gender prescriptions, that is, success thought to require behavior traditionally reserved for men. In the following study subjects reviewed and evaluated men and women who were all highly successful, but at jobs of different gender types. We expected the following:

Table 3
Study 1: Frequencies of Target Choices for Competence and Liking

Condition and target	Competence choice		Liking choice	
	Frequency	%	Frequency	%
Clear success				
Male target	8	38	19	83
Female target	13	62	4	17
Ambiguous success				
Male target	19	86	10	45
Female target	3	14	12	55

Hypothesis 1. Successful women as compared with successful men will be rated as less likable and more interpersonally hostile when the job is male in gender type but not when it is female or neutral in gender type.

Study 2

Method

Subjects and Design

Subjects were 63 undergraduates at a large northeastern university (24 women and 39 men) enrolled in an introductory psychology course in which over 90% of the students reported having had work experience. Their mean age was 19.5 years. As in Study 1, they participated in groups of 2–8, and they received credit toward their experimental requirements in return for their participation in the study. Two independent variables were manipulated—sex of stimulus person (male or female) and job gender type (male, female, or neutral). Subjects reviewed both a male and a female target in the 2×3 factorial design with repeated measures on the sex of stimulus person variable. Twenty-one subjects were randomly assigned to each of the three job gender-type conditions.

Procedure

The general procedure for this study followed that of Study 1. The experimenter explained to the subjects that they would be reviewing three employees, all of whom worked at the same job but at different affiliates of one large company. Each subject was given a packet of materials containing a description of the job, short biographical descriptions of the three employees, and questionnaires on which to give their reactions to them.

The contents of the stimulus packet largely followed the format of that used in Study 1. There was a job description summary containing information about the job, which in all cases was reported to be an Assistant Vice President (AVP) of Human Resources. The description of the job, together with the sheet listing the 10 potential individuals to be reviewed in the study (with the ones the subject was to review designated with check marks), was the vehicle for the job gender-type manipulation described in the next section. The job description was followed by a background information sheet, which contained information about each target's history and background as well as his or her current status in the company. The history and background information was almost identical to that provided in Study 1, except for some minor changes in personal interests stated. The current status information was the same as that provided in the clear success conditions in Study 1, with the exception of the criteria for the company-wide evaluation. The criteria were altered to be relevant to an Assistant Vice President of Human Resources position (number of employees serviced, quality of workshops offered, etc.). In all cases, the employee was said to have undergone the company evaluation and to have been designated as a top performer. The same two versions of the report of the company performance review used in Study 1 were used again here.

After reviewing an employee's background information, subjects responded to a brief questionnaire. As with Study 1, materials were presented for three stimulus people, the last of whom was not of interest to us. This third individual's sex was consistent with the gender type of the job condition into which the subject had been placed. Also, as in Study 1, the order of presentation of the male and female target stimulus people was rotated, and the specific information provided about the male and female targets was alternated. The same information was always provided about the third, dummy stimulus person. After all the materials had been collected, the experimenter provided subjects with both a written and a verbal debriefing that explained the purpose of the experiment and the manipulations used in the study.

Experimental Manipulations

Sex of stimulus person. Information about the sex of the male and female targets was manipulated by the names used in the description of each employee, with the dummy target given a male name, female name, or name that could be either male or female in the male, female, and neutral job gender-type conditions, respectively.

Job gender type. The gender type of the job was manipulated in the job description summary. Although in all cases the position was presented as that of Assistant Vice President of Human Resources, the division in which the employee was said to be working differed, and the responsibilities said to be associated with the position also differed. Specifically, the stimulus person was said to be working in the Financial Planning Division (a male type position), the Employee Assistance Division (a female-type position), or the Training Division (a neutral-type position). These divisions were selected on the basis of pretesting with the same population that was to serve as subjects in the study; their ratings indicated each of the AVP jobs significantly differed from one another on a 9-point *masculine-feminine* scale. Responsibilities and task requirements for the AVP jobs also depended on the division:

Financial planning: Supervises a unit within Human Resources that provides financial planning information to employees. Helps inform employees about within-company benefit options through individual appointments and in-house workshops, and locates out-of-company sources that can aid them in mapping out long-term financial strategies for themselves and their families. Needs to be good with numbers and knowledgeable about banking, insurance, accounting, and bond and equity investment.

Employee Assistance: Supervises a unit within Human Resources that provides assistance to employees with personal and family problems. Helps counsel employees about mental health problems through individual appointments and in-house workshops, and refers employees to professionals who can aid them in coping with issues affecting their work performance. Needs to have good interpersonal skills, sensitivity to the concerns of others, and the ability to build trusting relationships.

Training: Supervises a unit within Human Resources that provides skill training to employees who seek to upgrade their positions within the company. Helps inform employees about job advancement opportunities through individual appointments and in-house workshops, and refers them to professionals who can aid them in developing long-term career goals. Needs to be a good communicator and knowledgeable about job and career planning.

A section labeled *Additional Responsibilities* included other responsibilities of the particular AVP job that were meant to reinforce the gender-type manipulation. These included tasks such as staying abreast of programs and practices within the industry concerning life insurance and mortgage assistance (Financial Planning AVP), on-site day care (Employee Assistance AVP), and paid leaves for taking courses (Training AVP). Finally, in a section labeled *Characteristics of AVPs*, the sex distribution of employees in the job was presented (86% male or female in the male- and female-type jobs, respectively, and 53% male in the neutral gender-type condition). Sex distribution was only one of several AVP characteristics reported; others included average age and education level of the typical job holder.

The names on the list of employees to be evaluated also were designed to reinforce the gender-type manipulation. In the male gender-type condition, 8 out of the 10 employees listed were men and only 2 were women, whereas in the female gender-type condition, 8 out of the 10 employees listed were women and only 2 were men. In the neutral gender-type conditions, 4 of the listed employees had male names, 4 had female names, and 2 had names that could be either male or female.

Dependent Measures

The same interpersonal reaction measures were collected as those in Study 1: the liking measures (both the scale and the comparative judgment) and the interpersonal hostility measure. The coefficient alpha for the liking scale was .76 and for the interpersonal hostility measure was .74.

Results

Manipulation Checks

Responses to the question of whether most people holding the job were men or women indicated that we had successfully manipulated the gender type of the job. All but 2 of the subjects in the male job conditions indicated that they believed job holders to be “mostly men,” all but 3 of the subjects in the female job conditions indicated that they believed the job holders to be “mostly women,” and all but 4 of the subjects in the neutral job conditions indicated that they believed the job holders to be “about equal numbers of men and women.” In addition, an ANOVA of responses to the question, “How surprised were you to find this individual in this job?” revealed a significant interaction effect, $F(2, 60) = 9.74, p > .01$, indicating that subjects experienced sex role incongruity in both the female and male job conditions. For the male job, subjects indicated more surprise to find a female employee ($M = 5.24$) than a male employee ($M = 3.29$), whereas for the female job, subjects indicated more surprise to find a male employee ($M = 4.38$) than a female employee ($M = 3.33$). There was no difference in surprise in the neutral job conditions, in which subjects were equally nonsurprised to find either a female or a male employee ($M = 3.43$ for both male and female employees).

Responses also indicated that, as we had intended, subjects viewed the employee as successful. Ninety-nine percent of all employees reviewed were rated as having been “very successful” (rather than “somewhat successful” or “not very successful”) in the current job.

Data Analysis

As in Study 1, there were two types of measures in this study—rating scales and dichotomous choices. A MANOVA was conducted on the two continuous dependent variable measures (liking and interpersonal hostility characterizations). Results revealed a significant main effect for job type, $F(4, 118) = 2.74, p < .001$, and a significant Target \times Job Type interaction, $F(4, 118) = 8.45, p < .001$. We then conducted univariate ANOVAs for each of the two dependent variable scales and did intercell comparisons both to clarify the results and to directly test our hypothesis. As in Study 1, we used Fisher’s LSD tests, with a $p < .05$ significance level, to make the intercell comparisons. The correlation between the two scale measures was $-.57 (p < .01)$.² The means and standard deviations are presented in Table 4. Responses to the forced-choice question concerning liking preferences were analyzed with chi-square tests. The frequency of liking choices is presented in Table 5.

ANOVAs on each of the dependent measures scales using subject sex as a factor indicated no significant main effects or interaction effects involving subject sex, and chi-square analyses of the liking preferences done separately for male and female subjects did not differ in their outcomes. Consequently, the re-

Table 4
Study 2: Means and Standard Deviations for Dependent Measures

Condition and target	Liking		Interpersonal hostility	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Male gender-typed job				
Male target	7.33	1.19	6.62	1.15
Female target	6.36	1.11	5.98	1.11
Female gender-typed job				
Male target	6.74	1.60	6.86	1.13
Female target	7.21	1.14	7.20	1.06
Neutral job				
Male target	7.14	0.94	6.52	1.30
Female target	7.76	0.89	7.10	0.95

Note. $n = 21$ in each condition. All ratings were done on 9-point scales, and the higher the mean, the more favorable the ratings, that is, the more likable and the less interpersonally hostile the rating.

sponses of male and female subjects were once again combined for all analyses.

Liking

Scale ratings. An ANOVA of subjects’ ratings on the liking scale only revealed a significant interaction effect between the sex of the employee and the gender type of the job, $F(2, 60) = 18.04, p < .01, \eta^2 = .38$. Subsequent intercell comparisons indicated that, as we had predicted, in the male job condition female employees were found to be significantly less likable than male employees. This contrasted with the other two job-type conditions in which female employees were found to be either significantly more likable (neutral job type) or marginally ($p < .06$) more likable (female job type) than male employees.

Comparative judgments. When the subjects were asked to indicate whom they thought they would like better—the male or the female employee—the pattern of their choices also supported our hypotheses. Chi-square analyses indicated that whereas in the male job condition subjects chose the female employee significantly less frequently than the male employee, $\chi^2(1, N = 19) = 4.26, p < .05$, this did not occur in the female and neutral job conditions. Instead, in these latter two conditions a greater number of subjects thought they would like the female employee better than the male employee, $\chi^2(1, N = 21) = 8.05, p < .01$, and $\chi^2(1, N = 21) = 3.86, p < .05$, for the female and the neutral job conditions, respectively.

Interpersonal Hostility

Analysis of ratings on the interpersonal hostility scale indicated a significant interaction effect, $F(2, 60) = 8.34, p < .01, \eta^2 = .22$. Intercell comparisons produced results that paralleled the liking ratings. Although the difference did not quite reach our predeter-

² When correlations were calculated separately for male and female targets, no difference in the pattern of correlations was found in either Study 2 or Study 3.

Table 5
Study 2: Frequencies of Target Choices for Liking

Condition and target	Frequency	%
Male gender-typed job		
Male target	14	74
Female target	5	26
Female gender-typed job		
Male target	4	19
Female target	17	81
Neutral job		
Male target	6	29
Female target	15	71

mined significance level, there was a clear tendency for women to be characterized as more interpersonally hostile than men when the job was male gender typed ($p < .06$). However, women were characterized as significantly less interpersonally hostile than men when the job was either female or neutral in gender type.

Discussion

The findings of Study 2 provided strong support for our hypotheses. Not only did we replicate the results of Study 1 concerning the effects of women's success on social reactions to them, but we also demonstrated that these effects were limited to situations in which the success was in a domain considered inconsistent with behavior suitable for a woman. The fact that negative reactions to successful women occurred only when the job was male in gender type, but not when it was female or neutral in gender type, argues for the idea that these negative reactions derive from disapproval for stereotype-based norm violation. Success for women is OK, it seems, unless it is in an area deemed off-limits for them.

The data also lend insight into the potential asymmetry of penalties for gender prescription violation. Successful men occupying female gender-typed jobs seemed to elicit the same type of negative ratings that were directed at successful women occupying male gender-typed jobs, and they also were chosen less often than women as the person subjects thought they would like better. However, this apparent negativity toward men was not exclusive to female gender-typed jobs; it also was evidenced in the neutral-job-type conditions. The similar pattern of findings in the female-job-type and the neutral-job-type conditions indicates that the observed negativity toward men was not really a reaction to men's gender prescription violation but rather part of a general tendency for our subjects to react more favorably to women than to men (except when the job was male in gender type). Our findings therefore suggest that the failure to act in accordance with gender-stereotypic norms does not uniquely produce social disapproval for men or, if it does, that the disapproval takes a different form than it does for women and therefore was not evident in our study. This latter point is worth noting; because a man's normative violation is not of communal oughts but of agentic oughts, the penalties that ensue may not be the same as those for women. Instead of resulting in dislike, for example, the violation may result in disrespect. This reasoning is consistent with work done by Rudman and her colleagues (Rudman, 1998; Rudman & Glick, 1999, 2001).

The first two studies demonstrated the negativity displayed toward successful women who deviate from normative prescrip-

tions and do not conform to the shoulds of gender stereotypes. But what is the effect of this negativity on outcome-oriented judgments? This was the focus of Study 3.

The objective of Study 3 was to explore the effect of being disliked on how individuals are evaluated and on the types of recommendations made about how they should be treated in work settings. The premise behind this study was that people who are disliked are at a serious disadvantage when evaluations are made and rewards distributed. Also, we contend that these negative reactions to disliked individuals result whatever the reason they are disliked. That is, men and women who are disliked should suffer similarly, although it is only women, not men, for whom work success uniquely leads to dislike.

In Study 3 we directly manipulated information about the likability of male and female employees who had recently completed a management training program, and we obtained evaluative reactions to them and recommendations for both salary and special career opportunities. Because of the nature of our dependent variables, our subjects all were individuals who work full time and therefore had experience in forming work-related judgments about others. It was our expectation that likability would affect judgments even when the reported level of an individual's competence was very high:

Hypothesis 1. Information about likability will have a significant effect on overall evaluations and reward recommendations made about both male and female employees regardless of how competent they are.

Study 3

Method

Subjects

One hundred thirty-one subjects participated in the study, of whom 86 were employees of a financial services company in the northeast and the remainder were students enrolled in a master of arts program who worked full time and attended school at night. For subjects in the corporate setting, the data were collected during company-sponsored training sessions, and for the student subjects, the data were obtained during class sessions. Both samples contained approximately equal numbers of men and women, with the overall sample containing 62 men and 65 women (4 respondents did not report their sex). The mean age of subjects was 31.4 years.

Design

The research design was a $2 \times 2 \times 2$ factorial with sex of employee (male, female), competence rating (high, low), and likability rating (high, low) as the three independent variables. Subjects were randomly assigned to a condition, resulting in 16 subjects in five conditions and 17 subjects in three conditions.

Procedure

The study was said to be part of a university-based research program investigating how people combine different sources and types of information in evaluating others in work settings. Subjects were told that they would be reviewing information about an individual who was 1 of 30 employees who had recently completed a yearlong management training program in a major consumer goods company. To ensure that the program was seen as male gender typed, we made the majority of the trainees male

(22 men and 8 women). The materials that subjects received explained that the management training program consisted of four training rotations lasting 3 months each: marketing, product development, operations, and sales. Trainees were said to have been assigned to a different department for each training rotation and to have received on-the-job training and direct supervision from the department head. During each rotation, performance and interpersonal assessments were reportedly completed for each trainee.

Subjects received a packet of materials, including a list of all the employees who were said to have participated in the training program and a background description of the particular employee whom they were to review. The background description indicated that the employee grew up in the southwestern part of the United States, attended a state university, and after four years of work in both finance and marketing, attended an eastern business school from which he or she had graduated with honors. It furthermore followed the employee's career history since receiving the master of business administration and indicated the employee's current position as an AVP supervising an operations department. Following the background description, there was a performance rating information sheet containing graphs depicting the employee's job ratings in the management training program and also a two-page questionnaire containing the dependent measures.

Subjects were instructed to complete the research questionnaire after they had reviewed all the materials in the packet. When subjects were finished, they were debriefed, the purpose of the study was explained, and they were thanked for their time.

Experimental Manipulations

Sex. The employee's sex was manipulated by the name used in the description of the employee.

Competence and likability. Employee competence and employee likability were manipulated by the information provided on the performance rating information sheet. Both variables were manipulated through bar graphs having a minimum value of 0 and a maximum value of 10. For each variable, the graph consisted of two bars, one depicting the average of the ratings of all trainees in the training program and the other depicting the rating of the particular employee being described. Thus, the graphs provided information not only about the absolute ratings of the employee but also about the ratings of the employee relative to his or her peers.

Competence ratings were said to be compilations of the assessments of all of the supervisors from the training program rotations in which the employee participated. Employees in the high-competence conditions were shown to have a competence rating of 9.1, and individuals in the low-competence conditions were shown to have a competence rating of 5.4; the average trainee rating was 6.8. Likability ratings were said to include the assessments of both supervisors and peers from the training program rotations to reflect likability in multiple contexts. The average rating of the trainee for likability was 7.1, and the employee's likability rating was 9.3 in the high-likability condition and 4.9 in the low-likability condition.

Dependent Measures

There were four key dependent measures, two reflecting evaluative reactions to the employee—overall evaluation and feelings about having the individual as one's manager—and two reflecting recommended personnel actions—recommendations for special career opportunities and salary recommendations. We obtained subjects' overall evaluations of the employee by creating a scale composed of responses to four questionnaire items (coefficient $\alpha = .89$): "Overall, how would you rate this individual?" (*very low–very high*); "How successful do you think this individual will be in this organization?" (*not at all successful–very successful*); "Rate this individual's potential to excel in his/her career" (*very low–very high*); and "To what degree do you recommend retaining this individual in the

organization?" (*not at all–very much*). We measured subjects' feelings about having the employee as their manager with the question, "How would you feel about working with this person as your manager?" (*not pleased–pleased*). Our measure of recommendation for special career opportunities was a scale composed of responses to two questionnaire items (coefficient $\alpha = .86$): "To what degree do you recommend placing this individual on the 'fast track'?" (*not at all–very much*); and "There are five highly prestigious upper-level positions available to the recent trainees. To what degree do you recommend this individual be placed in one of these five jobs?" (*not at all–very much*). Last, we obtained a recommendation for salary level by asking subjects to indicate which of five levels of potential salary they would recommend for the employee reviewed. With the exception of the salary recommendation, all questions were rated on 9-point response scales.

Results

Manipulation Checks

Question responses concerning how the employee was rated during the training program indicated that our competence and likability manipulations were effective. Subjects in the high-competence conditions rated the employee as far more competent ($M = 7.83$) than did subjects in the low-competence conditions ($M = 4.90$). Moreover, subjects indicated the employee they reviewed to be more likable in the high-likability conditions ($M = 8.11$) than in the low-likability conditions ($M = 4.16$). Both of these ratings were done on 9-point bipolar adjective scales.

Dependent Measures

A MANOVA was conducted on the four evaluation scales. Overall, the multivariate F was significant for competence, $F(4, 118) = 21.31, p < .01$, liking, $F(4, 118) = 9.40, p < .01$, and the interaction between competence and liking, $F(4, 118) = 3.92, p < .01$. Having determined these overall effects, we conducted univariate ANOVAs as well as intercell contrasts using Fisher's LSD tests (with the significance level set at $p < .05$). Correlations among the dependent measures appear in Table 6.³ Table 7 presents the condition means and standard deviations.

Analyses of variance on each of our four dependent measures scales including subject sex as a factor revealed no statistically significant main effects or interactions involving the subject sex variable; therefore, responses of male and female subjects once again were combined for all analyses.

Evaluative Reactions

Overall evaluation. An ANOVA of the overall evaluation scale indicated that in addition to a significant main effect for competence $F(1, 123) = 59.99, p < .01, \eta^2 = .32$, there was a significant main effect for liking, $F(1, 123) = 26.16, p < .01, \eta^2 = .18$. Even with the employees' competence established, likability made a difference: Across the board, those who were reported to be likable were evaluated more favorably than those who were reported to be not likable.

³ When correlations were calculated separately for male and female targets, no difference in the pattern of correlations was found.

Table 6
Study 3: Intercorrelations Between Dependent Measures

Dependent variable	1	2	3	4
1. Overall evaluation	—	.58**	.84**	.62**
2. Feelings about having as manager		—	.65**	.60**
3. Special opportunity recommendation			—	.71**
4. Salary recommendation				—

** $p < .01$.

Feelings about having the employee as a manager. An ANOVA of subjects' ratings of their feelings about the target employee as their manager revealed significant main effects for competence, $F(1, 123) = 20.21, p < .01, \eta^2 = .14$, and liking, $F(1, 123) = 27.70, p < .01, \eta^2 = .19$, and a significant interaction between them, $F(1, 123) = 5.84, p < .05, \eta^2 = .05$. Intercell comparisons, conducted for clarification, indicated that likability affected feelings about having that individual as one's manager in high-competence conditions but not in low-competence conditions.

Recommended Personnel Actions

Special career opportunities. An ANOVA of special career opportunities scale ratings revealed a significant main effect for both competence, $F(1, 123) = 81.57, p < .01, \eta^2 = .40$, and liking, $F(1, 123) = 17.35, p < .01, \eta^2 = .12$. Competent employees were more highly recommended for special opportunities than less competent employees; but even when they were competent, likable employees were more highly recommended for special opportunities than less likable employees.

Salary recommendations. The results of the ANOVA of salary recommendations yielded the same pattern of results as did the analysis of the special opportunities ratings. Results revealed a significant main effect for both competence, $F(1, 123) = 40.69, p < .01, \eta^2 = .25$, and liking, $F(1, 123) = 10.65, p < .01, \eta^2 = .08$. Not only were competent employees recommended for a higher salary than less competent employees, but likable employees, whether competent or not, were recommended for a higher salary than less likable employees.

Discussion

These results suggest that being disliked can have detrimental effects in work settings. Evidently, when making evaluations and judgments about personnel actions, competence is not the sole consideration. Negative social reactions to highly competent employees can adversely influence both overall evaluations and recommendations for how they should be treated. Thus, being disliked is likely to be not just unpleasant but also a hindrance for upwardly aspiring women.

These findings are consistent with the idea that affect may operate to bias performance ratings (Dipboye, 1985; Feldman, 1981; Ilgen & Feldman, 1983) and with research that demonstrates that liking can interfere with performance rating accuracy (Cardy & Dobbins, 1986). However, our results go further than this, demonstrating effects not only on performance ratings but also on the allocation of potential organizational rewards.

The fact that an unlikable individual is not viewed to be as worthy of salary increases or promotions as individuals deemed to be likable, and that this was found to be true regardless of whether the individual is a man or a woman, is important. But in terms of our concerns, what is most critical to remember is that whereas

Table 7
Study 3: Means and Standard Deviations for Dependent Measures

Target and condition	<i>n</i>	Evaluative reactions				Recommended personnel actions			
		Overall evaluation		Feelings about having as manager		Special career opportunities		Salary recommendation	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Competent									
Male target									
Liked	17	7.99	0.61	7.59	2.00	8.00	0.56	4.06	0.43
Not liked	16	7.13	1.21	5.31	2.58	6.41	1.95	3.56	0.96
Female target									
Liked	16	8.06	0.62	7.56	1.32	7.56	0.83	3.86	0.72
Not liked	16	6.72	1.06	4.81	1.05	6.41	1.16	3.56	0.63
Not competent									
Male target									
Liked	16	6.34	1.30	5.33	1.95	5.56	1.61	3.19	0.40
Not liked	17	5.96	1.29	4.77	1.79	4.72	1.81	2.65	0.86
Female target									
Liked	17	6.43	1.15	5.29	2.20	4.85	1.42	3.12	0.78
Not liked	16	5.34	0.94	4.00	1.55	4.34	1.35	2.75	1.00

Note. Ratings were done on 9-point scales except for the salary recommendations, for which we used a 5-point scale. The higher the means, the more favorable the ratings.

there are many things that lead an individual to be disliked, including obnoxious behavior, arrogance, stubbornness, and pettiness, it is only women, not men, for whom a unique propensity toward dislike is created by success in a nontraditional work situation. This suggests that success can create an additional impediment to women's upward mobility when they have done all the right things to move ahead in their careers.

General Discussion

The results of these studies provide support for the idea that success in traditionally male domains can have deleterious consequences for women. Information that women had been successful in these contexts resulted in them being less liked and more personally derogated as compared with equivalently successful men. Moreover, we have shown that negative feelings about successful women can have serious consequences. Being disliked was shown to strongly affect competent individuals' overall evaluations and recommended organizational rewards, including salary and special job opportunities.

Our data also are suggestive about when and why successful women are viewed negatively. We had posited that negativity toward successful women would be a penalty for their violation of gender-stereotypic norm prescriptions. We therefore expected that women's success would prompt disapproval only in situations in which the success signaled deviation from behavior deemed appropriate for them. In the second study we directly tested these ideas by varying the gender type of the job at which women were successful. The results, which demonstrated that penalties for success were exacted when the job was male gender-typed but not when it was female gender-typed or neutral in gender type, made clear that success is not in and of itself anathema for women. It is only when the success implies that gender-stereotypic norms have been violated that it induces social penalties.

It is interesting that in none of the three studies did female subjects react differently to the stimulus targets than did male subjects. This finding attests to the universality of gender-stereotypic norms and of the tendency to penalize individuals who violate them and is not at all consistent with the view that women are less biased than men in rating other women. Contrary to the results of some recent studies (Carli, LaFleur, & Loeber, 1995; Rudman, 1998), we found absolutely no indication that women and men raters differed in their leniency in enforcing gender-based norms or that they used different standards in determining whether a woman's behavior was gender inconsistent.

It should be recognized that two of our studies used undergraduates as subjects, and although they were more likely than many undergraduates to have had work experience, the nature of their work experience may limit the generalizations that we can make from our data. Moreover, our studies all used a paper-and-pencil format, did not necessitate actual interaction between the subjects and those whom they were rating, and sometimes used single-item measures. These, too, are potential limitations of the research. Also, the information provided about the stimulus managers was purposely sparse and uncomplicated to enable us to clearly and precisely manipulate our independent variables. But this sparseness may have facilitated the use of stereotypes because of the absence of individuating information. This all strongly suggests that our ideas should be tested in ongoing settings, perhaps by

obtaining ratings of women and men in different types of job categories and industries who are widely known to be successful.

Nonetheless, the importance of our results for understanding how women are regarded when they succeed at male gender-typed tasks should not be minimized. There are many instances in which those who work in organizations have distinct impressions of others with whom they have never worked or even met—impressions based on inference and word of mouth but not experience. Also, even when impressions of women in organizations are based on information, the information available often is not particularly detailed or elaborated. Despite this, however, there is a strong need for research that validates and extends our findings in actual work contexts if researchers are to fully understand the effect of success on working women.

Our results imply that success in nontraditional areas is double-edged for women. When acknowledged as successful they no longer are saddled with the image of being incompetent, but they may also pay a price. The price is social rejection, taking the form of both dislike and personal derogation, and it appears to have definite consequences for evaluation and recommendations about reward allocation. The highly negative interpersonal characterizations concerning orientation toward others that we uncovered here have analogues in the work world, where terms such as *bitch*, *ice queen*, *iron maiden*, and *dragon lady* are invoked to describe women who have successfully climbed the organizational ladder. Moreover, these characterizations provide some insight into why, despite their success, high-powered women often tend not to advance to the very top levels of organizations. As with Ann Hopkins, whose denial of partner status at Price Waterhouse was eventually reviewed by the U.S. Supreme Court (Price Waterhouse v. Hopkins, 1989), women may well be applauded as competent and accomplished but may also be seen as personally abhorrent. The results of Study 3 suggest that such individuals are unlikely to be viewed as suitable additions to the upper management team, whatever their skills and abilities.

Future research is needed to identify whether there are conditions under which success at traditionally male jobs does not have the detrimental consequences for women that we demonstrated in our investigations. It is not clear, for example, whether being seen as agentic and being viewed as communal are antithetical or if there are conditions under which both can exist in perceptions of an individual. If such coexistence is possible, then a woman's success in traditionally male domains need not always be a violation of her "shoulds," and perhaps negative reactions need not result. Moreover, it is not clear whether the interpersonal hostility perceptions that we found to result from a woman's success are part of a larger subset of characteristics associated with successful women, nor is it clear whether these perceptions are a result or a cause of the dislike directed at them. Finally, in light of recent work done by others (Berdahl, Magley, & Waldo, 1996; Stockdale, Visio, & Batra, 1999; Waldo, Berdahl, & Fitzgerald, 1998), it would be interesting to further examine men's gender-stereotypic norm violation and the reactions it provokes. Research addressing these questions is currently underway.

References

- Bartol, K. M., & Butterfield, D. A. (1976). Sex effects in evaluating leaders. *Journal of Applied Psychology*, 61, 446-454.

- Berdahl, J. L., Magley, V. J., & Waldo, C. R. (1996). The sexual harassment of men? *Psychology of Women Quarterly*, 20, 527–547.
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5, 665–692.
- 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.
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672–678.
- Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology*, 59, 941–951.
- Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology*, 68, 1030–1041.
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 151–192). Boston: McGraw-Hill.
- Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. *Academy of Management Review*, 10, 116–127.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15, 543–558.
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly*, 15, 203–216.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127–148.
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1993). A brief rejoinder: Accuracy and objectivity on behalf of the APA. *American Psychologist*, 48, 55–56.
- Hagen, R. L., & Kahn, A. (1975). Discrimination against competent women. *Journal of Applied Social Psychology*, 5, 362–376.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior*, Vol. V (pp. 269–298). Greenwich, CT: JAI Press.
- Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality*, 10, 3–26.
- Heilman, M. E. (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. (1995). Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality*, 10, 237–252.
- Heilman, M. E., Block, C., Martell, R., & Simon, M. (1989). Has anything changed?: Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74, 935–942.
- Horner, M. (1970). Femininity and successful achievement: A basic inconsistency. In J. Bardwick, E. M. Douvan, M. S. Horner, & D. Gutmann (Eds.), *Feminine personality and conflict* (pp. 45–76). Belmont, CA: Brooks/Cole.
- Horner, M. (1972). Toward an understanding of achievement-related conflicts in women. *Journal of Social Issues*, 28, 157–175.
- Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. *Research in Organizational Behavior*, 5, 141–197.
- Jago, A. G., & Vroom, V. H. (1982). Sex differences in the incidence and evaluation of participative leader behavior. *Journal of Applied Psychology*, 67, 776–783.
- Lyness, K. S., & Judiesch, M. K. (1999). Are women more likely to be hired or promoted into management positions? *Journal of Vocational Behavior*, 54, 158–173.
- Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. *Academy of Management Review*, 5, 267–276.
- Porter, N., & Geis, F. L. (1981). Women and nonverbal leadership cues: When seeing is not believing. In C. Mayo & N. Henley (Eds.), *Gender and nonverbal behavior* (pp. 39–61). New York: Springer-Verlag.
- Price Waterhouse v. Hopkins, 490 U.S. 228 (1989).
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, 74, 629–645.
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, 77, 1004–1010.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57, 743–762.
- Stockdale, M. S., Visio, M., & Batra, L. (1999). The sexual harassment of men: Evidence for a broader theory of sexual harassment and sex discrimination. *Psychology, Public Policy, and Law*, 5, 630–664.
- Swim, J. K., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105, 409–429.
- Tosi, H. L., & Einbinder, 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.
- Waldo, C. R., Berdahl, J. L., & Fitzgerald, L. F. (1998). Are men sexually harassed? If so, by whom? *Law and Human Behavior*, 22, 59–79.
- Wiley, M. G., & Eskilson, A. (1985). Speech style, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles*, 12, 993–1007.

Received July 9, 2002

Revision received May 13, 2003

Accepted July 7, 2003 ■