

# Same Behavior, Different Consequences: Reactions to Men's and Women's Altruistic Citizenship Behavior

Madeline E. Heilman and Julie J. Chen  
New York University

In 2 experimental studies, the authors hypothesized that the performance of altruistic citizenship behavior in a work setting would enhance the favorability of men's (but not women's) evaluations and recommendations, whereas the withholding of altruistic citizenship behavior would diminish the favorability of women's (but not men's) evaluations and recommendations. Results supported the authors' predictions. Together with the results of a 3rd study demonstrating that work-related altruism is thought to be less optional for women than for men, these results suggest that gender-stereotypic prescriptions regarding how men and women should behave result in different evaluative reactions to the same altruistic behavior, depending on the performer's sex.

Stereotypes about women are prescriptive as well as descriptive. Not only do they specify what women are like, but they also dictate norms about how women should behave (Burgess & Borgida, 1999; Eagly & Karau, 2002; Heilman, 2001). The studies reported here examine how these gender role prescriptions can affect reactions to women as compared with reactions to men in work settings. Our particular focus is on altruistic citizenship behavior. Specifically, we propose that the same act of work-related altruism will prompt different evaluations and recommendations, depending on the sex of the performer.

Altruism has been identified as one of several dimensions of organizational citizenship behavior—behavior that involves prosocial activity in the workplace that increases the effective functioning of the organization (Organ, 1988). Organizational citizenship behaviors are not part of an individual's formal job duties or responsibilities, and they often are perceived by those in the workplace to be both voluntary and discretionary. Although they are not explicitly specified in the formal job description, organizational citizenship behaviors have frequently been shown to favorably affect individuals' performance evaluations and the determination of their deservingness for organizational rewards (Allen & Rush, 1998; Mackenzie, Podsakoff, & Fetter, 1991, 1993; Motowidlo & Van Scotter, 1994; Werner, 1994) and to play a significant role in decisions that impact career advancement and success (Podsakoff, Mackenzie, Paine, & Bachrach, 2000).<sup>1</sup>

Altruism (called "helping behavior" by some researchers) is a central dimension of organizational citizenship behavior (e.g., Organ, 1988; Podsakoff et al., 2000). It involves assisting others with organizationally relevant tasks—going the "extra mile" to help others out with a work-related problem. Being a helper is central to female gender stereotype prescriptions, which dictate

that women be nurturing and socially oriented (communal) rather than competitive and achievement oriented (agentic) (Eagly & Mladinic, 1989; Eagly & Steffen, 1984; Heilman, 2001). Behaviors deemed appropriate for women therefore involve supportiveness and being concerned about the well-being of others. These prescribed behaviors not only provide a blueprint for action for women themselves but also create expectations on the part of others about how women are likely to behave. They therefore are likely to lead to the expectation that, when given the opportunity, women will engage in altruistic citizenship behavior, not avoid it.

Violations of normative role prescriptions tend to be penalized (Cialdini & Trost, 1998), and violations of women's gender role prescriptions in work settings are no exception. There is evidence that women are evaluated unfavorably as compared with men when they engage in stereotypic male behaviors such as using autocratic or directive leadership styles (Butler & Geis, 1990; Eagly, Makhijani, & Klonsky, 1992), presenting themselves in a self-promoting manner (Rudman, 1998), using a task-oriented nonverbal style (Carli, LaFleur, & Loeber, 1995), or simply being successful managers (Eagly & Johnson, 1990; Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman, Block & Martell, 1995; Heilman, Block, Martell, & Simon, 1989). But women behave inconsistently with stereotype-based gender role prescriptions not only when they behave "like men" but also when they fail to behave "as women should." Therefore, not engaging in altruistic citizenship behavior, because it violates the female prescription to be communal, is also likely to prove costly for women in work settings.

<sup>1</sup> The fact that organizational rewards have been found to be associated with citizenship behavior has raised questions about whether it truly is discretionary and should be classified as extrarole behavior or contextual performance (Borman & Motowidlo, 1993; Organ, 1997; Van Dyne, Cummings, & McLean Parks, 1995). However, the classification of citizenship behavior is peripheral to our concern. Rather, the ideas presented here are concerned with *perceptions* of citizenship behavior as discretionary because they are not part of the formal job expectations.

---

Madeline E. Heilman and Julie J. Chen, Department of Psychology, New York University.

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

Moreover, even if women behave consistently with female role prescriptions and engage in altruistic citizenship behaviors, they are not likely to induce reactions equally favorable as reactions to men who engage in the very same behaviors. Because such prescription-consistent actions are routinely expected for women, they are not apt to be regarded as particularly noteworthy; thus, when they occur, they are likely to be disregarded or ignored. Consequently, work behavior that is applauded in men may not result in equivalent recognition or acclaim for women when the behavior is, as the case with altruistic citizenship behavior, one that is directly specified by the female gender role prescription. In such instances, women are unlikely to be as highly regarded or as highly rewarded as men who engage in the identical behavior.

According to our reasoning, then, women who fail to perform altruistic citizenship behaviors when the opportunity arises are likely to provoke strong negative reactions—they are behaving contrary to the specifications of the female role prescription. This stands in contrast to men whose similar failure to act altruistically is unlikely to be of great consequence because being helpful is not normatively required of them. Furthermore, women who *do* perform altruistic citizenship behaviors are unlikely to be noticed, or, if they are, their actions may not be considered to be anything special, whereas men's altruistic behavior is more likely to stand out and appear to be exceptional. So, women are likely to benefit less than men from performing altruistic citizenship behaviors and are likely to be penalized more than men for their failure to act altruistically.

In the following study, undergraduate students were asked to review the performance profile of an employee and then evaluate his or her performance and make recommendations about organizational rewards. The employees were either men or women and were depicted as having chosen to engage or having chosen not to engage in altruistic citizenship behavior when confronted with the opportunity to do so. Also included for purposes of control was a condition in which no information about altruistic citizenship behavior was presented. We expected that women, who are expected to fulfill their prescribed role behaviors by being helpful, would not benefit when they engaged in altruistic citizenship behavior but would be harshly treated when they did not. In contrast, we expected that men, who are not normatively required to be helpful, would benefit when they engaged in altruistic citizenship behavior but would suffer no penalty when they did not. We therefore hypothesized the following:

*Hypothesis 1.* Engaging in altruistic citizenship behavior will enhance men's performance evaluations and reward recommendations but will not affect those of women.

*Hypothesis 2.* Withholding altruistic citizenship behavior will be detrimental to women's performance evaluations and reward recommendations but will not affect those of men.

In addition, we collected participants' perceptions of the employee's competence and interpersonal civility to control for their possible effects on our dependent variable measures.

## Study 1

### Method

#### Participants

One hundred thirty-five male and female undergraduate students enrolled in introductory psychology courses at a large northeastern university participated in this study in exchange for course credit. Their mean age was 19.9 years ( $SD = 3.08$ ). Two thirds of the participants were women.

#### Design

The study was a  $2 \times 3$  factorial between-subjects design, with the independent variables being the performance of altruistic citizenship behavior (performed, not performed, and no information about performance) and sex of the target being rated. Participants were randomly assigned to the six conditions, with 22 participants in each of the male target conditions and 23 participants in each of the female target conditions.

#### Procedure

Each participant was given a folder of research materials, the first page of which was a cover sheet explaining the purpose of the study. The study was said to be focused on performance evaluation methods and, in particular, was concerned with examining the effectiveness of 360-degree feedback procedures. We provided some background information about the 360° feedback process and briefly described its purpose and advantages as well as the procedures used to implement it. Participants were told that during the research session, they would be reviewing an employee's performance feedback materials and that the materials they would see had been completed by either the employee's supervisor, coworker, or subordinate. Unknown to them, all participants received performance feedback materials that had been completed by the employee's coworker. We selected the coworker to be the information source because we thought it most plausible that the coworker would have had experiences relevant to altruistic behavior and would choose to share them.

The next page, labeled *Employee Information Form*, contained background information about the employee's work history with the company. This information included the employee's work department name (Purchasing) and job title (Level III, Administrator). This job was chosen to be gender neutral—not particularly male or female in gender type, and pre-testing verified that this was the case. Also included was the employee's starting date, indicating length of tenure in the company (5 years) and tenure in the current position (4 years). Also included was a photograph of the employee's four-person work group, consisting of two women and two men of approximately the same age (late 20s), dressed in professional attire; a red arrow designated the employee who was to be rated.

Participants were then presented with a feedback rating form, ostensibly completed by the employee's coworker. The first section of the form contained a list of various work skills and attributes, on which the employee was rated for proficiency on a 5-point scale ranging from 1 (*poor*) to 5 (*excellent*). These skills and attributes included planning and organizing, follow through, dependability, efficiency, accuracy, accepting of responsibility, versatility, capacity to work, and emotional stability. The employee was rated "good" on seven of the scales and "excellent" on the two others. The second section of the feedback form contained the coworker's open-ended report of episodes that "characterize the employee's typical work behavior."

The final portion of the materials was a questionnaire that asked participants to evaluate the employee's job performance and make recommendations for organizational rewards. It also asked for attribute ratings of the employee. After completing the questionnaire, participants were fully debriefed and the purpose of the study explained.

### Independent Variable Manipulations

*Sex of target employee.* The sex of the employee to be rated was varied by the name and gender-relevant adjectival references on the employee information form and on the feedback materials. The designation of the target employee in the photograph was meant to further support this manipulation. Pretesting was done, using the same population from which our participant sample was drawn, to ensure that the male and female targets were perceived to be equivalent in age, intelligence, friendliness, cheerfulness, and professionalism. Testing done subsequent to our data collection also showed the male and female target to be equivalent in attractiveness.

*Altruistic citizenship behavior.* Information about the performance or nonperformance of altruistic citizenship behavior was provided on the "episode" portion of the coworker's feedback rating form, which contained brief reports describing experiences with the employee that were meant to be revealing about the employee's typical pattern of behavior. First, there was an episode reported, appearing in all experimental conditions, recounting events that had occurred at a stressful meeting that had revealed the employee's alertness and conscientiousness. In the condition in which no information about performing altruistic citizenship behavior was provided, participants read only the report of this first episode.

Participants in conditions in which the target employee either performed or did not perform the altruistic citizenship behavior read a report of a second episode, also thought to be provided by the target employee's coworker. In creating this episode, efforts were made to depict a situation that would be relevant and meaningful to our undergraduate research participants:

Once I was in a panic because I had to make copies of some presentation materials for an important meeting the next morning. The copy machine broke down on me and would not collate or staple the pages. It was 5:15 and all the support staff was gone, and everyone else was preparing to go out for another coworker's birthday dinner. We'd all been looking forward to it. I ran around looking for help to manually collate and staple the 500 pages.

What followed differed, depending on the experimental condition. Participants in the condition in which the target employee had performed the altruistic behavior read, "When Cathy (Kevin) learned what had happened, she (he) immediately volunteered to help me even though she (he) would miss part of the dinner."

In contrast, participants in the condition in which the target employee did not perform the altruistic behavior read, "When Cathy (Kevin) learned what had happened, she (he) said she (he) could not help me because she (he) was on her (his) way to the party but suggested I try to find a copy shop that was still open."

Immediately following these statements, to reinforce the idea that the behavior depicted was typical of the employee's behavior in general, was the sentence "That's the way Cathy (Kevin) is."

### Dependent Measures

There were two primary dependent variables, performance evaluation and reward recommendations. In addition, several attribute measures were included. Scales were constructed for each measure.

*Performance evaluation.* Our measure of performance evaluation was a composite scale consisting of the following three items: "Overall, how would you rate this employee's performance over the past year?" "In your opinion, how likely is it that this employee will advance in the company?" and "Give your assessment of the individual's likelihood of success." Each item was measured on a 7-point response scale ranging from 1 (*excellent*) to 7 (*average*) (reverse coded) for the first item and from 1 (*very unlikely*) to 7 (*very likely*) for the other two items. The three items combined to form a scale with a reliability of  $\alpha = .82$ .

*Reward recommendations.* Following Allen and Rush (1998), we created a reward recommendation scale by asking participants to give their recommendations for four types of common organizational rewards (salary increase, promotion, high-profile project, and bonus pay). Although we used the same measure, the response scales were slightly modified for the present study so they would have seven points, ranging from 1 (*would definitely recommend*) to 7 (*would definitely not recommend*). Allen and Rush (1998) reported a reliability of  $\alpha = .90$ ; the scale used for the present study produced a reliability of  $\alpha = .88$ .

*Attribute ratings.* Participants were asked to rate the employee on items comprising two attribute scales: competence and interpersonal civility. Ratings on four adjectives were combined to form an overall measure of competence: competent, productive, effective, and decisive ( $\alpha = .80$ ). Ratings on three adjectives were combined to form a measure of interpersonal civility: nasty, selfish, and manipulative (all reverse coded), with the reliability for the scale being  $\alpha = .79$ . 7-point response scales were used for all adjective ratings ranging from 1 (*very little*) to 7 (*very much*).

Correlations among the dependent variable measures appear in Table 1.

## Results

### Manipulation Checks

To check on our manipulation of altruistic citizenship behavior, we created a measure of the employee's helpfulness from ratings on three attributes: helpful, caring, and sympathetic ( $\alpha = .92$ ). An analysis of variance (ANOVA) verified that the altruism manipulation was successful. We found a main effect for altruism,  $F(2, 128) = 103.93, p < .001, \eta^2 = .62$ . Follow-up *t* tests indicated that participants in the condition in which the target performed the helping behavior ( $M = 6.29$ ) rated the target as significantly more helpful compared with participants in the condition in which the target did not perform the helping behavior ( $M = 3.39$ ), and ratings in each of these conditions differed from those in which no information about helping behavior was provided ( $M = 5.54$ ).

### Dependent Measures

Preliminary ANOVA analyses, including participant sex as an additional independent variable, indicated that the sex of the participant had no significant main effects, nor did it interact with either of the study's independent variables in responses on the primary outcome measures of overall performance and reward recommendation ratings, nor on ratings of the target's competence. Moreover, although there was a significant main effect indicating that female participants generally rated targets more favorably in terms of interpersonal civility compared with male participants,  $F(1, 123) = 11.24, p < .01$ , there were no significant interactions with participant sex in ratings of interpersonal civility. We there-

Table 1  
Means, Standard Deviations, and Correlations Among  
Dependent Variable Measures: Study I

Dependent variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. Performance evaluation	5.07	1.10	—	.66**	.38**	.25**
2. Reward recommendations	5.11	0.91		—	.55**	.35**
3. Competence	5.78	0.72			—	.24**
4. Interpersonal civility	5.48	1.12				—

\*\*  $p < .01$ .

fore combined responses across male and female participants in all subsequent analyses.

We conducted a multivariate analysis of variance on the two primary dependent measures—performance evaluation ratings and reward recommendation ratings. The multiple *F* was significant for target sex,  $F(2, 128) = 3.30, p < .05$ , for altruism condition,  $F(4, 258) = 6.06, p < .001$ , and for the predicted Target Sex  $\times$  Altruism interaction,  $F(4, 258) = 3.03, p < .05$ . We then conducted univariate ANOVAs and, to test our hypotheses directly, intercell contrasts. We also conducted ANOVAs on the two attribute rating scales, and they subsequently were taken as covariates to control for their effects on both performance evaluations and reward recommendations. We tested all intercell contrasts with Fisher’s least significant difference (LSD) method, with the significance level set at  $p < .05$ . Table 2 presents the relevant means and standard deviations.

*Performance Evaluation.* A  $2 \times 3$  ANOVA revealed a significant main effect for target sex,  $F(1, 129) = 5.58, p < .05, \eta^2 = .04$ , as well as a significant main effect for altruism condition,  $F(2, 129) = 5.25, p < .01, \eta^2 = .08$ . Also, consistent with our hypotheses, there was a significant Target Sex  $\times$  Altruism interaction,  $F(2, 129) = 4.80, p < .05, \eta^2 = .07$ .

We conducted intercell contrasts to clarify the interaction effect. As predicted by our hypotheses, the tests revealed a difference in how women and men were evaluated when they did and did not perform the helping behavior that was requested. Specifically, when they did not help, the man’s job performance was rated no differently, but the woman’s performance was rated significantly lower than when no information about helping was provided. In contrast, when they did perform the helping behavior, the man’s performance ratings were significantly higher, and the woman’s performance ratings were no different than when no information about helping was provided.

Because of the gender-neutral nature of the stimulus job, we had not anticipated a difference in the ratings of the male and female

target in the no-information condition, and none was found. However, there was a significant difference in the performance ratings of the male and female target both when the helping behavior was performed and when it was withheld. As our ideas would suggest, regardless of what they reportedly chose to do, women, when in a situation in which altruistic citizenship behavior was at issue, always were rated more negatively than were men.

*Reward recommendations.* An ANOVA revealed significant main effects for target sex,  $F(1, 129) = 4.95, p < .05, \eta^2 = .04$ , and altruism,  $F(2, 129) = 13.32, p < .001, \eta^2 = .17$ , and a significant interaction between the two,  $F(2, 129) = 5.22, p < .01, \eta^2 = .08$ . Fisher’s LSD tests were conducted to clarify the interaction; the pattern of results that emerged for reward recommendations was the same as that found for the performance evaluations.

Again, as predicted, the performing and not performing of altruistic citizenship behavior had different consequences for the ratings given to men and women. When they did not engage in helping behavior, there were no different recommendations made about the male targets but significantly more negative recommendations made about the female targets, as compared with targets about whom no information about helping was available. In contrast, when the targets did engage in helping behavior, men were given significantly more favorable recommendations, whereas women were given no different recommendations than those about whom no information about helping was available.

Analogous to the results for evaluations of job performance, although the difference in ratings between the male and female targets in the no-information condition was not significant, the male target was more highly recommended for organizational rewards compared with the female target both when the altruistic citizenship behavior was provided and when it was withheld.

*Attribute ratings.* An ANOVA of the competence scale revealed no significant main effects for either target sex,  $F(1, 29) = 0.16, ns$ , or altruism condition,  $F(2, 129) = 1.93, ns$ . There also

Table 2  
Means and Standard Deviations: Study 1

Condition	Performance evaluation		Reward recommendations		Competence		Interpersonal civility	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Male target								
Performs helping behavior	5.85 <sub>a</sub>	1.09	5.84 <sub>a</sub>	0.77	6.02 <sub>a</sub>	0.69	6.29 <sub>a</sub>	0.74
Does not perform helping behavior	5.06 <sub>b</sub>	1.28	4.94 <sub>b</sub>	0.78	5.59 <sub>a</sub>	0.86	4.77 <sub>b</sub>	1.08
No helping information provided	4.94 <sub>b,c</sub>	1.21	5.03 <sub>b</sub>	0.89	5.81 <sub>a</sub>	0.64	5.64 <sub>c</sub>	0.86
Female target								
Performs helping behavior	4.96 <sub>b</sub>	0.81	5.18 <sub>b</sub>	0.82	5.74 <sub>a</sub>	0.77	6.22 <sub>a</sub>	0.90
Does not perform helping behavior	4.35 <sub>c</sub>	0.70	4.35 <sub>c</sub>	0.81	5.63 <sub>a</sub>	0.67	4.64 <sub>b</sub>	1.25
No helping information provided	5.29 <sub>b</sub>	0.97	5.36 <sub>b</sub>	0.78	5.90 <sub>a</sub>	0.66	5.36 <sub>c</sub>	0.61

*Note.* The higher the mean, the higher the performance evaluation, the higher the reward recommendations, the more competent, and the more interpersonally civil the target was rated. Ratings were done on 7-point scales, with  $n = 22$  for each of the male target conditions and  $n = 23$  for each of the female target conditions. Means within a column with different subscripts differ significantly at  $p < .05$ , as indicated by the Fisher’s least significant difference procedure.

was no significant interaction effect,  $F(2, 129) = 0.91$ , *ns*; all targets were rated as being equally competent. However, the analysis of the interpersonal civility data revealed a main effect for altruism,  $F(2, 129) = 30.96$ ,  $p < .001$ ,  $\eta^2 = .32$ , such that targets who performed the helping behavior were rated more favorably than targets for whom no information about helping was given, and both were rated more favorably than targets who chose not to help. There was no significant main effect for target sex,  $F(1, 129) = 0.99$ , *ns*, or for the two-way interaction,  $F(2, 129) = 0.14$ , *ns*.

*Covariance analyses.* Analyses of covariance (ANCOVAs) were conducted to determine whether the hypothesized interaction between our independent variables would remain statistically significant when either competence ratings or interpersonal civility ratings were covaried out. Results indicated that, when the competence ratings were taken as a covariate, the interaction effect for both the performance evaluation measure,  $F(2, 128) = 4.35$ ,  $p < .05$ , and for the reward recommendation measure,  $F(2, 128) = 5.49$ ,  $p < .01$ , remained significant. A parallel set of ANCOVAs were also conducted using interpersonal civility ratings as a covariate. Results indicated that, for the reward recommendation measure, the interaction effect remained significant,  $F(2, 128) = 5.66$ ,  $p < .01$ , even after interpersonal civility was covaried out. For the performance evaluation measure, interpersonal civility as a covariate was not significant, so no further analysis was conducted. Thus, taking the competence ratings or the interpersonal civility ratings as covariates had virtually no effect on the statistical significance of the original interaction effect in the ANOVA.

### Summary of Results

These results strongly support our hypotheses, indicating a differential reaction to women and men when they have, or have not, performed altruistic citizenship behaviors in a work setting. In very clear terms, women were shown to be judged more negatively than men whatever they did with respect to helping behavior: When they helped, they were not awarded the high regard bestowed upon men, and when they did not help, only they, not men, paid the price in terms of performance evaluations and reward recommendations.

Compelling as these findings are, the nature of our research participants in Study 1 may limit their generalizability. Extrapolating from the behavior of college students to the behavior of those who are organizational employees is highly risky because of the different frames of reference they bring to the research session and the potentially different value they place on the process of performance evaluation and organizational reward allocation. For that reason, we decided to replicate our study with a sample of research participants who, although students, were older than the undergraduates in Study 1 and were all currently working full-time in organizations. The hypotheses were the same as those for Study 1.

## Study 2

### Method

#### Participants

Ninety-nine Master of Business Administration (MBA) and Master of Arts (MA) students enrolled in business and psychology courses at two

large northeastern universities participated in the study as part of a class exercise. Of the participants, 58 were male and 41 were female, with a mean age of 26.7 years ( $SD = 3.17$ ). Participants had an average of 5.5 years of work experience ( $SD = 3.75$ ), with 79.6% having had managerial experience. The participants with managerial experience had an average of 3.2 years of experience ( $SD = 2.79$ ).

### Design

As in Study 1, the design was a  $2 \times 3$  factorial between-subjects design, with the independent variables being the performance of altruistic citizenship behavior (performed, not performed, and no information about performance) and sex of the target being rated. Participants were randomly assigned to the six conditions, resulting in 16 participants in each of the male target conditions and 17 participants in each of the female target conditions.

### Procedure

The procedure for Study 2 was identical to that for Study 1. However, the episode used to establish the altruism manipulation was different than the one used in the former study, although the key elements were the same. We made this change to increase the sophistication of the situation depicted because of the more work-savvy participants in our sample. The episode read,

Once I was preparing materials for an important presentation I was giving the next morning. My computer had been infected with a virus earlier in the week, and I had lost several of my presentation files. I spent the rest of the week recreating those files and was faced with a lot of work to do at the last minute, the most urgent of which involved incorporating the week's expenses and earnings into my presentation. I was in a panic because I still had to retrieve the financial reports from each department, and I knew I would never have time to double check that all the figures were correct before creating my concluding presentation slides. It was 7:30 and all the support staff was gone, and everyone else was preparing to go out for another coworker's birthday dinner. We'd all been looking forward to it. I ran around looking for help to retrieve the reports and check the figures.

As with Study 1, there were different endings of this episode, depending on the altruism condition. In the conditions in which the altruistic behavior occurred, it read, "When Cathy (Kevin) learned what had happened, she (he) immediately volunteered to help me out even though she (he) would miss part of the dinner."

In the conditions in which the target did not perform the altruistic behavior, the ending read, "When Cathy (Kevin) learned what had happened, she (he) said she (he) could not help me because she (he) was on her (his) way to the party but suggested I try to find someone in Accounting who was still there."

As in Study 1, the episode ended with a final sentence meant to reinforce the idea that the episode reported was one that depicted behavior typical of the employee: "That's the way Cathy (Kevin) is."

### Dependent Measures

As in Study 1, the key dependent variables were performance evaluation and reward recommendations, and we also collected measures of competence and interpersonal civility. The scales for each measure in Study 2 were identical to those constructed for Study 1. The reliability was  $\alpha = .74$  for the performance evaluation scale,  $\alpha = .91$  for reward recommendations,  $\alpha = .82$  for competence, and  $\alpha = .88$  for interpersonal civility. Correlations among the dependent variable measures appear in Table 3.

Table 3  
Means, Standard Deviations, and Correlations Among  
Dependent Variable Measures: Study 2

Dependent variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. Performance evaluation	5.09	0.80	—	.76**	.46**	.54**
2. Reward recommendations	4.72	0.92		—	.45**	.43**
3. Competence	5.53	0.67			—	.32**
4. Interpersonal civility	5.02	1.28				—

\*\*  $p < .01$ .

## Results

### Manipulation Checks

An ANOVA verified that the helping manipulation was successful. Helpfulness, measured by combining the ratings on the helpful, caring, and sympathetic attribute scales ( $\alpha = .93$ ), indicated a main effect for altruism,  $F(2, 93) = 217.38, p < .001, \eta^2 = .82$ . Follow-up  $t$  tests indicated that participants in the condition in which the target helped ( $M = 6.43$ ) rated the target as significantly more helpful compared with participants in the condition in which the target did not help ( $M = 3.18$ ), and ratings in each of these conditions differed from those in which no information about helping behavior was provided ( $M = 4.67$ ).

### Dependent Measures

Preliminary ANOVAs, including participant sex as a third independent variable, revealed no significant main effects or interactions in the responses of male and female participants for any of the outcome measures. Therefore, as in Study 1, we combined responses across male and female participants in all subsequent analyses.

Our strategy for data analysis was the same as that used in Study 1. Results of the multivariate analysis of variance conducted

on the performance evaluation ratings and the reward recommendation ratings revealed a multiple  $F$  significant for target sex,  $F(2, 92) = 3.13, p < .05$ ; altruism,  $F(4, 186) = 8.07, p < .001$ ; and, most importantly, for the Target Sex  $\times$  Altruism interaction, predicted by our hypotheses,  $F(4, 186) = 3.90, p < .01$ . Thus, we then conducted univariate ANOVAs as well as intercell contrasts to test our specific hypotheses. Additionally, we conducted ANOVAs on the competence and interpersonal civility ratings and subsequently included them as covariates in ANCOVAs. We tested all intercell contrasts with Fisher's LSD method, with the significance level set at  $p < .05$ . Table 4 presents the relevant means and standard deviations.

*Performance evaluation.* An ANOVA revealed a significant main effect for target sex,  $F(1, 93) = 5.77, p < .05, \eta^2 = .06$ , as well as a significant main effect for altruism,  $F(2, 93) = 19.21, p < .001, \eta^2 = .29$ . There was also a significant Target Sex  $\times$  Altruism interaction,  $F(2, 93) = 8.17, p < .001, \eta^2 = .15$ .

Intercell contrasts were conducted to clarify the interaction effect. Consistent with our hypotheses, and similar to the results for Study 1, the tests revealed a difference in how women and men were evaluated when they did and did not provide the help that was requested. Specifically, when they did not help, the man's job performance was rated no differently, but the woman's performance was rated significantly lower than when there was no information given about helping. In contrast, when they did help, the man's performance ratings were significantly higher, and the woman's performance ratings were no different than when there was no information given about helping.

As in Study 1, there was no significant difference in the ratings of the male and female target in the no-information condition, but there was a significant difference in the performance ratings of the male and female target both when helping was provided and when it was withheld. Thus, regardless of what they reportedly chose to do, women were always rated more negatively than were men when in a situation in which helping was at issue.

Table 4  
Means and Standard Deviations: Study 2

Condition	Performance evaluation		Reward recommendations		Competence		Interpersonal civility	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Male target								
Performs helping behavior	5.98 <sub>a</sub>	0.63	5.55 <sub>a</sub>	0.77	5.69 <sub>a</sub>	0.62	6.38 <sub>a</sub>	0.48
Does not perform helping behavior	4.94 <sub>b</sub>	0.52	4.53 <sub>b,c</sub>	0.81	5.25 <sub>a</sub>	0.66	4.00 <sub>b</sub>	0.83
No helping information provided	4.83 <sub>b</sub>	0.89	4.64 <sub>b</sub>	0.77	5.73 <sub>a</sub>	0.92	5.06 <sub>c</sub>	0.84
Female target								
Performs helping behavior	5.24 <sub>b</sub>	0.55	4.79 <sub>b</sub>	0.80	5.74 <sub>a</sub>	0.69	6.29 <sub>a</sub>	0.51
Does not perform helping behavior	4.33 <sub>c</sub>	0.55	4.03 <sub>c</sub>	0.85	5.40 <sub>a</sub>	0.35	3.41 <sub>b</sub>	0.58
No helping information provided	5.25 <sub>b</sub>	0.64	4.81 <sub>b</sub>	0.94	5.38 <sub>a</sub>	0.63	5.02 <sub>c</sub>	0.75

*Note.* The higher the mean, the more favorable the rating. Ratings were done on 7-point scales, with  $n = 16$  for each of the male target conditions and  $n = 17$  for each of the female target conditions. Means within a column with different subscripts differ significantly at  $p < .05$ , as indicated by the Fisher's least significant difference procedure.

*Reward recommendations.* An ANOVA revealed significant main effects for target sex,  $F(1, 93) = 4.74, p < .05, \eta^2 = .05$ , and altruism,  $F(2, 93) = 9.55, p < .001, \eta^2 = .17$ . The interaction effect did not reach standard levels of significance,  $F(2, 93) = 2.73, ns$ . However, given the strong trend indicated ( $p = .07$ ), and the consistency of the means with the predicted pattern of data obtained both in Study 1 and in the analysis of the performance evaluations in Study 2 (see Table 4), we conducted intercell contrasts to test our specific hypotheses. The results were similar to those found for the performance ratings.

Once again, the performing and not performing of altruistic behavior had different consequences for the ratings given to men and women. When they did not help, there were no different recommendations made about the male targets but significantly more negative recommendations made about female targets, as compared with targets about whom no information about helping behavior was available. Also as expected, men who helped were given higher recommendations, whereas women who helped were given no different recommendations than those about whom no information about helping was available.

There again was a tendency to rate men more highly than women whether they were said to have engaged or to have not engaged in altruistic behavior. Specifically, the reward recommendation ratings were significantly higher for men than for women when the requested help was said to have been provided, and, although the difference did not meet conventional significance levels, ratings tended to be higher for men than for women when the requested help was said to have been withheld. There was, however, no evidence of a difference between the reward recommendation ratings given to male and female targets when no information about altruistic citizenship behavior was provided.

*Attribute ratings.* An ANOVA of the competence scale revealed no significant main effects for either target sex,  $F(1, 93) = 0.15, ns$ , or altruism condition,  $F(2, 93) = 2.86, ns$ . There also was no significant interaction effect,  $F(2, 93) = 1.31, ns$ ; all targets were rated as being equally competent. However, the analysis of the interpersonal civility data revealed a main effect for altruism,  $F(2, 93) = 122.89, p < .001, \eta^2 = .73$ , such that targets who helped were rated more favorably than targets for whom no information about helping was given, and both were rated more favorably than targets who chose not to help. The main effect for target sex,  $F(1, 93) = 3.01, ns$ , and for the two-way interaction,  $F(2, 93) = 1.65, ns$ , was not statistically significant.

*Covariance analyses.* Both the competence ratings and the interpersonal civility ratings were included in subsequent ANCOVAs. Results indicated that regardless of taking either of these attribute scales as a covariate, the hypothesized Target Sex  $\times$  Altruism interaction effect remained statistically significant for each of the two focal dependent variable scales. That is, when competence ratings were covaried out, the interaction remained statistically significant for both the performance evaluation scale,  $F(2, 92) = 15.68, p < .001$ , and the reward recommendation scale,  $F(2, 92) = 5.48, p < .01$ . When the interpersonal civility ratings were covaried out, the interaction remained statistically significant for the performance evaluation scale,  $F(2, 92) = 7.76, p < .01$ . Because the covariate of interpersonal civility was not significant for the reward recommendation scale, no further analysis was conducted. Thus, taking competence and interpersonal civility as

covariates had virtually no effect on the statistical significance of our original interaction.

### Discussion: Studies 1 and 2

The results of Study 2 almost completely replicate the results of Study 1, indicating that reactions to the altruistic citizenship behavior (or lack of it) of men and women were the same regardless of whether respondents were undergraduate students or individuals who were working full-time. Together, the results of the two studies support our predictions regarding differential performance evaluations and reward recommendations for men as compared with women when altruistic citizenship information is provided. Specifically, the performance of altruistic citizenship behavior enhanced the favorability of evaluations and reward recommendations regarding men but not women, whereas the withholding of altruistic citizenship behavior diminished the favorability of evaluations and reward recommendations regarding women but not men. Thus, women neither were given as much credit for their altruism nor treated as tolerantly for their lack of it as were men who behaved identically. As we had suspected, whether they did or did not engage in altruistic citizenship behavior, women tended to be devalued as compared with men.

These results are consistent with the idea that different altruistic citizenship expectations exist for men and women. Also consistent with this idea was the finding that men for whom no information about helping was provided elicited similar reactions as did men who did not help, whereas women for whom no information about helping was provided elicited similar reactions as did women who did help. In the absence of concrete information, there appears to be the expectation that women will be altruistic citizens, but the same expectation does not hold for men.

Ratings on the attribute scales helped rule out some other potential interpretations of our data. The results suggested that the differing pattern of evaluations and reward recommendations for women and men was not because of differing perceptions of their competence or their interpersonal civility. Analyses that enabled us to control for the effects of these perceived attributes on our dependent measures indicated that when competence or interpersonal civility ratings were covaried out, the predicted results concerning performance evaluation and reward recommendations remained the same. Thus, it appears that it was not how the individual was perceived as a result of his or her behavior but rather the degree to which the behavior was viewed as consistent with gender-stereotypic prescriptions that determined the ultimate evaluation or reward recommendation. When consistent, the behavior appears not to have had much impact on decision making, but when inconsistent, the behavior appears to have been weighted heavily. This supports the idea that the same behavior, even when it induces identical perceptions of interpersonal attributes, can have different consequences for evaluation because of its fit or lack of fit with expected behavior—in this case, because of prescriptive gender stereotypes.

The idea that different normative role prescriptions exist about the way in which men and women should respond to the same situation, and that these prescriptions lead to different consequences for them, although they perform the same altruistic behavior, is strongly supported by the results of Studies 1 and 2. But direct evidence that the prescriptive “shoulds” for men and women

differ when altruism is concerned is not provided in these two studies. This was the objective of Study 3.

### Study 3

Although organizational citizenship behaviors are often thought of as being discretionary, our argument rests on the idea that engaging in altruistic citizenship behavior is, in actuality, not always considered to be optional. This idea is consistent with arguments presented by Morrison (1994); Organ (1997); Van Dyne, Cummings, and McLean Parks (1995); and Tepper, Lockhart, and Hoobler (2001); that is, organizational citizenship behaviors are not universally extra-role. If altruism is, as we have argued, a “should” for women because of gender prescriptions to be communal, then helping behavior will be seen as a job requirement for women, albeit a tacit requirement, whereas it is regarded as discretionary for men. As suggested by Kidder and McLean Parks (2001), and again by Kidder (2002), as well as by Ehrhart and Godfrey (2003), altruistic citizenship behaviors may be considered to be more in-role than extra-role for women.

In Study 3, we sought to determine the extent to which altruistic citizenship behaviors are thought to be required or optional for women and men employees holding the same job. It was our expectation that altruistic citizenship behaviors will be seen as less optional and more required for women than for men.

In conducting this research, we also were able to explore whether there are other types of citizenship behaviors that are considered to be required for women but discretionary for men, and whether there were citizenship behaviors that were just the opposite—considered to be required for men but discretionary for women. Accordingly, we investigated citizenship behaviors that involve agentic behavior as well as ones that involve communal behavior.

### Method

#### Participants

Forty-one MBA students at a northeastern university participated in this study as part of a class exercise. They ranged in age from 22 to 29 years old ( $M = 25.39$ ,  $SD = 1.60$ ). Of the participants, 71% were men, and 29% were women. Participants had an average of over 5 years of work experience ( $M = 5.15$ ,  $SD = 1.51$ ), with 88% having had managerial experience. Of those participants with managerial experience, individuals had an average of 2.17 years of experience ( $SD = 1.18$ ).

#### Design

Study 3 was a two-group between-subjects design, with sex of the target being rated as the independent variable. Participants were randomly assigned to the two conditions, with 20 participants in the female target condition and 21 participants in the male target condition.

#### Procedure

The research was said to be about the expectations people have about work performance. The cover page of the research packet handed out to the participants contained the rationale for the study. It read,

There are a variety of informal behaviors that occur at work that are not written into the formal job description. Engaging in these behaviors is often optional, but that may vary, depending on the situation.

Our goal is to begin to identify the different types of expectations people have about these informal work behaviors, and how these expectations vary, depending on job type (e.g., managerial vs. non-managerial, service vs. nonservice), jobholder (e.g., older vs. younger workers), and type of work organization (e.g., nonprofit vs. profit making, large vs. small).

Accordingly, participants were told that they would be reviewing a description of one specific job, jobholder, and work organization, and they were to read the description and indicate the extent to which they thought each of the work behaviors that followed is considered to be required or optional for “this individual working in this job.”

At the top of the next page was a photo of the jobholder for whom the judgments were being made, along with information about the individual’s educational background, job history, and current job description. The jobholder was described as having graduated from college 8 years previously and to be currently working in a middle management position in the Purchasing department of a multinational manufacturing company. The job description included a list of responsibilities typically required of a purchasing manager (e.g., plan, manage, and monitor expense accounts for business units; supervise subordinates in approving and authorizing purchase requests; establish and maintain vendor relationships; research product pricing). Following this set of background information was a list of “informal” work behaviors and instructions directing respondents to indicate the extent to which each behavior is considered to be required or optional for the jobholder to perform.

#### Independent Variable Manipulations

Sex of the target was manipulated both by the name of the jobholder and by the photograph provided. Photos were  $2 \times 3$ -in. ( $5 \times 8$ -cm) portraits and depicted either a man or a woman in business attire. The male and female targets were the same as those used in Studies 1 and 2; their portraits were created from the stimulus photos from the earlier studies.

#### Dependent Measures

The primary dependent measure of interest was the extent to which the respondent thought it was required or optional for the target to perform altruistic citizenship behaviors. Three such behaviors were included in the list: “putting in extra time to help a coworker with a work-related problem,” “resolving conflicts between coworkers,” “helping new employees settle into the job,” two of which were based on items in Moorman and Blakely’s (1995) measure of altruistic citizenship behavior. In addition, there were 10 other behaviors listed, included for “filler” purposes, many of which were organizational citizenship behaviors other than altruism. Each item was measured on a 7-point response scale ranging from 1 (*definitely required*) to 7 (*completely optional*).

### Results

Preliminary ANOVA analyses, including participant sex and managerial experience as additional independent variables, indicated no differences on the basis of these participant demographic characteristics. We therefore combined responses across male and female participants and across managerial experience in all subsequent analyses.

A series of *t* tests were conducted to compare the difference between ratings for male and female targets on each of the dependent measure scale items. The relevant means, standard deviations, and *t* statistics appear in Table 5.

#### Altruistic Citizenship Behaviors

Results of the *t* tests indicated that there was a significant difference in the ratings for each of the three altruistic citizenship



Table 5  
Means, Standard Deviations, and *t* Statistics: Study 3

Dependent variable	Female target		Male target		<i>t</i> (39)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Putting in extra time to help coworkers with work-related problems	1.95	1.15	3.19	0.87	-3.91***
Resolving conflicts between coworkers	2.50	0.76	3.52	0.68	-4.55***
Helping new employees settle into the job	1.85	0.37	3.29	0.64	-8.72***
Alerting upper management to potentially troublesome issues	2.40	1.00	1.57	0.60	3.25**
Keeping others informed about industry trends and changes	2.60	1.00	2.00	1.05	1.87 <sup>a</sup>
Working extra hours during busy times	3.00	0.73	2.29	1.06	2.51*
Making suggestions about important work processes	2.05	0.89	1.57	0.81	1.81 <sup>a</sup>
Attending company-sponsored social events	3.60	0.88	3.67	1.46	-0.18, <i>ns</i>
Speaking well of the organization to outsiders	2.00	0.97	2.14	1.42	-0.37, <i>ns</i>
Attending optional training sessions	4.00	0.86	4.19	1.29	-0.55, <i>ns</i>
Being punctual	1.90	0.72	2.00	1.10	-0.34, <i>ns</i>
Keeping one's work area tidy	2.00	0.56	3.19	0.75	-5.73***
Maintaining a positive attitude	1.70	0.47	2.67	0.97	-4.04***

*Note.* Twenty participants rated the female target, and 21 participants rated the male target. The female target was coded as 0, and the male target was coded as 1; therefore, *t* values with a negative sign indicate that the woman was rated as significantly more required to perform the activity than the man.

<sup>a</sup>*p* < .07. (These trends in the data are reported given the exploratory nature of the results and their implications for future research.)

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

behaviors, with the behavior rated as more required and less optional for the female than the male target. Thus, "putting in extra time to help coworkers with work-related problems" was rated as being significantly more required of women than of men, as was "resolving conflicts between coworkers" and "helping new employees settle into the job." These results strongly supported our hypothesis.<sup>2</sup>

### Other Behaviors

Results of the *t* tests conducted on the other 10 stimulus behaviors were exploratory. These behaviors had been included primarily to provide "cover" for the three altruistic behaviors that were of interest to us so that our interest would not be too apparent and therefore create demand. We therefore had no hypotheses about them. Nonetheless, because we chose to include behaviors that fall within the boundaries of types of organizational citizenship behavior other than altruism, an examination of these results is potentially informative.

Most interesting was the pattern of ratings that occurred with items that can be thought of as falling into the dimensions of organizational citizenship behavior that involve agentic behaviors, known as civic virtue and individual initiative (Podsakoff et al., 2000). This pattern was the reverse of that resulting from analyses of the altruism ratings. Thus, "alerting upper management to potentially troublesome issues," which may be considered to be civic virtue-type citizenship behavior, was rated as being required more of men than of women. Similarly, "working extra hours during busy times," which may be considered to be individual initiative-type citizenship behavior, also was rated as being more required of men than of women. These results are consistent with Ehrhart and Godfrey's (2003) findings that men are significantly more associated with civic virtue and individual-initiative citizen-

ship behaviors than are women. Additional support for this point is provided by the ratings of "keeping others informed about industry trends" (civic virtue-type citizenship behavior) and "making suggestions about important work processes" (individual initiative-type citizenship behavior), both of which tended to be rated as being required more of men than of women (*p* < .07 for each).

Of the remaining behaviors, two that may be considered to be loyalty-type citizenship behaviors ("attending company sponsored social events" and "speaking well of the organization to outsiders"), one that may be seen as being a self-development type of citizenship behavior ("attending optional training sessions"), and one that may be categorized as compliance citizenship behavior ("being punctual") were not rated differently for men and women. However, two additional behaviors, both typically considered part of the prescriptive gender stereotype for women, and of these, one perhaps falling in the category of sportsmanship, were rated as being more required for women than for men: "keeping one's work area tidy" and "maintaining a positive attitude."

### Discussion: Study 3

The results of Study 3 make clear that altruistic citizenship behaviors are not seen as equally discretionary for everyone. On the contrary, they were regarded as more optional for men than for women. This finding helps support our proposition that gender-

<sup>2</sup> When the three altruistic citizenship behavior items were combined into one scale ( $\alpha = .84$ ), the data pattern was identical. The mean rating for the female target was 2.10 (*SD* = 0.59), whereas the mean rating for the male target was 3.33 (*SD* = 0.61), and the difference between them was highly significant,  $t(39) = -6.53, p < .001$ .

stereotypic prescriptions, by affecting expectations about what women “should” do, underlie the differential reactions to them documented in Studies 1 and 2. For if altruism is seen as a requirement for women, then engaging in it is not particularly noteworthy, whereas not engaging in it is. Moreover, if altruism is seen as discretionary for men, then engaging in it is special and recognition-worthy, but not engaging in it is of little consequence. These results help verify the underpinnings of the process we had posited to mediate the effects we had predicted and obtained in our two earlier studies.

It is interesting to note that the results of our exploratory analyses imply that not all organizational citizenship behaviors are more optional for men than for women. Indeed, behaviors typically considered part of the organizational citizenship dimensions of civic virtue and individual initiative, which are agentic in character, seemed to be viewed as more optional for women than for men. This finding, which coincides with ideas presented by Kidder (2002), Kidder and McLean Parks (2001), and Ehrhart and Godfrey (2003), suggests that in some instances, men, too, may be disadvantaged by gender role prescriptions when citizenship behavior is performed or withheld. According to our reasoning, the critical issue is not the sex of the target, but rather the constraints of gender role prescriptions.

### General Discussion

This set of studies demonstrates that the same altruistic behavior can result in different performance evaluations and reward recommendations when the individual is a man or a woman. Studies 1 and 2 indicate that providing work-related help resulted in more favorable reactions to men but had little effect on reactions to women, and the withholding of work-related help resulted in unfavorable reactions to women but had little effect on reactions to men. Women, it seems, are truly disadvantaged when it comes to altruism: When they have acted altruistically, they do not benefit, and when they have failed to act altruistically, they are penalized as compared with identically behaving men. Whatever they do, women wind up less highly regarded than their male counterparts.

This pattern of results is consistent with our contention that the prescriptive nature of gender stereotypes, because it dictates that women should be communal, and therefore altruistic, is the impetus for the differential reactions to women’s and men’s altruistic citizenship behavior. The results of Study 3, which demonstrated that altruism is considered to be less discretionary and more required for women than for men, support this idea. For it is only if it is thought to be discretionary that engaging in citizenship behavior is likely to have a positive impact on evaluations and judgments. If it is seen as “required,” then penalties for failure to engage in citizenship behavior are reasonable and predictable.

The results reported here may have implications for a broader range of behavior than organizational citizenship. They suggest that other behaviors prescribed by female gender stereotypes, such as being approachable or working cooperatively, will have similar consequences for personnel judgments and decisions concerning women. If, indeed, it is the case that women generally are penalized for not performing their communal “shoulds” and not given credit when they do perform them, then our findings may be more far-reaching than our results indicate. Research is presently underway to determine how robust our findings are and the degree to

which they are informative about evaluations on the basis of other aspects of work performance.

It is interesting to note that in none of the three studies was there a difference found in responses of male and female research participants. This finding suggests that gender-stereotypic norms are universal as are reactions to their violation, and is not at all supportive of the view that women are more lenient than men in evaluating other women. This null finding is consistent with the results of several studies about reactions to women’s violations of gender role prescriptions (e.g., Butler & Geis, 1990; Heilman et al., 2004), but inconsistent with the results of others (e.g., Carli et al., 1995; Rudman, 1998). The conditions that regulate when a perceiver’s sex makes a difference in the enforcement of gender-based norms have yet to be identified.

Although this research was primarily focused on gender issues, the findings also help to further our understanding of organizational citizenship behavior. The results demonstrate that the same citizenship behavior is sometimes considered to be in-role and at other times to be considered extra-role, depending on the type of citizenship behavior and whether a man or a woman is performing it. Thus, they lend support to arguments that organizational citizenship behaviors are not always perceived to be discretionary and that the way in which any one citizenship behavior is viewed can vary, depending on circumstance. This has definite implications for the likely impact of these behaviors on evaluations and decisions about deservingness for organizational rewards.

Although our research indicates different reactions to organizational citizenship behavior when the behavior is performed by a woman or a man, there still are many questions remaining. For instance, in our first two studies, we examined reactions to men and women performing one specific type of citizenship behavior—altruism. It is important to determine whether similar results occur when studying other citizenship behaviors that coincide with female role prescriptions and whether different results occur when studying organizational citizenship behaviors that are not part of female role prescriptions. Also, it would be interesting to determine whether an effect analogous to the one found for women in our studies is found for men in instances when being a good citizen entails acts that are considered to be agentic and therefore are prescribed for men in our culture—for example, acts of civic virtue or individual initiative, or helping behaviors that require assertiveness and physical prowess.

Between-subjects designs were used for all of our studies, but it is a within-subjects world; it therefore is important to replicate this work using designs in which participants evaluate both men and women. Moreover, because our research was conducted using a “paper people” methodology, there is a definite need to replicate these findings in settings in which people actually witness and/or are the potential beneficiaries of citizenship behaviors performed by others; reactions may differ when one’s own or another’s personal well-being is immediately at stake. Also, although there are data that indicate that reactions to individuals performing organizational citizenship behaviors are similar whether the research is conducted in the laboratory or in field settings (e.g., Allen & Rush, 1998), it is important to determine whether the particular effects we uncovered in this study would occur in actual work situations, in which more and richer information about an individual’s behavior over time is available to raters.

Despite these many additional questions to be addressed, the results of this investigation are suggestive about the way in which prescriptive aspects of gender stereotypes may subtly affect the evaluation of men and women in work settings. The findings also are cause for concern. They indicate that the outcomes for men and women who perform exactly the same altruistic citizenship behavior can be dramatically different, with only the men, not the women, benefiting from their socially responsible actions.

## References

- Allen, T. D., & Rush, M. C. (1998). The effects of organizational citizenship behavior on performance judgments: A field study and a laboratory experiment. *Journal of Applied Psychology, 83*, 247–260.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personality selection* (pp. 71–98). San Francisco: Jossey-Bass.
- 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.
- 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.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin, 108*, 233–256.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573–598.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*, 3–22.
- 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., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46*, 735–754.
- Ehrhart, M. G., & Godfrey, E. G. (2003, April). *The role of schemas in gender and organizational citizenship behavior research*. Poster session presented at the 18th Annual Conference for the Society of Industrial and Organizational Psychology, Orlando, FL.
- 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.
- Kidder, D. L. (2002). The influence of gender on the performance of organizational citizenship behaviors. *Journal of Management, 28*, 629–648.
- Kidder, D. L., & McLean Parks, J. (2001). The good soldier: Who is (s)he? *Journal of Organizational Behavior, 22*, 939–959.
- Mackenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons' performance. *Organizational Behavior and Human Decision Processes, 50*, 123–150.
- Mackenzie, S. B., Podsakoff, P. M., & Fetter, R. (1993). The impact of organizational citizenship behavior on evaluations of salesperson performance. *Journal of Marketing, 57*, 70–80.
- Moorman, R. H., & Blakely, G. L. (1995). Individualism–collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior, 16*, 127–142.
- Morrison, E. W. (1994). Role definitions and organizational citizenship behavior: The importance of the employee's perspective. *Academy of Management Journal, 37*, 1543–1567.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475–480.
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10*, 85–97.
- Podsakoff, P. M., Mackenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26*, 513–563.
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypical impression management. *Journal of Personality and Social Psychology, 74*, 629–645.
- Tepper, B. J., Lockhart, D., & Hoobler, J. (2001). Justice, citizenship, and role definition effects. *Journal of Applied Psychology, 86*, 789–796.
- Van Dyne, L., Cummings, L. L., & McLean Parks, J. M. (1995). Extra role behaviors: In pursuit of construct and definitional clarity (a bridge over muddied waters). In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 17, pp. 215–285). Greenwich, CT: JAI Press.
- Werner, J. M. (1994). Dimensions that make a difference: Examining the impact of in-role and extrarole behaviors on supervisory ratings. *Journal of Applied Psychology, 79*, 98–107.

Received August 22, 2003

Revision received April 19, 2004

Accepted April 27, 2004 ■