Sex Roles, Vol. 37, Nos. 9/10, 1997

Context Effects on Self-Perceptions of Feminine and Masculine Qualities

Judi Beinstein Miller,[1] Jennifer Lewy,[1] and Emily Peckham[1]

Activation of gender-related experiences in memory was predicted to enhance self-perceptions of stereotypically feminine and masculine qualities. In two experiments with undergraduate students—approximately 6% African American, 12% Asian American, 74% Caucasian, 6% Latino, and 2% of other backgrounds in each—the effects of recalling either expressive or instrumental experiences are evaluated in relation to subjects' self-ratings on related feminine and masculine traits. Results show consistent increases in ratings of feminine traits in expressive conditions and in ratings of masculine traits in instrumental ones, but these tendencies occur primarily among subjects with initially low self-ratings. Results are interpreted as evidence that context can influence gender-related self-evaluation by activating related experience in memory and enhancing its salience for identity.

Early gender research measured femininity and masculinity as stable personality traits that differentiated females and males (Morawski, 1985). More recent scholarship has questioned the stability of femininity and masculinity in identity and suggested an alternative view, that self-perceptions of gender are somewhat fluid and influenced by context (Deaux & Major, 1987; Lorber, 1994). Social interactions, according to this viewpoint, can encourage displays of gender-appropriate behavior, which provide evidence of a person's femininity and masculinity (West & Zimmerman, 1987). Contexts that promote or limit these displays can thereby increase or decrease self-perceptions of gender appropriateness. The purpose of the current research was to evaluate one process by which context might influence gender identity, the activation of gender-congruent experience.

[1]Department of Psychology, Oberlin College, Oberlin, OH 44074.

0360-0025/97/1100-0723$12.50/0

Effects of Context on Gender Displays

Assumptions about gender have changed since the first part of the 20th century, when femininity and masculinity were construed as unidimensional, bipolar, and essential aspects of human beings (Morawski, 1985). Today we recognize the multidimensionality and potential independence of feminine and masculine attributes (cf. Constantinople, 1973; Spence, 1984; Helgeson, 1994). Yet despite evidence for the social construction, and hence malleability of gender attributes (Lorber, 1994), there has been little empirical attention to their variation in identity by situation. An important step in this direction has been the recognition that gender-appropriate behavior can be evoked through self-fulfilling prophecies. Deaux and Major (1987) propose a model in which situational cues activate gender expectations and individuals, acting in anticipation of their expectations, evoke gender-appropriate behavior. Circumstances that encourage the treatment of individuals as women or men can encourage them to behave accordingly (Geis, 1993). In these cases confirmation of social expectations is set in motion by contextual stimuli—for example, when a person's office and clothing elicit respect and submissiveness from others and result in the person behaving dominantly. West and Zimmerman (1987) have argued in a related vein that gender is continually created in social interaction. Successful displays of gender entail the fitting of gender-appropriate behavior to context and successful performers are conferred a gender-appropriate character.

Effects of Gender Displays on Gender Impressions

Perception of others' gender role conformity—for example, their occupation—can bias inferences about their other gender-related attributes—for example their gentleness, sensitivity, independence, and self-confidence (Deaux & Lewis, 1984). However, the extent to which self-perceptions are biased by gender-appropriate role behavior is unclear. Actor–observer biases predict that we attribute our own behavior to the constraints of situations and that of others to their dispositions (Jones & Nisbett, 1972). We might be less likely to diagnose ourselves than others by single situations due to greater awareness of variety in our own responses than in theirs. Instances of our own behavior might therefore carry less inferential weight in judgment than do instances of others' behavior.

In contrast, research on construct accessibility suggests that gender self-assessments will shift in accordance with gender-congruent behavior (cf. Bargh, Lombardi, & Higgins, 1988; Sedikides & Skowronski, 1990; Srull & Wyer, 1980). By this line of reasoning, constructs available in memory are used to draw inferences about the self, and one's behavior can activate

related constructs in memory. Gender-related behavior might activate gender-related constructs in memory and influence the experience of the self. Indeed, feminine and masculine constructs have been viewed as chronically accessible components of the identities of gender-schematic persons (e.g. Markus, Crane, Bernstein, & Siladi, 1982; Rubini & Antonelli, 1992). As demonstrated by gender-related biases in information processing, gender can become a primary organizing principle for seeing the world (Bem, 1993). Contexts that promote gender-appropriate behavior might thereby indirectly influence gender self-perceptions.

According to research on construct accessibility, feminine and masculine constructs can be situationally activated, even among gender aschematics (Stangor, 1988). Yet little is known about the effects of gender-appropriate behavior on self-evaluation. One exception is research in which subjects evaluate their expressive and instrumental qualities in different social roles. Results from these studies demonstrate differences in subjects' self-ratings, depending on whether they receive standard instructions to rate themselves generally, or instructions to rate themselves in social, sexual, parental, student, or working roles (Dailey & Rosenzweig, 1988; Lawrance, Taylor, & Byers, 1996; Rosenzweig & Dailey, 1991, Uleman & Weston, 1986). However, it is difficult to know from this research whether differences in response are based on social expectations associated with the roles or on subjects' actual role behavior, and whether a single role-related experience would influence ratings of feminine and masculine qualities.

Contextual Activation of Femininity and Masculinity

We designed two studies to determine whether situational activation of gender-related experiences would influence self-ratings on stereotypically masculine and feminine qualities. We asked undergraduate students to rate themselves on a series of feminine and masculine qualities early in the semester as part of a questionnaire battery, and several weeks later, prior to rating themselves again, asked them to tell us about previous expressive or instrumental experiences while they played the role of a job applicant. This design enabled us to determine whether recall of gender-related experiences would enhance subsequent self-ratings of gender-related qualities. We selected expressive experiences of interpersonal responsiveness and instrumental experiences of self-assertion because these reflect major dimensions of scales that are intended to measure psychological femininity and masculinity (Spence, 1984). We predicted that retrieval of gender-congruent experiences would increase self-assessments of gender-related qualities. This prediction is consistent with research in which self-perceptions of extraversion are manipulated by questions about extraversion (Riggs, Mon-

ach, Ogburn, & Pahides, 1983) and in which self-esteem is manipulated by inducements to behave in self-enhancing and self-deprecating ways (Jones, Berglas, Rhodewalt, & Skelton, 1981; Rhodewalt & Agustsdottir, 1986). We predicted that activation of expressive experiences would increase self-ratings on stereotypical feminine qualities and that activation of instrumental experiences would increase self-ratings on stereotypical masculine ones.

EXPERIMENT 1

Method

Subjects. Subjects were 92 introductory psychology students (46 males and 46 females) from a midwestern liberal arts institution, who volunteered in return for partial course credit. Although data about ethnicity were not collected, their ethnic composition can be estimated by class enrollment figures, according to which approximately 6% were African American, 12% Asian American, 74% Caucasian, 6% Latino, and 2% of other backgrounds. Most students in the class (90%) were in their first or second year. At the beginning of the semester, students rated themselves on 16 gender-related traits, as part of a battery of questionnaires. Several weeks later they were randomly assigned to one of two conditions (expressive or instrumental) to play the role of job applicant in videotaped interviews. Two students did not complete the self-ratings and 5 did not complete the interviews. Of the remaining subjects, 22 females and 22 males participated in the expressive condition, 22 females and 19 males in the instrumental condition. Subjects were interviewed one at a time.

Pretest and Posttest Questionnaires. The pretest questionnaire contained 24 items from the Bem Sex Role Inventory (BSRI, Bem 1974)—8 feminine, 8 masculine, and 8 neutral items. Seven of the feminine items (i.e. affectionate, compassionate, eager to sooth hurt feelings, gentle, sensitive to needs of others, sympathetic, and tender) and 7 of the masculine items (i.e. acts as leader, aggressive, assertive, dominant, has leadership ability, independent, and willing to take a stand) were selected for their consistently strong loadings on dimensions of expressiveness and instrumentality in prior factor analyses (Ballard-Reisch & Elton, 1992). The items "feminine" and "masculine" were also included on the questionnaire. The 8 neutral items (adaptable, conventional, friendly, helpful, moody, reliable, tactful, and unpredictable) were included to disguise the purpose of the test. Subjects rated themselves on each item using 7-point Likert scales. They also responded to 5 other multiitem tests, which were unrelated to the purposes of this experiment.

Internal consistency (Cronbach's alpha) for pretest ratings of the 8 feminine items was .81 for females and .87 for males. Alphas for the 8 masculine items were .80 and .82, respectively. The mean of ratings on feminine items was computed for a summary femininity score and the mean of ratings on masculine items was computed for a summary masculinity score. Femininity scores were higher for females (M = 5.44, SD = 0.73) than for males (M = 4.92, SD = 0.91), $t(83) = 2.96$, $p < .01$, but there was no gender difference in masculinity scores (females M = 4.55, SD = 0.84 and males M = 4.75, SD = 0.87), $t(83) = 1.09$, p .20.

The posttest questionnaire, on which subjects rated themselves following the interview, contained the same 8 feminine, 8 masculine, and 8 neutral items as on the pretest. These items were interspersed among an additional 12 gender-neutral items from the BSRI (conceited, conscientious, happy, inefficient, jealous, likable, secretive, sincere, solemn, theatrical, truthful, unsystematic) to change the appearance of the questionnaire from its pretest form and disguise the purpose of the experiment. The posttest questionnaire also requested self-ratings of subjects' performance in the interview. After the BSRI items, subjects rated their comfort in the interview, the quality of their role playing, and their difficulty in answering the interview questions, also on 7-point scales.

Alphas for posttest ratings on the 8 feminine items were .82 for females and .82 for males. Alphas for the 8 masculine items were .80 and .81, respectively. Posttest femininity and masculinity scores were computed as for pretest scores. As on the pretest, femininity scores were higher for females (M = 5.61, SD = 0.70) than for males (M = 5.06, SD = 0.78), $t(83) = 3.42$, $p < .001$, and there was no gender difference in masculinity scores (females M = 4.80, SD = 0.81 and males M = 4.84, SD = 0.88), $t(83) = 0.19$, $p > .80$.

Correlations between pretest and posttest scores were moderately strong. Correlations between femininity scores were .70 ($p < .001$) for females and .74 ($p < .001$) for males, and correlations between masculinity scores were .75 ($p < .001$) for females and .85 ($p < .001$) for males.

The alpha for subjects' ratings of their comfort, quality of role playing, and difficulty in answering the interview questions (the latter scored in a reverse direction) was .81. The mean of these ratings was computed as an average performance score, which was moderately good (M = 4.00, SD = 1.36). An ANOVA of performance scores by subject gender and condition yielded no statistically significant main or interaction effects, $F(1,81) = 2.75$, $p > .10$, $F(1,81) = 0.37$, $p > .50$, and $F(1,81) = 0.01$, $p > .90$ respectively.

Interview Materials. The job description was for a "psychiatric assistant" who was expected to work closely in recreational and group activities with adolescent and young adult patients on a psychiatric hospital ward. Beyond

this general expectation, the job requirements differed by condition. In the expressive condition the job required someone who could work as part of a team, compromise with staff members, follow orders, and speak soothingly to patients. The person was expected to be responsive to the needs of patients, listen to their concerns, share their feelings, and encourage conversation about their personal experiences. In the instrumental condition the job required someone who could work on his/her own, take initiative in making decisions, speak authoritatively to patients, and give them orders. The person was expected to be firm about patients' routines, capable of focusing them on constructive activity rather than self-concern, and able to separate feelings for the patients from their treatment.

The first two interview questions were the same in both conditions. Subjects were asked to state their name and address and then to describe a situation (e.g., a job) in which they had been responsible for other people. The questions that followed depended on the subject's assigned condition and were intended to elicit recall of experiences consistent with the job requirements. In the expressive condition they responded to the following questions, intended to activate stereotypically feminine experiences.

1. Please describe a job you've had, or other kind of situation, where you had to work cooperatively as part of a team.
2. Please describe a job you've had, or other kind of situation, where responsiveness to other people was critical.
3. Please describe a job you've had, or other kind of situation, that required you to encourage the sharing of feelings.

In the instrumental condition they responded to the following questions, intended to activate stereotypically masculine experiences.

1. Please describe a job you've had, or other kind of situation, where you had to work on your own and take initiative in decision-making.
2. Please describe a job you've had, or other situation, where firm guidance of other people was critical.
3. Please describe a job you've had, or other kind of situation, that required you to not let your feelings interfere with your work.

Procedure. Three weeks after the start of classes, the course instructor circulated sign-up sheets for the experiment. The experiment was described on the signup sheet as a study of impression making in employment interviews. Subjects would play the role of a job applicant while being videotaped so that the effect of their verbal and nonverbal behavior could be compared. Volunteers were expected to allow 40 minutes for the experimental session.

All interview sessions took place in the Psychology Department's social psychology lab, which had been outfitted with a sofa, coffee table, and arm chairs. The lab was adjacent to a control room from which the interviews were unobtrusively videotaped.

Upon arrival, volunteers were greeted in a waiting room by a student experimenter, who reminded them that this was a study of impression making in job interviews. She explained that we were investigating the verbal and nonverbal behaviors that inexperienced applicants use to create favorable impressions and indicated how helpful their participation would be. She then specified that they would be getting a job description and should imagine that they really wanted the job. The interview would entail questions about their own, real life experiences in situations relevant to the job. They were reminded that the interview would be videotaped and then asked for consent to proceed. All consented to continue.

Having consented to proceed, subjects were given the job description and, 3 minutes later, the interview questions, each typed on a separate card. They received one of two written job descriptions, which assigned them to condition, and one of two sets of interview questions that depended on their assigned condition. Job descriptions were systematically alternated after the first random assignment, and were alternated separately for females and males to maintain equal numbers in the two conditions.

The experimenter escorted the subjects to the interview room after providing them with the interview questions. She seated them on the sofa and asked them to read and answer one question at a time, in the order given, after she left the room. When they finished answering the last question she returned, announced that the experiment was over, thanked them for their participation, and escorted them to a debriefing room. Prior to debriefing, she asked them to provide some additional information about themselves that would aid interpretation of their interview tape. She then gave them the posttest questionnaire, which contained the BSRI and performance items, and asked them to rate themselves on each.

After subjects completed the posttest questionnaire, the experimenter asked them to evaluate the experiment and record whatever they thought had been its specific purpose. Afterwards, she debriefed them by explaining that we were interested in the way that situations can change a person's sense of self. She pointed out that many of the items on the posttest questionnaire had also been part of the survey administered on the first day of class. She indicated that we wanted to determine whether their ratings would change as a consequence of the job qualifications they provided in their interview responses. She asked them not to talk with anyone about the study until all sessions had been completed and thanked them once more for their help.

There were no subjects who expressed knowledge of the experiment's purpose. Many wrote that they were unsure or did not know, but then added a possible idea or two. Subjects' ideas about the experiment fell into a limited number of categories. Some suggested that we were investigating successful interview behavior, or the behaviors people use when they are trying to be convincing. Others thought the interview was a ruse for studying responses to discomfort, anxiety, or nervousness. Still others made use of the posttest personality traits in their ideas and suggested we could be studying the personality correlates either of effective interview performance or of anxiety management. The interview situation was clearly stressful for many subjects and their discomfort was reflected in these responses.

Supplementary Analyses of Videotapes. Since awareness of being videotaped could have made subjects anxious and interfered with their interview responses, it was desirable to obtain a measure of their apparent comfort and confidence in the interview. Videotapes were therefore analyzed for nonverbal indicators of confidence. Two independent coders, blind to subjects' self-ratings but not to condition, coded each interview—except for one, which was blank due to technical difficulties—for evidence of expansive as opposed to contracted posture, body relaxation as opposed to tension, loud as opposed to quiet voice, direct as opposed to hesitant manner, and direct as opposed to indirect gaze. These dimensions were coded on 3-point scales, as bipolar opposites. Additionally, interviewees' use of hand movements to illustrate their answers was coded on a 3-point scale ("hardly any" to "many"). Coding reliabilities (Spearman–Brown) ranged from .77 to .94. The median reliability was .82. The internal consistency of these ratings was estimated at .71 (Cronbach's alpha). Ratings for each judgment were averaged across coders and the mean of averaged judgments was computed for a confidence score. Confidence scores were higher for males ($M = 2.27$, $SD = 0.38$) than for females ($M = 1.84$, $SD = 0.35$), $t(82) = 5.38$, $p < .001$.

It is important to note that subjects' gender and condition were obvious from the videotapes and could have therefore biased coding decisions. Subjects' self-ratings of feminine and masculine traits were not available during coding, however, and had no influence on these decisions.

RESULTS

Interview Effects. The hypothesis that gender-related self-ratings would increase from pretest to posttest in gender congruent conditions was tested in a gender rating (femininity vs. masculinity scores) by test (pretest vs. posttest) by gender (female vs. male) by condition (expressive vs. instru-

Table I. Mean Pretest and Posttest Femininity and Masculinity Scores by Condition, Experiment 1[a]

	Expressive Condition		Instrumental Condition	
BSRI Score	Mean	(*SD*)	Mean	(*SD*)
Femininity				
Pretest	5.13	(0.82)	5.25	(0.90)
Posttest	5.38	(0.65)	5.30	(0.92)
Change	0.25*	(0.55)	0.05	(0.62)
Masculinity				
Pretest	4.70	(0.77)	4.60	(0.95)
Posttest	4.74	(0.85)	4.91	(0.83)
Change	0.04	(0.48)	0.31*	(0.56)

[a]*Note.* Change scores represent increases (on 7-point scales) from pretest to posttest, $*p < .05$ and $***p < .001$; interaction contrasts for changes in femininity and masculinity scores are significant at $p < .05$ and $p < .10$ in the instrumental condition and expressive condition respectively; expressive condition $n = 44$ and instrumental condition $n = 41$.

mental) mixed analysis of variance (ANOVA), the first two factors within subjects and the second two between subjects. The predicted outcome was for femininity scores to increase more in the expressive condition than in the instrumental condition and for masculinity scores to increase more in the instrumental condition than in the expressive condition. The predicted Gender Rating × Test × Condition interaction was statistically significant, $F(1,81) = 9.02, p < .01$. Separate analyses by condition yielded a consistent Gender Rating × Test interaction contrast in the instrumental condition, $F(1,40) = 6.88, p < .05$. A similar although statistically insignificant trend was obtained in the expressive condition, $F(1,43) = 3.23, p < .08$. Comparisons of femininity scores yielded a statistically significant increase from pretest to posttest in the expressive condition, $t(43) = 2.97, p < .01$, but not in the instrumental condition, $t(40) = 0.61, p > .50$. A similar comparison of masculinity scores yielded a statistically significant increase in the instrumental condition, $t(40) = 3.53, p < .001$, but not in the expressive condition, $t(43) = 0.63, p > .50$ (see Table I).

A main effect of test was also obtained, $F(1,81) = 12.16, p < .001$; posttest scores were generally higher than pretest scores, a result consistent with condition effects.

Effects of Subject Gender. The higher order Gender Rating × Test × Condition × Gender interaction was not statistically significant, $F(1,81) = 0.68, p > .40$, and indicated that the predicted Gender Rating × Test × Condition interaction occurred for both females and males. Other ANOVA results included a main effect of gender rating, $F(1,81) = 21.80, p < .001$; femininity scores were generally higher than masculinity scores. This effect was qualified by a gender rating × gender interaction, $F(1,81) = 8.25, p < .01$, indicating that the difference in femininity and masculinity scores was

greater among females than among males. This result is consistent with gender differences in pretest and posttest femininity and masculinity scores reported earlier, that females and males scored similarly on masculine traits, but that females scored higher than males on feminine traits.

Additional Correlates of Subjects' Changes in Femininity and Masculinity Scores. Although the average increase in femininity scores in the expressive condition was only 0.25 (SD = 0.55), and the average increase in masculinity scores in the instrumental condition was only 0.31 (SD = 0.56), it is important to remember the variety of individual differences that could have influenced subjects' ratings. Standard deviations for changes in femininity scores indicated a range of average deviation from -0.30 to 0.80 in the expressive condition; standard deviations for changes in masculinity scores indicated a range of average deviation from -0.25 to 0.87 in the instrumental condition. Thus some subjects demonstrated decreases in scores whereas others demonstrated substantial increases, indicating that a variety of individual differences were probably influencing the magnitude of change in scores.

One factor that influenced the magnitude of change was subjects' initial level of self-ratings. Correlations between changes in femininity scores (posttest score—pretest score) and pretest femininity scores were -.43 ($p < .01$) and -.54 ($p < .001$) for females and males respectively. Correlations between changes in masculinity scores (posttest score—pretest score) and pretest masculinity scores were -.40, $p < .01$ and -.25, $p > .10$. This consistent pattern of correlation—the lower the pretest score, the greater its increase —suggests that the interview had its strongest influence on subjects whose initial ratings of feminine and masculine qualities were low.

Condition effects on subjects with low pretest gender ratings can be demonstrated further with repeated measures ANOVAs of pretest and posttest scores, using condition and level of pretest score (dichotomized by median splits) as predictors. The analysis of femininity scores yielded a main effect of test, $F(1,81) = 6.73$, $p < .05$ and a Test × Femininity Level interaction effect, $F(1,81) = 10.85$, $p < .001$. However, the Test × Condition × Femininity Level interaction was statistically insignificant, $F(1,81) = 0.02$, $p > .80$. As indicated in Table II, femininity scores generally increased from pretest to posttest, except among subjects in the instrumental condition who had high pretest femininity scores, and the increase was significant primarily among subjects in the expressive condition who had low pretest femininity scores. Post hoc Tukey tests indicated that subjects with low pretest femininity scores in the expressive condition increased their self-ratings significantly more than did subjects with high pretest scores in the instrumental condition. In fact, subjects with high pretest scores in the instrumental condition demonstrated a decrease in self-ratings, although the difference was statistically insignificant, $t(18) = -1.51$, $p > .10$.

Table II. Mean Pretest and Posttest Femininity and Masculinity Scores by Condition and Level of Pretest Score, Experiment 1[a]

	Expressive Condition		Instrumental Condition	
	Pretest Scores		Pretest Scores	
BSRI Score	Low	High	Low	High
Femininty	(n = 23)	(n = 21)	(n = 22)	(n = 19)
Pretest	4.58	5.75	4.59	6.01
Posttest	5.01	5.79	4.84	5.85
Change	0.43_a**	0.04_{ab}	0.25_{ab}	-0.16_b
Masculinity	(n = 25)	(n = 19)	(n = 20)	(n = 21)
Pretest	4.15	5.41	3.82	5.33
Posttest	4.16	5.51	4.37	5.42
Change	0.01_b	0.10_b	0.55_a***	0.09_b

[a]Pretest scores for femininity are used as predictors in comparisons of femininity scores and pretest scores for masculinity are used as predictors in comparisons of masculinity scores; change scores represent increases (on 7-point scales) from pretest to posttest, **$p < .01$, and ***$p < .001$; change scores with different subscripts in a row differ at $p < .05$, by the Tukey range test.

The analysis of masculinity scores yielded a main effect of test, $F(1,81) = 10.09$, $p < .01$, a Test × Condition interaction effect, $F(1,81) = 6.50$, $p < .05$, and a Test × Condition × Masculinity Level interaction effect, $F(1,81) = 6.21$, $p < .05$. As indicated in Table II, the general increase in masculinity scores from pretest to posttest was due to increases in the instrumental condition among subjects with low pretest masculinity scores. Increases among these subjects were significantly greater than among the other three groups, as indicated by post hoc Tukey tests.

A second factor associated with the magnitude of change scores was subjects' apparent confidence in the interview. However, this factor correlated primarily with changes in masculinity scores in the instrumental condition. The more confidence the subjects displayed in this condition, the greater the increase in their masculinity scores. Correlations were .47 ($p < .05$) for females and .42 ($p < .09$) for males. Other correlations between subjects' confidence and change scores ranged from −.17 to .26, and were statistically insignficant, $ps > .20$. Similarly, correlations between subjects' pretest scores and apparent confidence ranged from −.18 to .26 and were statistically insignificant, $ps > .20$. Pretest levels of femininity and masculinity appeared not to influence confidence.

In contrast, subjects' performance scores did not correlate significantly with changes in their femininity and masculinity scores, either in the separate conditions ($ps > .20$) or among subjects with low pretest scores in the separate conditions ($ps > .30$), but correlated instead with pretest femininity and masculinity scores. Positive correlations were obtained with pretest femininity scores in the expressive condition ($r = .38$, $p < .01$) and with

pretest masculinity scores in the instrumental condition ($r = .37, p < .05$). Correlations for pretest femininity scores in the instrumental condition and for pretest masculinity scores in the expressive condition were small (.19 and .18 respectively) and statistically insignificant ($ps > .20$). Separate correlations for males and females were consistent with this pattern. Higher pretest levels of feminine and masculine qualities appeared to make the interview easier for subjects in gender-congruent conditions.

Main Components of Change in Femininity and Masculinity Scores. Item analyses were conducted to determine the main components of change in femininity and masculinity scores. Expressive condition ratings of feminine items and instrumental condition ratings of masculine items were compared at pretest and posttest to determine the extent to which each had changed. Repeated measures ANOVAs were conducted for each item, using level of pretest rating (dichotomized by a median split) as a predictor. These comparisons provide more precise analyses of the effects of low pretest scores than do comparisons based on dichotomized femininity and masculinity pretest scores. Comparisons based on dichotomized femininity and masculinity pretest scores aggregate the rating increases of subjects who had low pretest ratings on some items and high pretest ratings on others. These comparisons consequently dilute the magnitude of rating increases. In contrast, when comparisons are based on dichotomized pretest ratings of individual items, a wider range of change in ratings can be demonstrated, as is indicated in Table III.

There were three clear patterns that emerged from these analyses. First, subjects with low pretest ratings demonstrated statistically significant or marginally significant rating increases on nearly every item, an indication that the interview questions had activated a broad set of gender associations. Second, rating increases were stronger for some items than for others. Expressive condition results were strongest for the items, compassionate, sensitive, and sympathetic, and instrumental condition results were strongest for the items, aggressive, assertive, dominant, and independent. Third, subjects with relatively high pretest ratings on individual items tended to rate themselves lower on those items at posttest, although the differences were statistically significant in only 3 out of 16 comparisons.

Discussion

Results from this experiment confirmed the prediction that gender-related enhancement of identity would be greater in gender-congruent than in gender-noncongruent conditions. Self-ratings on feminine traits increased more in the expressive condition, relative to the instrumental condition, and self-ratings on masculine traits increased more in the

Table III. Changes in Ratings of Gender Items in Gender-Congruent Conditions, Experiment 1[a]

Gender Item	Change in Rating		Test Effect F	Test × Pretest Interaction, F
	Low Pretest	High Pretest		
	Expressive condition			
Affectionate	0.44	−0.21	0.88	3.60
Compassionate	0.88***	−0.30*	1.84*	23.97***
Eager to soothe	0.50	−0.25	0.02	5.32*
Feminine	0.58***	−0.11	4.73*	6.20*
Gentle	0.44	−0.21	1.19	4.88*
Sensitive	1.37***	−0.08	18.17***	31.43***
Sympathetic	1.22***	−0.08	12.32***	23.51***
Tender	0.59**	−0.19	2.37	8.26**
	Instrumental condition			
Acts as leader	0.44	−0.21	1.45	2.94
Aggressive	1.15**	−0.52*	1.76	14.39***
Assertive	1.05***	−0.05	9.54**	10.45**
Dominant	0.94*	0.24	6.82*	3.01
Has leadership ability	0.35	−0.13	0.30	3.11
Independent	0.95***	−0.40*	5.32*	28.37***
Masculine	0.68**	0.00	5.34*	6.18*
Willing to take a stand	0.81*	−0.16	2.04	9.52**

[a]Pretest ratings of femininity items are used as predictors in comparisons of femininity ratings and pretest ratings of masculinity items are used as predictors in comparisons of masculinity ratings; *n*s at each pretest level vary somewhat, depending on the dichotomization of pretest ratings; change scores represent increases (on 7-point scales) from pretest to posttest, $*p < .05$, $**p < .01$, and $***p < .001$.

instrumental condition, relative to the expressive condition. Although increases, especially of femininity scores in the expressive condition, were not large in the total sample, they did provide evidence that gender-related constructs, when activated, would have an influence on self-perception.

Results also suggest that contexts of role enactment, and their activation of particular competencies, might carry greater weight in an individual's situational sense of self than does identification with the general role. In this experiment subjects interviewed for a job that was labeled the same way, regardless of condition. All subjects initially learned that the position was for a psychiatric assistant, to work in recreational and group activities on a hospital ward with young adults and adolescents. Consequently all subjects expected to be working with groups of patients, helping them in one or another kinds of activities. However, the job description diverged thereafter, focusing in the expressive condition on responsive and empathic aspects of the role, and in the instrumental condition on controlling and independent aspects of the role. The interview questions were written to access experiences that exemplified these qualities, and subjects, both male and female, shifted their self-ratings accordingly. Expressive condition sub-

jects' increases in femininity scores were due mainly to increases in ratings of compassion, sensitivity, and sympathy rather than ratings of tenderness or affection. Instrumental condition subjects' increases in masculinity scores were due mainly to increases in ratings of aggressiveness, assertiveness, dominance, and independence, rather than ratings of leadership ability. This pattern of change suggests a correspondence between job context, construct activation, and self-perception, although gender qualities less relevant to the job were also activated, as indicated by rating increases on other gender items. Indeed the small and marginally significant increase in femininity scores among instrumental condition subjects with low pretest femininity scores might reflect an association of expressive qualities with caretaking positions, even when these are described in instrumental terms.

That increases in femininity and masculinity scores occurred primarily among subjects with initially lower scores can be interpreted in two ways. One interpretation concerns subjects' integration of past self-assessments and current self-assessments due to the interview experience. Those with lower pretest scores might have performed related roles less often or less well in the past than did subjects with higher scores. If differences in pretest scores reflected past performance then the interview experience, relative to past experience, could have provided more concentrated or extreme examples of gender-related behavior for subjects with low than with high pretest scores. The difference between past and current experience would have resulted in a stronger upward averaging of self-assessments among the former than the latter. Thus the interview's selective activation of gender-consistent experience would have provided greater evidence for the former than the latter to revise their ratings upwards.

A second explanation is related to subjects' understanding of experimental instructions and impression management. Although the experimenter announced the end of the experiment prior to administering the posttest questionnaire, subjects might have continued to think that their job-related impressions were being evaluated. Under these circumstances subjects with lower scores might have changed them more, to sustain the correct impression. Uleman and Weston (1986) reported a similar tendency, for parents with low scores on the BSRI scales to increase their ratings relative to parents with high scores, when rating themselves as parents in comparison to rating themselves globally. They recognized that this tendency could be due to the social desirability of conveying a good impression. Yet they also reported that there was no correlation between parents' increased ratings on the BSRI and their observed play with their children. Had their increased ratings been due to impression management, then their observed play might have reflected this attempt. In the current experiment also there was additional evidence that posttest questionnaire responses

were probably not contrived to sustain a role-consistent impression. Had subjects attempted to sustain role-consistent impressions on the posttest questionnaire, then their performance scores might have reflected the amount of change in their femininity and masculinity scores. Positive correlations might have been expected between performance ratings and changes in femininity and masculinity scores since both would signify needs for social desirability. Yet no statistically significant correlations were obtained between change scores and performance scores either among the total sample or among those whose pretest scores were relatively low. Moreover, impression management attempts would have been reflected in average performance scores, which were taken at the same time as the posttest BSRI ratings. The overall mean of these ratings was 4.0 on 7-point scales ranging from "not at all good" to "very good," and the standard deviation of ratings was 1.4. Thus a majority rated their performance between 2.6 and 5.4, not very good to fairly good, and subjects with low pretest scores rated themselves even lower. It seems unlikely that subjects would have rated their comfort in the interview, ease in answering questions, and quality of performance this low if they were trying to maintain a favorable impression. Although an impression management explanation cannot be ruled out altogether, it is not substantiated by these results. An explanation based on selective activation of gender-consistent experience is probably better than one based on impression management.

The tendency for high pretest ratings to decrease somewhat on the posttest can be interpreted as further evidence that results were not an artifact of impression management. Although high ratings cannot increase by as much as low ratings, it seems unlikely that they would decrease if subjects were trying to create a job-relevant impression. Part of the tendency for ratings to decrease is probably a consequence of their regression to the mean. However, it is also possible that subjects with high pretest ratings found their current, restricted role performance less indicative of expressive and instrumental qualities than they found their ordinary behavior. Similarly, it is possible that rating increases were due, at least in part, to regresssion to the mean. However, the increases among subjects with low pretest ratings were consistently significant and larger than were decreases among those with high pretest ratings. The stronger effects of condition among subjects with low pretest ratings is difficult to explain completely by regression to the mean.

We also found that nonverbal expressions of confidence were associated with increases in masculinity scores, but primarily in the instrumental condition. One explanation for this result is that subjects' nonverbal displays constituted additional evidence for their instrumentality. Their self-awareness of behaving confidently could have acted as feedback and

enhanced their sense of instrumentality. That males received higher confidence scores than females might reflect the impact of stereotypical associations between masculinity and confidence. The instrumental context could have intensified the effect of confidence on subjects' self-perceptions.

It is also possible that displays of confidence were simply by-products of finding appropriate experiences to share in the interview, and that the correlation between confidence scores and increases in masculinity scores is spurious. The activation of instrumental past experience could have produced both changes in masculinity ratings and confident interview behavior. It is difficult to evaluate the relative adequacy of these explanations without additional research.

There are still other individual differences that could have influenced subjects' posttest femininity and masculinity scores—for example, their experiences in the weeks intervening between pretest and posttest. It is therefore not surprising that the effect of condition in the total sample was small.

EXPERIMENT 2

Subjects' awareness of the camera in Experiment 1 could have made them uncomfortable and influenced their masculinity scores. We therefore conducted a second experiment using written materials rather than videotaped interviews. We additionally changed the position title, job description, and experience-related questions to evaluate the generalizability of our results.

The method of Experiment 2 was similar to that of Experiment 1 in most respects, except that it was recast as a job application study and was conducted with groups of subjects rather than on a one-by-one basis. Subjects filled out job applications on which they provided evidence of experience consistent with either expressive or instrumental job requirements, although the position was the same for all subjects. Posttest femininity and masculinity scores were compared by condition, relative to pretest scores.

Method

Subjects. Subjects were 111 introductory psychology students (47 males and 64 females) who volunteered in return for partial course credit. Subjects came from the same midwestern liberal arts institution as in Experiment 1. Their ethnic distribution was similar to that in Experiment 1, approximately 6% African American, 12% Asian American, 74% Caucasian, 6% Latino, and 2% of other backgrounds, according to class enrollment figures. At the beginning of the semester they rated themselves on

16 gender-related traits, as part of a battery of questionnaires. Several weeks later they were randomly assigned to one of two conditions (expressive or instrumental) to play the role of job applicant. Four did not complete the applications. Of the remaining subjects, 33 females and 22 males participated in the expressive condition, 30 females and 22 males in the instrumental condition. Subjects filled out job applications in groups of 10–25. Eight experimental sessions were conducted.

Pretest and Posttest Questionnaires. The pretest questionnaire contained the same BSRI items as in Experiment 1 and, also as in Experiment 1, subjects rated themselves on each item using 7-point Likert scales. Internal consistency (Cronbach's alpha) for pretest ratings of the 8 feminine items was .79 for females and .84 for males. Alphas for the 8 masculine items were .81 and .79, respectively. As in Experiment 1, the mean of feminine items was higher for females (M = 5.52, SD = 0.69) than for males (M = 5.04, SD = 0.86), $t(105) = 3.23, p < .01$, and there was no gender difference in masculinity scores (females M = 4.49, SD = 0.83 and males M = 4.74, SD = 0.84), $t(105) = 1.49, p > .10$.

The posttest questionnaire contained the same BSRI items as in Experiment 1 and subjects again rated themselves on each item using 7-point scales. The posttest questionnaire also requested self-ratings of subjects' performance on the job application. After the BSRI items subjects rated their comfort in filling out the application, the quality of their self-presentation, and their difficulty in answering the job application questions, also on 7-point scales.

Alphas for posttest ratings on the 8 feminine items were .88 for females and .81 for males. Alphas for the 8 masculine items were .82 and .78, respectively. As on the pretest, femininity scores were higher for females (M = 5.49, SD = 0.89) than for males (M = 5.17, SD = 0.77), $t(105) = 1.93, p < .06$, although this difference was only marginally significant, and there was no gender difference in masculinity scores (females M = 4.69, SD = 0.86 and males M = 4.89, SD = 0.86), $t(105) = 1.21, p > .20$.

Correlations between pretest and posttest scores were moderately strong. Correlations between femininity scores were .72 ($p < .001$) for females and .81 ($p < .001$) for males, and correlations between masculinity scores were .86 ($p < .001$) for females and .89 ($p < .001$) for males.

The alpha for performance ratings was only .60, due to low correlations between ratings of the application's difficulty and ratings on self-presentation and comfort. Performance scores were nonetheless computed from ratings of self-presentation, comfort, and difficulty (the latter ratings scored in a reverse direction) to maintain consistency with procedures from Experiment 1. Subjects generally reported moderately good performance (M = 4.33, SD = 1.20), although 4 subjects did not complete the ratings. An

ANOVA of performance scores by subject gender and condition yielded a marginally significant main effect of condition, $F(1,99) = 3.76$, $p < .06$. Expressive condition subjects reported better performance than instrumental condition subjects ($M = 4.56$, $SD = 1.25$ and $M = 4.09$, $SD = 1.12$, respectively). Effects of subject gender and its interaction with condition were statistically insignificant, $F(1,99) = 1.59$, $p > .20$ and $F(1,99) = 0.30$, $p > .50$ respectively.

Job Application Materials. The job description was for a college staff position that involved work on many different projects and therefore required flexibility and adaptability. The initial description was the same in both conditions but the specific job requirements differed by condition. In the expressive condition the job required someone who could relate well with, be responsive to, and appreciate the feelings of others. The person was also expected to exercise kindness and understanding, especially when dealing with complaints. In the instrumental condition the job required someone who could direct projects, work alone, and compete for limited backing from college officials. The person was also expected to present and defend projects at staff meetings.

All subjects began by recording their name and address and by briefly describing an experience in which they had to adjust to a novel situation. The questions that followed depended on subjects' assigned condition and were intended to elicit recall of experiences consistent with job requirements. In the expressive condition they responded to the following questions, intended to activate stereotypically feminine responses. Two examples of each experience were requested.

1. Briefly describe an experience in which you had to be responsive to the needs of others.
2. Briefly describe an experience in which you had to establish warm and friendly relationships with people you did not know.
3. Briefly describe an experience in which you had to take someone else's perspective in order to understand them and avoid hurting their feelings.
4. Briefly describe an experience in which you calmed someone who was upset.

In the instrumental condition they responded to the following questions, intended to activate stereotypically masculine responses. Two examples of each experience were requested.

1. Briefly describe an experience in which you took charge of a situation and provided guidance for other people.
2. Briefy describe an experience in which you worked on your own, without supervision.

3. Briefly describe an experience in which you presented and defended something you had worked on or believed in.
4. Briefly describe a situation in which you competed strongly for what you wanted.

Procedure. The procedure was similar to that of Experiment 1. Three weeks after the start of classes, the course instructor circulated signup sheets for the experiment. The experiment was described on the signup sheet as a study of impression making on job applications. Subjects would play the role of a job applicant so that their responses to job application questions could be compared. They were again expected to allow 40 minutes for the experimental session.

All experimental sessions took place in a classroom in the psychology building. Students sat at desks where they were administered the job application questionnaire by a student experimenter.

Upon arrival they were greeted by the student experimenter, who reminded them that this was a study of impression making on job applications. She distributed questionnaire packets that contained both the job description and job application questions, and then read the instructions on the first page of the packet. Subjects were told that they would read a job description and should imagine that they were applying for the job, which they really wanted. They would be filling out an application on which they provided experiences from their own lives that were relevant to the job. They were to answer the application questions as completely and spontaneously as possible, drawing on any of their life experiences. The instructions on the first page were the same for all subjects. The job description on the second page assigned them to condition and was systematically alternated for males and females after the first random assignment. The experimenter indicated that they would have as much time as they needed to complete the application.

When subjects had completed the job application they were given the posttest questionnaire. Afterward they wrote down what they thought had been the specific purpose of the experiment and then were debriefed. Most subjects thought we were investigating the kinds of experiences that people consider impressive, the personality characteristics of people with different kinds of experiences, or the personality correlates of impression-leaving. None expressed knowledge of the experiment's true purpose. The debriefing explanation was the same as in Experiment 1.

Results

Job Application Effects. As in Experiment 1, the hypothesis that gender-related self-ratings would increase from pretest to posttest in gender-

Table IV. Mean Pretest and Posttest Femininity and Masculinity Scores by Condition, Experiment 2[a]

	Expressive Condition		Instrumental Condition	
BSRI Score	Mean	(*SD*)	Mean	(*SD*)
Femininity				
Pretest	5.35	(0.85)	5.30	(0.75)
Posttest	5.53	(0.81)	5.18	(0.87)
Change	0.19*	(0.54)	-0.13	(0.57)
Masculinity				
Pretest	4.48	(0.79)	4.71	(0.88)
Posttest	4.69	(0.73)	4.86	(0.98)
Change	0.21***	(0.44)	0.15*	(0.41)

[a]*Note.* Change scores represent increases (on 7-point scales) from pretest to posttest, $^{*}p < .05$ and $^{***}p < .001$; interaction contrasts for changes in femininity and masculinity scores are significant at $p < .01$ in the instrumental condition and insignificant in the expressive condition, $p > .80$; expressive condition $n = 55$ and instrumental condition $n = 52$.

congruent conditions was tested in a gender rating (femininity vs. masculinity scores) by test (pretest vs. posttest) by gender (female vs. male) by condition (expressive vs. instrumental) mixed ANOVA, the first two factors within subjects and the second two between subjects. Although the predicted Gender Rating × Test × Condition interaction was statistically significant, $F(1,103) = 4.40$, $p < .05$, separate analyses by condition yielded a significant Gender Rating × Test interaction contrast in the instrumental condition, $F(1,51) = 8.70$, $p < .01$, but not in the expressive condition, $F(1,54) = 0.06$, $p > .80$. In the instrumental condition there was an increase in masculinity scores, $t(51) = 2.65$, $p < .05$, and a statistically insignificant decrease in femininity scores, $t(51) = -1.61$, $p > .10$. However, in the expressive condition both femininity and masculinity scores increased, $t(54) = 2.57$, $p < .05$, and $t(54) = 3.49$, $p < .001$, respectively (see Table IV).

A main effect of test, $F(1,103) = 8.77$, $p < .01$, was qualified by a Test × Condition interaction, $F(1,103) = 6.79$, $p < .01$; posttest scores were higher than pretest scores and the difference was greater in the expressive than in the instrumental condition.

Effects of Subject Gender. Also as in Experiment 1, the higher order Gender Rating × Test × Condition × Gender interaction was not statistically significant, $F(1,103) = 1.42$, $p > .20$, and indicated that the predicted Gender Rating × Test × Condition interaction occurred for both females and males. Other ANOVA results consistent with Experiment 1 included a main effect of gender rating, $F(1,103) = 41.16$, $p < .001$, that was qualified by a Gender Rating × Gender interaction, $F(1,103) = 8.98$, $p < .01$. The tendency for femininity scores to be higher than masculinity scores was stronger for females than for males. The main effect of gender rating was qualified also by a Gender Rating × Test interaction, $F(1,103) = 5.43$, $p <$

Table V. Mean Pretest and Posttest Femininity and Masculinity Scores by Condition and Level of Pretest Score, Experiment 2[a]

	Expressive Condition		Instrumental Condition	
	Pretest Scores		Pretest Scores	
BSRI Score	Low	High	Low	High
Femininty	(n = 25)	(n = 30)	(n = 28)	(n = 24)
Pretest	4.58	5.98	4.77	5.92
Posttest	5.00	5.98	4.77	5.65
Change	0.42_a**	0.00_b	0.00_b	-0.27_b
Masculinity	(n = 30)	(n = 25)	(n = 23)	(n = 29)
Pretest	3.92	5.16	3.94	5.32
Posttest	4.25	5.22	4.04	5.51
Change	0.33_a***	0.06_a	0.10_a	0.19_a*

[a]Pretest scores for femininity are used as predictors in comparisons of femininity scores and pretest scores for masculinity are used as predictors in comparisons of masculinity scores; change scores represent increases (on 7-point scales) from pretest to posttest, *$p < .05$, **$p < .01$, and ***$p < .001$; change scores with different subscripts in a row differ at $p < .05$, by the Tukey range test.

.05, indicating that the difference in femininity and masculinity scores was larger at pretest than at posttest.

Additional Correlates of Subjects' Changes in Femininity and Masculinity Scores. The average increases in femininity and masculinity scores were somewhat smaller in Experiment 2 than in Experiment 1, but the variation in change scores was comparable. The average increase in femininity scores was 0.19 (SD = 0.54) in the expressive condition and -0.13 (SD = 0.57) in the instrumental condition. The average increase in masculinity scores was 0.15 (SD = 0.41) in the instrumental condition and 0.21 (SD = 0.44) in the expressive condition.

Subjects' initial ratings of feminine and masculine qualities explained some of the variation in change scores. Correlations between pretest scores and change scores were negative in the expressive condition (femininity $r = -.39, p < .01$, and masculinity $r = -.42, p < .001$), as in Experiment 1, but were small and statistically insignificant in the instrumental condition (femininity $r = -.16, p > .20$, and masculinity $r = .02, p > .80$). The effect of pretest scores was evaluated further, as in Experiment 1, with repeated measures ANOVAs of pretest and posttest scores, using condition and level of pretest score (dichotomized by median splits), as predictors. The analysis of femininity scores yielded a Test × Femininity Level interaction effect, $F(1,103) = 9.24, p < .01$, and a Test × Condition interaction effect, $F(1,103) = 11.06, p < .001$. As indicated in Table V, these effects resulted in a significant increase in femininity scores in the expressive condition among subjects with low pretest femininity scores. Increases among these subjects were significantly greater than among the other three groups, as indicated

by post hoc Tukey tests. Femininity scores decreased marginally in the instrumental condition among subjects with high pretest scores.

The analysis of masculinity scores yielded a main effect of test, $F(1,103) = 19.90$, $p < .001$, and a Test × Condition × Masculinity Level interaction effect, $F(1,103) = 5.04$, $p < .05$. However, as indicated in Table V, although masculinity scores increased from pretest to posttest, there were no statistically significant differences among subgroups in the amount of change according to post hoc Tukey tests.

Although there were no objective measures of subjects' confidence in Experiment 2 as in Experiment 1, it was possible to evaluate effects of their performance self-ratings. As in Experiment 1, correlations between performance scores and change scores were small and statistically insignificant ($ps > .10$). Also as in Experiment 1, correlations between performance scores and pretest femininity and masculinity scores were positive in gender-congruent conditions, but the correlations were statistically insignificant, $r = .21$ and $r = .22$ respectively, $ps > .10$.

Main Components of Change in Femininity and Masculinity Scores. Item analyses were conducted to determine the main components of change in femininity and masculinity scores. Expressive condition ratings of feminine items and instrumental condition ratings of masculine items were compared at pretest and posttest to determine the extent to which each had changed. As in Experiment 1, repeated measures ANOVAs were conducted for each item, using level of pretest rating (dichotomized by a median split) as a predictor.

These analyses produced the same three patterns of results as in Experiment 1, although results for masculinity items were not as strong in this experiment (see Table VI). First, subjects with low pretest ratings demonstrated statististically significant rating increases on most items, especially in the expressive condition. Second, rating increases were stronger for some items than for others, although the strongest increases differed from those in Experiment 1. In this experiment, expressive condition results were strongest for the items, affectionate, eager to sooth, and tender, and instrumental condition results were strongest for the items, assertive and has leadership ability. Third, subjects with relatively high pretest ratings on individual items tended to rate themselves lower on those items at posttest, although the differences were statistically significant in only 3 out of 16 comparisons. As in Experiment 1, ratings on the item, feminine increased in the expressive condition among subjects with low pretest ratings. A similar increase in ratings on the item, masculine, did not occur in the instrumental condition. In fact, subjects with high pretest ratings in this condition decreased their ratings significantly at posttest.

That femininity and masculinity scores both increased in the expressive condition suggested that some aspect of this condition might have enhanced

Table VI. Changes in Ratings of Gender Items in Gender-Congruent Conditions, Experiment 2[a]

Gender Item	Change in Rating		Test Effect F	Test × Pretest Interaction, F
	Low Pretest	High Pretest		
	Expressive condition			
Affectionate	0.76***	0.04	11.51***	8.51**
Compassionate	0.63*	-0.16	0.36	8.38**
Eager to soothe	0.75**	-0.19	2.60	11.97***
Feminine	0.44*	-0.30	0.80	6.64*
Gentle	0.16	-0.50*	0.84	5.58*
Sensitive	0.71*	-0.06	3.38	8.54**
Sympathetic	0.63*	-0.26*	0.00	12.56***
Tender	1.21***	0.20	19.36***	14.61***
	Instrumental condition			
Acts as leader	0.53*	0.17	6.52*	2.21
Aggressive	0.48*	-0.47	0.49	5.71*
Assertive	0.67**	0.23	10.58**	3.03
Dominant	0.36	-0.19	0.24	3.02
Has leadership ability	0.65***	-0.13	3.53	16.76***
Independent	0.22	-0.06	0.68	0.31
Masculine	0.23	-0.65**	2.40	10.49**
Willing to take a stand	0.41*	0.04	4.69*	2.83*

[a]Pretest ratings of femininity items are used as predictors in comparisons of femininity ratings and pretest ratings of masculinity items are used as predictors in comparisons of masculinity ratings; *n*s at each pretest level vary somewhat, depending on the dichotomization of pretest ratings; change scores represent increases (on 7-point scales) from pretest to posttest, $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

subjects' self-worth in general. To test this possibility we compared pretest and posttest self-ratings on gender-neutral items that had moderately consistent responses. Two principal components analyses, one of pretest responses and one of posttest responses, each yielded a first factor on which the items of friendly, helpful, reliable, and tactful had moderately strong loadings. Internal consistencies (alpha) at pretest and posttest were .67 and .63, respectively. Mean ratings on these items were computed as pretest and posttest niceness scores and compared in a test (pretest vs. posttest) by condition (expressive vs. instrumental) mixed ANOVA, the first factor within subjects and the second between subjects. The analysis yielded no statistically significant effects (*p*s > .10), indicating that condition effects were specific to femininity and masculinity scores.

Discussion

Results from Experiment 2 partially replicated those from Experiment 1. As in Experiment 1, femininity scores increased from pretest to posttest

in the expressive condition and this effect occurred primarily among subjects with initially low femininity scores. Unlike results from Experiment 1, masculinity scores increased significantly in both conditions. These results suggest that some aspect of the expressive condition was empowering for subjects. Perhaps a sense of interpersonal efficacy was imparted to expressive condition subjects as they recalled experiences of responsiveness, relationship management, and perspective taking. The main components of expressive condition changes in masculinity scores were ratings on the items, acts as leader, assertive, and independent, which emphasize self-sufficient aspects of job performance. When considered in relation to the main components of expressive condition changes in femininity scores—ratings on the items, affectionate, eager to sooth, and tender—these could reflect interpersonal management skills required by the expressive condition job description, for example, the ability to influence behavior while exercising kindness, warmth, and understanding. The main components of instrumental condition changes in masculinity scores were ratings on the items, assertive and leadership ability. These items emphasize only the directive aspects of job performance required by the instrumental condition job description.

It is also possible that some aspect of the instrumental condition inhibited subjects' sense of instrumentality. Perhaps the requirement to defend their work constrained subjects' sense of independence and masculinity. Perhaps it was more difficult for instrumental condition subjects than for expressive condition subjects to retrieve the required experiences. Instrumental condition subjects were asked to recall experiences in which they took charge of situations, presented and defended their beliefs, and competed strongly for their goals. Such experiences may be more extreme and accrue later in adolescence than do those based on interpersonal relations. Even those subjects with high pretest masculinity scores might have been challenged by job application questions in the instrumental condition. This could be one reason why pretest masculinity scores did not correlate significantly with change scores in this condition and why ratings on the item, masculine, decreased significantly in the instrumental condition among subjects with high pretest ratings. If the requested instrumental experiences were uncommon even for subjects with high pretest masculinity scores, then their performance ratings and changes in masculinity would have been closer to those with lower pretest scores. That performance scores were somewhat lower in the instrumental condition than in the expressive condition supports the idea that instrumental condition questions required less common and more difficult to retrieve experiences than did expressive condition questions. Moreover, subjects in Experiment 2 were asked for two examples of each experience instead of only one, as in Experiment 1. The requirement to retrieve two, possibly less available expe-

riences could have inhibited increases by conveying limited evidence for related characteristics. Schwarz, Strack, Bless, Klumpp, Rittenauer-Schatka, and Simons (1991) provide evidence that subjects' sense of assertiveness can be enhanced by asking them to recall 6 instances of self-assertion, but that asking them to recall 12 instances diminishes their sense of assertiveness. If job application questions were in fact more demanding of instrumental condition subjects than of expressive condition subjects, then these could have have facilitated increases in masculinity and femininity scores in the expressive condition but hindered their increase in the instrumental condition. Future research could evaluate the effects of placing increasing demands on memory for gender-related experiences. Would excessive demands for evidence of feminine and masculine qualities reduce subjects' self-ratings on these and related qualities? Are there other ways in which memory constrains the effects of context on gendered aspects of self?

Although the rating changes from this experiment were not as consistent as those from Experiment 1, they nonetheless replicate a general tendency for subjects with low pretest ratings to increase their ratings at posttest, especially in gender-congruent conditions. We think that these results reflect genuine shifts in self-assessment rather than attempts at impression management, because performance ratings were again not very high and because there was evidence of decreased ratings among subjects with high pretest scores. We think that these trends would not have been obtained had subjects been trying to manipulate an appropriate impression at posttest.

GENERAL DISCUSSION

Taken together, results from these experiments provide evidence that context can affect gender-related self-evaluation by activating related experience in memory. In both experiments activation of expressive and instrumental experiences influenced self-ratings on stereotypically feminine and masculine qualities. Since increases in self-ratings occurred primarily among subjects with initially low femininity and masculinity scores, these effects are probably due to shifts in the salience of gender-related experiences for identity. Subjects with initially high self-ratings must have been typically aware of these experiences, whereas subjects with initially low ratings must have been less aware. Activation of related experience in memory would consequently benefit the latter more than the former. This result was obtained in both experiments, although the activation of instrumental experiences in Experiment 2 produced weaker results. In this case, the type of experience demanded might have been uncommon and extreme, even

for those with chronically accessible instrumental experiences. Activation of less common and more extreme experiences might therefore increase the self-ratings of even those with initially high scores. However, if contextual demands do not activate the required experience—for example, if the experience is too uncommon or extreme—then a decrease in related self-ratings might be obtained. Failure to retrieve the required experiences can be taken as evidence for the absence of these qualities.

Although increases in femininity and masculinity scores were obtained in both experiments, their main components differed. Increases in femininity scores emphasized subjects' compassionate, sensitive, and sympathetic qualities in Experiment 1, but their affectionate, soothing, and tender qualities in Experiment 2. We think that these differences reflect differences in the experiences requested of subjects and the context of their recall. In Experiment 1, subjects recalled situations in which they were responsive to people while encouraging the sharing of feelings; in Experiment 2 they recalled situations in which they were responsive while exercising kindness, warmth, and understanding. Similarly, increases in masculinity scores reflected different qualities of identity in Experiments 1 and 2. In Experiment 1, increases in masculinity scores emphasized subjects' aggressive, assertive, dominant, and independent qualities; in Experiment 2 they emphasized (in the instrumental condition) their assertiveness and leadership ability. These qualities are congruent with the experiences requested of subjects—in Experiment 1 for guidance of others and initiative in decision making, and in Experiment 2 for taking charge of situations. Primary components of rating changes thus appear to have stemmed from the job-related experiences that subjects recalled whereas secondary components appear to have stemmed from a general network of gender associations that the experiences activated.

Another way to interpret results from these experiments is with reference to expressive and instrumental opportunities for identity development in females and males. Sex role stereotypes have traditionally polarized the expressive and instrumental roles associated with women and men, and a major social explanation for sex differences in expressiveness and instrumentality, when they are found, has been related to sex differences in opportunities to develop these competencies (e.g. Eagly, 1987). Although results from this experiment are based on single, short-term experiences only, they are nonetheless instructive about the potential impact of chronic differential access to such opportunities. Self-ratings on feminine qualities increased significantly more in contexts that activated expressive strengths; self-ratings on masculine qualities increased significantly more in contexts that activated instrumental strengths. Females and males did not differ in this regard. If subjects' reflections on gender-related experiences and abili-

ties in only one interview can shift their self-perceptions somewhat, then routine reminders of similar experiences and abilities might have even larger effects. If occupational and other life contexts provide routine reminders of expressive and instrumental abilities to women and men equally, they are likely to promote the homogenization of gender self-perceptions. If they provide routine reminders of expressive abilities to women and instrumental abilities to men, they are likely to promote gender polarization.

The expressive and instrumental constructs activated by contexts of everyday life can thus play an important part in development and maintenance of gender-related identities. If women are channeled into one type of context and men into another, then the difference in experience is likely to be reflected in their identities. That we consistently obtained sex differences in femininity scores but not in masculinity scores is consequently instructive. We think this pattern reflects continued asymmetry in women's and men's access to expressive experiences and the interpersonal skills that are thereby fostered.

There are at least three objections that could be raised regarding the validity of our results. First, it could be argued that the manipulations and posttest questionnaire cued subjects to the purpose of the experiment since both were gender related. The job interview and application questions requested information about expressive and instrumental experiences and subjects' self-ratings might have reflected their desire to be cooperative. We have tried to provide evidence that subjects were not cued by the manipulations and posttest ratings. There were no subjects in either experiment who expressed knowledge of our purpose. Even those who thought we were investigating personality correlates of interview and application performance suggested that our interest was in the effects of personality on performance rather than the effects of performance on personality. Moreover, not a single subject mentioned gender or gender-related reasons for the study.

It could also be argued that our results reflect subjects' beliefs that the experiment was not yet over when they filled out the posttest questionnaire. The increases in their self-ratings on condition-congruent gender items could therefore reflect continued attempts to create an appropriate impression. We have provided evidence that subjects' performance ratings were not consistent with increases in their self-ratings and that they generally did not rate their performance high. In addition, subjects with high pretest ratings demonstrated decreases on the posttest. We think that neither of these results would have been obtained had subjects been trying to manipulate their impressions at posttest.

Finally it could be argued that subjects' changes in self-assessment were fleeting, with little consequence for their subsequent behavior. This

could be a valid objection, particularly in regard to subjects who return to environments that do not activate consistent expectations. However, our major purpose in these experiments was to show that gender self-assessments could be shifted by activation of gender-related experiences in memory, and we were able to demonstrate these shifts, at least temporarily, by asking subjects to retrieve expressive and instrumental experiences. Future research might evaluate the behavioral and longerterm consequences of activating gender-related experiences. For example, will subjects respond to a person with problems more compassionately after retrieving an expressive experience and more directively after retrieving an instrumental one? How long will it take for the effects of a retrieved experience to decay? Can its effects be prolonged by multiple, consistent instances of retrieval? Our results were also based on relatively small and homogeneous samples of white undergraduates. Future studies could try to replicate the results and evaluate their relevance for subjects of different ethnicities and ages. Such studies would provide insight into the variable strength of gender associations and the contexts of everyday life that promote their development.

REFERENCES

Ballard-Reisch, D. & Elton, M. (1992). Gender orientation and the Bem Sex Role inventory: A psychological construct revisited. *Sex Roles, 27*, 291-306.

Bargh, J. A., Lombardi, W. J., & Higgins, E. T. (1988). Automaticity of chronically accesssible constructs in person x situation effects on person perception: It's just a matter of time. *Journal of Personality and Social Psychology, 55*, 599-605.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155-162.

Bem, S. L. (1993). *The lenses of gender.* New Haven, CT: Yale University Press.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80*, 389-407.

Dailey, D. M., & Rosenzweig, J. (1988). Variations in men's psychological sex role self-perception as a function of work, social and sexual life roles. *Journal of Sex and Marital Therapy, 14*, 225-240.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46*, 991-1004.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review, 94*, 369-389.

Eagly, A. H. (1987). *Sex differences in social behavior: A social role interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Geis, F. L. (1993). Self-fulfilling prophecies: A social psychological view of gender. In A. E. Beall & R. J. Sternberg (Eds.), *The psychology of gender.* New York: Guilford Press.

Helgeson, V. S. (1994). Prototypes and dimensions of masculinity and femininity. *Sex Roles, 31*, 653-682.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, NJ: General Learning Press.

Jones, E. E., Berglas, S., Rhodewalt, F., & Skelton, J. A. (1981). Effects of strategic self-presentation on subsequent self-esteem. *Journal of Personality and Social Psychology, 41*, 407-421.

Lawrance, K., Taylor, D., & Byers, E. S. (1996). Differences in men's and women's global, sexual, and ideal-sexual expressiveness and instrumentality. *Sex Roles, 34*, 337-357.

Lorber, J. (1994). *Paradoxes of gender.* New Haven, CT: Yale University Press.

Markus, H., Crane, M., Bernstein, S., & Siladi, M. (1982). Self-schemas and gender. *Journal of Personality and Social Psychology, 42*, 38-50.

Morawski, J. G. (1985). The measurement of masculinity and femininity: Engendering categorical realities. *Journal of Personality, 53*, 196-223.

Rhodewalt, F., & Agustsdottir, S. (1986). Effects of self-presentation on the phenomenal self. *Journal of Personality and Social Psychology, 50*, 47-55.

Riggs, J. M., Monach, E. M., Ogburn, T. A., & Pahides, S. (1983). Inducing self-perceptions: The role of social interaction. *Personality and Social Psychology Bulletin, 9*, 253-260.

Rosenzweig, J. M., & Dailey, D. M. (1991). Women's sex roles in their public and private lives. *The Journal of Sex Education and Therapy, 17*, 75-85.

Rubini, V., & Antonelli, E. (1992). Self-gender schemata and the processing of social information. *European Journal of Personality, 6*, 359-370.

Schwarz, N., Strack, F., Bless, H., Klumpp, G., Rittenaur-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 1-8.

Sedikides, C., & Skowronski, J. J. (1990). Towards reconciling personality and social psychology: A construct accessibility approach. *Journal of Social Behavior and Personality, 5*, 531-546.

Spence, J. T. (1984). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. In B. A. Maher & W. B. Maher (Eds.), *Progress in experimental personality research* (Vol. 13) New York: Academic Press.

Srull, T. K., & Wyer, R. S., Jr. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology, 38*, 841-856.

Stangor, C. (1988). Stereotype accessibility and information processing. *Personality and Social Psychology Bulletin, 14*, 694-708.

Uleman, J. S., & Weston, M. (1986). Does the BSRI inventory sex roles? *Sex Roles, 15*, 43-62.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender and Society, 1*, 125-151.