Double Standards in the Evaluation of Men and Women*

MARTHA FOSCHI
University of British Columbia

This article presents the results from two expectation-states studies on gender and double standards for task competence. The emergence of such standards under several experimental conditions is investigated. In both studies, men and women, participating in opposite-sex dyads, worked first individually and then as a team in solving a perceptual task. As predicted, results from Experiment 1 show that although subjects of both sexes achieved equal levels of performance, women were held to a stricter standard of competence than men. This difference was more pronounced when the referent of the standard was the partner rather than self. Experiment 2 investigates the extent to which the double standard is affected by level of accountability for one’s assessments. Results show a significant difference by sex of referent of standard when accountability was low, but not when it was increased. In both studies, measures of perceived competence in self and in partner reflected reported standards, as predicted. Theoretical and practical implications of these findings are discussed.

THEORETICAL BACKGROUND

In task-oriented groups, the processes whereby individuals assign competence to each other have crucial consequences for their future interaction. The standards used to judge that competence, in turn, play a key role in these processes: because such standards are norms specifying the level and type of outcome required to infer ability, the inference varies depending on the standard used. For example, a score of 70% is sufficient evidence for ability if the standard is 60% or higher, but that score becomes an unconvincing performance if the standard is at least 80%. A similar example may be constructed for lack of ability. Because the assignment of task competence is directly related to achieving status and influence in a group, it is important to understand how standards are set.

In many tasks, standards are not clearly defined beforehand. As a result, factors unrelated to the task (e.g., interpersonal liking, control motives, or group prejudices) can affect what standards will be used. The research presented here links standards to group prejudices. Specifically, the objective is to study whether, under certain conditions, different standards for competence are used to evaluate members of different social categories—even when they perform at the same level. The work focuses on gender as the basis of such a double standard. In other words, given that women often are assigned lower levels of ability than men, to what extent is this assignment the result of the application of a double standard, which is stricter for the female performers? What are some of the variables that affect when this occurs? The expectation states research program, particularly its branch on status characteristics, provides the theoretical background.

Expectation states theory investigates the development of power and prestige hierarchies in task groups (Berger et al. 1977; Berger, Wagner, and Zelditch 1985; Wagner and Berger 1993; Webster and Foschi 1988). The theory has a long-standing tradition, and empirical findings provide strong support for its predictions. (For assessments, see, for example, Deaux 1985; Wiley 1986.) A central concept in this program is that of a “status characteristic,” any valued attribute implying task competence. Such characteristics consist of at least two states (e.g., either high or low level of mechanical ability, either limited or extensive formal education), one of which is evaluated more positively than the.
other. These attributes are also defined as ranging from specific to diffuse, depending on their perceived applicability. A specific characteristic is associated with well-defined performance expectations; a diffuse characteristic carries, in addition, predictions about performance in a wide, indeterminate variety of tasks. In many societies, gender,1 ethnicity, and socioeconomic level constitute diffuse status characteristics. Thus women, for example, often are expected not only to be inferior to men in various specific skills, but also to be inferior in general competence. (For discussions and reviews of gender as a status attribute, see Ridgeway 1993; Wagner 1988.)

The assignment of task ability may be made directly (from performance evaluations) or indirectly (on the basis of status characteristics). This assignment, in turn, results in performance expectations—stable beliefs about how well a group member will carry out a given task in the future. Such expectations are theoretical constructs mediating the relationship between status characteristics and the power and prestige order of the group. This order is defined in terms of the following interrelated behaviors: unequal distribution in the offer and acceptance of performance opportunities, the type of evaluations received, and the rates of influence exerted. Expectation states theory focuses on explaining how this order originates and is maintained while interventions based on the theory identify ways of redressing the inequalities.

The present research is concerned with situations in which expectations are based on both status and performance evaluation. Specifically, I am interested in those cases in which actors differ in status but perform at the same level (either well or poorly). For theoretical as well as applied reasons, the most interesting case occurs when these two items of information constitute very different grounds for competence. Thus let us assume that the performance evaluations are objective (i.e., the operation of biases has been blocked) and that the status characteristic in question is diffuse. According to expectation states theory, the resulting expectations will be a combination of all the information that the actor considers relevant to the situation.

As an example, I assume gender to be the diffuse attribute. Thus, when a man succeeds, two consistent pieces of information (status and level of performance) are available and a definite inference of competence results. Success by a woman, however, represents an inconsistent combination; therefore a weaker inference of ability ensues. On the other hand, failure will be viewed as a consistent outcome for a woman but not for a man. Consequently this outcome will be interpreted as indicating lack of ability more strongly in the female performer than in her male counterpart. For a more fully detailed presentation of these hypotheses, including scope conditions, see Berger et al. (1977); experimental tests (e.g., Pugh and Wahrman 1983; Wagner, Ford, and Ford 1986) provide clear empirical support.

Foschi (1989) proposes both an elaboration and an extension of the above formulation. The proposal incorporates ideas from (1) attribution work examining how the perceived causes of success and failure are affected by the performer's membership in a social category (Deschamps 1983; Hansen and O'Leary 1985; Whitley, McHugh, and Frieze 1986) and (2) expectation states research on the effects of standards on the interpretation of performance outcomes (Foschi and Freeman 1991; Foschi, Warriner, and Hart 1985). At the core of this theory is the notion that differences in inferred ability persist in spite of equal performances because different standards for competence are applied to higher- and to lower-status members. It is useful to classify standards in terms of their strictness (see Foschi and Foddy 1988). A strict standard for ability requires more evidence of competence (e.g., a larger number of correct responses, attained over a larger number of more difficult tasks) than does a lenient standard. Conversely, a strict standard for lack of ability tolerates less evidence of incompetence than does a lenient standard. Moreover, a situation involves a “universal” standard if the same set of requirements is applied to all performers; if this is not the case, “double” or even “multiple” standards obtain. In Foschi (1989) I propose that, under certain conditions, a status characteristic that differentiates the performers into two classes activates the use of double standards for competence and for lack of competence, both of which are stricter.

---

1 I use the term sex to refer to biological differences between men and women, and gender for cultural aspects of these differences. See Foschi, Lai, and Sigerson (1994) for a discussion of this point.
for lower-status performers. The higher the inconsistency between status and outcome, the stricter the standard.

I propose the status characteristics and multiple standards theory for a situation in which a person (self) works on a joint task with a partner (other). The propositions are stated from self's point of view and apply to any status characteristic—specific or diffuse, individually as well as in combination with others. Here I present a summary for the case of gender, and assume the following scope conditions: (1) self values the task and is motivated both to do it well and to arrive at correct assessments of the two performers' competence; (2) self is aware that the partner is of the opposite sex (i.e., sex of performer is a salient factor in the situation); (3) self treats gender as a diffuse status characteristic (i.e., an indication of women's inferior competence); (4) self knows the results of each person's prior individual performance (assessed by a third party) and believes these to be unbiased, but no previously set and agreed-upon standards are available by which to infer ability (or lack thereof) from those results; (5) both persons perform at the same level; and (6) self has no other grounds on which to base assessments of task competence.

Foschi (1989) specifies that self may define the task in one of the following ways: (1) masculine, (2) feminine, (3) explicitly dissociated from gender, or (4) not explicitly defined in relation to gender. It is predicted that in (1), self will tend to activate a stricter standard for the female performer, for competence as well as lack of competence. This also will occur in (4), where gender and task become related through “status generalization.” In this process, a status attribute becomes relevant to the task at hand unless there is specific information to the contrary. In such a case, the double standards will be less pronounced than when the task is masculine, but they will manifest themselves nevertheless. No double standards benefiting the male performer are predicted for (2) or (3).

Once a double standard is activated, it affects the degree of ability that is inferred. The application of a more lenient standard to the man ensures that more ability is assigned to him, regardless of level of performance, than to the woman with the same record. Thus double standards contribute to the maintenance of the initial, status-based assignment of competence and are another component of the power and prestige order of the group. The practice is both subtle (because it does not involve either devaluing or overvaluing the performance directly) and not necessarily conscious (because an actor does not have to formulate such standards explicitly in order to use them). For an alternative but compatible formulation on double standards, also proposed within expectation states theory, see Foddy and Smithson (1989); also see Biernat and Manis (1994) and Biernat, Manis, and Nelson (1991) for their work on stereotypes and shifting standards.

Multiple standards for competence that benefit the higher-status performer are common in a variety of everyday task settings, ranging from informal groups to formal work contexts. The social psychological literature contains several descriptions of their opera-

---

2 I use the expression double standard to refer either to one person who is assessing the performances by himself or herself and by a specific partner in a particular context, or to the average requirement applied to self and to other by a number of individuals. I use the plural double standards, however, if I wish to emphasize that a double standard of different magnitude may be activated, either by each of several individuals or by the same person when making assessments across various circumstances and/or partners.

3 Other types of gender-based double standards also exist, such as differences in codes of morality and criteria for physical attractiveness. Similarly, double (or even multiple) standards for competence may be based on attributes other than status characteristics (e.g., level of interpersonal liking, or personal qualities such as friendliness). In general, a multiple standard refers to the use of different criteria for assessing the same trait (or behavior or performance) in different categories of people. The present research is concerned only with gender-based double standards in inferring ability. For a review of various types of multiple standards, see Foschi (1992).

4 Also, in some conditions, the lower-status person is treated with a more lenient (but not explicitly stated) standard than the higher-status counterpart, and is told that his or her performance is a sufficient demonstration of ability when in fact it is not. Such statements are rarely the result of a genuine error in assessment but rather are made for other purposes; for instance, to avoid conveying a poor evaluation or to meet quotas. For this reason, such patronizing standards fall outside the scope of this article. They are worth investigating, however, because depriving the lower-status actor of a truthful appraisal is yet another practice through which the status quo can be maintained. For discussions of this and other types of more lenient standards for the lower-status person, see, for example, Blalock (1979: chap. 4) and Epstein (1970:978).
tion, particularly regarding gender, as well as a number of studies providing indirect evidence of this practice (see Foschi 1992 for a review). To my knowledge, however, only Biernat et al. (1991), Foddy and Smithson (1989), and Foschi (1989) have proposed formal theories about such standards. (Also see Jasso and Webster 1995 for a theoretical analysis combining ideas from the multiple standards and the distributive justice formulations.) The work presented here is part of an ongoing research program designed to test Foschi’s (1989) theory. Two studies already have been completed in this program, namely Foschi et al. (1994) and Foschi, Sigerson, and Lembesis (1995); both address double standards in an extended context where self is an evaluator of others but not a performer. Results from the first study show that, given an average outcome by a male and a female performer, men but not women favored the former. In the second study, subjects of both sexes gave the male performer an advantage over his female counterpart when the two had achieved an average outcome, but this advantage was reversed when their performances were outstanding. In both studies double standards were measured indirectly. The two experiments reported here, on the other hand, assess double standards directly and investigate their occurrence in the original self-and-other context of expectation states research. The objectives are (1) to test key hypotheses on the activation of double standards and the resulting levels of perceived competence, and (2) to investigate the role of two additional selected factors in those processes: whether the standard is set for self or for other, and the level of accountability for self’s assessments.

EXPERIMENT 1

OBJECTIVES

This study concerns a situation in which two factors, assumed to affect the magnitude of the double standard, are held constant. These are sex linkage of task and level of performance. Thus the task was defined as masculine, and both actors performed at an average rather than an extreme level. These two design decisions maximize the likelihood that distinct double standards will be activated: should this practice not appear under these conditions, it would be unlikely to appear at other levels of these variables. Thus, regarding sex linkage, I considered it crucial to be able to demonstrate first that double standards indeed emerged under the facilitating conditions of an explicitly masculine task. The results then could be used later as a baseline for other sex-linkage conditions. Similarly, I chose an average level of performance on the assumption that extreme levels, because they constitute a more definite indication of either ability or lack of ability, would be more likely to escape status effects than would average outputs. (See Foschi et al. 1995. For discussions of this point in relation to gender effects in particular, see Epstein 1970; Kalin and Hodgins 1984; Lott 1985; Nieva and Gutek 1980; Wallston and O’Leary 1981.)

Let us then assume opposite-sex dyads, and consider them from each person’s point of view. Thus, either self is a man and his partner is a woman, or self is a woman and her partner is a man. The task context meets all the other conditions specified so far, and both persons perform successfully. The hypotheses to be tested concern sex of performer and that person’s role in the dyad (either self or other). I propose the following two hypotheses:

Hypothesis 1. (Gender). When a man and a woman perform at the same level of success as their respective opposite-sex partners, the woman’s performance will tend to be assessed with a stricter standard for ability than the man’s. As a result, the level of competence inferred about the woman will be lower than that inferred about the man.

A major assumption of expectation states theory is that assignments of competence are relative to the actors in a given situation; that is, they result from a comparison between these actors (Berger et al. 1977: chap. 4; Berger et al. 1985). Hence, in the dyads studied here, the man should experience the same competence advantage over his female partner whether he occupies the role of self or of other. In other words, although differences due to role may exist (stemming, for

---

5 Although Foschi (1989) largely overlaps with the original status characteristics theory, the two differ slightly in scope. The inclusion of scope conditions (4) and (5) makes the former more limited than the latter, although relaxing these in future work should be a relatively straightforward matter. On the other hand, by including situations where self is an evaluator of others but not a performer, Foschi (1989) extends status characteristics theory.
example, from self-enhancing motives), status is assumed to be the stronger factor under the specified scope conditions (also see Foschi et al. 1994). Thus:

**Hypothesis 2.** (Role). The process described in Hypothesis 1 will occur whether the actors under consideration occupy the role of self or of other in their respective dyads.

In other words, the female performer will be assessed by herself as well as by her male partner with a stricter ability standard than will her male counterpart. Conversely, the male performer will be assessed by himself as well as by his female partner with a more lenient ability standard.

**METHOD**

**Subjects and Experimenters**

Subjects were 72 men and 72 women, all undergraduates from the faculties (schools) of Arts and Science at the University of British Columbia. Average ages (with standard deviations in brackets) were 18.58 [0.96] for the men and 18.80 [1.11] for the women. Participation was voluntary and each person was paid $8 for the session. A pool of subjects was obtained by recruiting in large first- and second-year classes; those students who had taken courses in psychology beyond an introductory level and/or who had participated in social psychology experiments were excluded as prospective subjects.

Each participant was teamed with another of the opposite sex, and teams were assigned at random to one of three conditions: (1) higher score for self than for other; (2) higher score for other than for self; or (3) no score for either person. The study thus was a 2 (sex of subject and partner) x 3 (feedback condition) design, with 24 subjects per cell. Each session was conducted by one of two female research assistants of similar age. Special attention was paid to maintaining uniformity of appearance and delivery of instructions across sessions.

**Procedures and Materials**

For comparability, procedures were a variant of the standardized experimental situation developed for expectation states research (Berger et al. 1977:43–48). Instructions and questionnaires either were adapted from reliable instruments used in previous research or were pretested for this study. The two participants were seated individually at adjacent stations equipped with personal computers said to be linked to each other. The stations were separated by a partition and subjects were precluded from both seeing and talking with each other. The experimenter stated that the purpose of the study was to investigate performance on a “contrast sensitivity” task in two simulated work environments: an individual and a team setting. Subjects were informed that only the members of the research staff would see their individual responses, and that their names would be kept confidential. Instructions were worded so as to motivate participants to do well (i.e., to be “task-oriented”). Each team was informed that its two members were of the same year and faculty at the university but of different sex. Contrast sensitivity was described as a recently discovered ability and of high potential value to a variety of tasks. Reliable research was said to have shown it to be mainly intuitive and relatively specific. Thus the subjects heard that “although no significant relationship has so far been established between it and attributes such as mathematical skill or artistic ability, men have been found to be generally far more accurate than women at solving contrast sensitivity problems.”

The contrast sensitivity task consists of several trials. On each trial, subjects view a rectangular area covered to about the same extent by smaller rectangles of two different colors. Subjects must decide which of the two colors is predominant in the overall pattern. The task is actually ambiguous to allow for the measurement of acceptance/rejection of influence. It is a reliable instrument that has been used extensively in expectation states studies.

A computerized version developed explicitly for the present work was used here. Brief instructions for the task as well as the visual stimuli themselves were presented on the screen. The stimuli were white and red on a black background. The computer program gave the subject 10 seconds to look at each pattern, prompted him or her for a response, and, after five seconds, showed the next pattern. Subjects were assured that, although the task appeared difficult, there was always a correct answer. Computerization of the contrast sensitivity task offers many advantages, including (1) eliminating the possibility of
error by the experimenter in communicating prearranged feedback to the subjects, recording their responses, and computing influence rates, and (2) more control by the experimenter over the presentation time for each pattern as well as over the time elapsed between patterns. It also enhances the credibility of the cover story and results in a highly engaging task.

An overview of the experiment appears in Table 1. During the first part of the study, participants in the experimental conditions ((1) through (4)) worked individually on the task and made decisions on 20 patterns. At the end of this series, the scores obtained by the two persons (trial by trial as well as overall) appeared on both computer screens. They showed either 11 correct answers by self and 13 by the partner, or the reverse. Next, each person received a printout of these scores. Then each subject completed a written questionnaire that included several manipulation checks and a measure of his or her own standards for the higher-scoring person “to definitely have contrast sensitivity ability.”

I chose the scores of 11 and 13 to indicate an equally average level of success by both persons, for the reasons presented earlier. Such a level also has the advantage of giving subjects a wider range of choices for setting their own ability standard than would a more definite success. The latter, in addition, could have been interpreted by the subjects as implicitly informing them of what that standard should be. On the other hand, because reporting that both persons received exactly the same scores would likely have created suspicion, the two scores were made to differ slightly. This difference also focuses attention on the better performer and highlights the fact that, for this person, outcome and sex represent either a consistent or an inconsistent combination. As discussed earlier, degree of consistency is assumed to contribute to the use of a double standard.

One item in the questionnaire prompted subjects to record both scores; the immediately following item asked them to state their own ability standard for the better performer, as a percentage of correct answers required. I assumed that, given the scores received, the subjects would most likely activate and be concerned with this standard. Also, the question about it would follow logically (and quite unobtrusively) from the previous question on scores.6

During the second series of trials, the two subjects were instructed to work as a team and to try to arrive at a correct choice in each trial. The intention of thus creating a “collective orientation” in each person was to make him or her assess the relative competences of self and other. Subjects were told that a team would be awarded two points whenever both persons were correct. As an added incentive for teamwork, each of the six teams with the most points would win a $20 prize.

For consistency with other expectation states experiments with a similar design, the contrast sensitivity task varied slightly during this phase. It now involved two patterns per trial; subjects had 10 seconds to decide which of the two rectangular areas contained more white. The same ability as in the first phase was said to be involved. After a subject made a decision, the partner’s “choice” was relayed. The feedback was manipulated to result in 20 disagreements and five agreements on the initial choices. Each subject then made his or her own final selection. In the disagreement trials, this entailed either remaining with self’s initial choice or changing it to agree with the partner’s. The former decision is referred to as an s-response. The proportion of s-responses operationalizes influence rejection, a variable which, in this setting, is associated reliably with perceived competence in self and in other.

At the conclusion of this series, subjects completed a second questionnaire that included further manipulation checks and one additional measure of relative competence. This instrument also served to assess any misunderstandings and/or suspicions regarding the procedures. Next, subjects were interviewed individually to check further on these issues, and then were debriefed.

The control conditions ((5) and (6) in Table 1) excluded the individual performance phase and the corresponding scores. Because double

---

6 If subjects had been asked to set an ability standard for each of two persons with closely matching scores, they probably would have produced mechanical answers indicating the same standard for both. On the other hand, the small degree of difference between these scores did not justify asking for standards for ability in one performer and for standards for lack of ability in the other.
standards are assumed to be applied to such scores, the question about standards was not asked either, and subjects worked directly in the team setting. In all other respects, procedures were the same as in the experimental conditions. The purpose of the control groups was to provide baseline measures of competence inferred from only two items of information, namely sex of self and of other, and sex linkage of task.

In sum, I expected the results to show that the female better performer would be judged by a stricter ability standard than her male counterpart, regardless of their roles in the dyad. The same sets of scores thus would be interpreted differently, depending on sex of self and of other. As a result, measures of perceived competence would reflect a higher level for the male better performer than for his female counterpart, again regardless of role. This sex difference would be larger in the control groups than in the experimental conditions, because in the latter the effects of gender-related information would be moderated by the equality in the two persons’ performance levels.

RESULTS

On the basis of the information obtained through the postexperimental questionnaire and interview, 15 subjects (eight men and seven women) were excluded from the analysis. This figure represents 10.4% of the total number of participants. Rejection rules were conservative and were formulated beforehand, and this percentage compares well with the exclusion rates of similar experiments. The excluded subjects may be classified as follows: seven volunteered well-articulated suspicions about the conformity aspects of the study, six misunderstood crucial sections of the instructions and/or the written questions, and eight showed a clear lack of task orientation and/or collective orientation. (Several subjects fell into more than one of these categories.) The analysis that follows includes only the 129 retained subjects.7

Manipulation Checks

The postexperimental questionnaire consisted of a variety of items, some of which were fillers added to maintain the realism of the cover story. Included in this instrument were six five-point bipolar scales used to assess the subjects' perceptions of the task. The means and (in brackets) standard deviations corresponding to these items were: creative (1)—routine (5): 2.79 [1.01]; important (1)—unimportant (5): 2.96 [0.88]; easy (1)—difficult (5): 3.93 [0.89]; intuitive (1)—learned (5): 2.04 [0.80]; and masculine (1)—feminine (5): 2.97 [0.49]. Analysis of variance for each attribute showed no statistically significant effects from either type of dyad8 or feedback condition, as anticipated. The figures regarding the first four attributes are within expected ranges and indicate that

7 In line with commonly accepted use in statistical analysis, I call a result significant if \( p \leq 0.05 \), of marginal or borderline significance if \( p > 0.05 \) and \( \leq 0.10 \), and nonsignificant otherwise. In the latter case, I omit the presentation of details from the statistical tests. The expression \( p = .000 \) results from rounding off to three decimal places.

8 I use sex composition of dyad, type of dyad, sex of self and other, and sex of subject and partner interchangeably. Because the partner was always of the opposite sex, for brevity sometimes I simply refer to sex of subject.
the procedures had created perceptions of the task as intended. The data on the fifth item, however, show that, contrary to the instructions, subjects on average considered the ability to be gender-neutral. I examine this point in the discussion section below.

The questionnaire also included similar scales to assess the subjects’ dispositions toward their task. Averages and standard deviations were: interested (1)—uninterested (5): 2.35 [0.91]; motivated (1)—unmotivated (5): 2.43 [0.88]; and involved (1)—uninvolved (5): 2.29 [0.88]. As can be seen, values for all three items again fell within expected ranges. Also as anticipated, neither independent variable had any statistically significant effects. Finally, responses to two other questionnaire items indicate that all subjects recalled exactly the scores received by themselves and by their partner, and correctly identified the sex of the latter.

**Dependent Variables**

I begin by examining the results on standards for the higher-scoring person, shown in Table 2. As predicted, ANOVA on these data shows a significant main effect from sex of better performer (or referent of the question on standards) \( M \) (male referent) = 68.05, \( M \) (female referent) = 73.45, \( F \) (1, 82) = 8.49, \( p = .005 \), while neither this person’s role nor the interaction was significant. Simple contrasts reveal a strong effect of sex of referent in the “other” conditions \( F \) (1, 82) = 5.75, \( p = .019 \) but a marginally significant effect in the “self” groups \( F \) (1, 82) = 3.02, \( p = .086 \). Neither of the other two contrasts (between “self” and “other” conditions, by sex of referent) was significant, as expected.

**Perceived competence** in self and other was measured primarily through level of rejection of influence from the partner, calculated as the proportion of times a subject did not change his or her initial answer. The higher this figure, the higher self’s perceived competence advantage over other. In addition to this behavioral measure, I obtained one auxiliary indicator: perceived task ability in self relative to the partner. I used a bipolar scale, with responses ranging from “self much worse” (1) to “self much better” (5). Results from both measures appear in Table 3.

Analysis of variance on rejection of influence shows significant results as expected: from type of dyad \( M \) (male subject) = .597, \( M \) (female subject) = .486, \( F \) (1, 123) = 31.93, \( p = .000 \), from feedback condition \( M \) (subject worse than partner) = .476, \( M \) (subject better than partner) = .592, \( M \) (no scores) = .562, \( F \) (2, 123) = 12.95, \( p = .000 \), and from the interaction \( F \) (2, 123) = 3.59, \( p = .031 \). Simple contrasts on the comparisons of direct relevance to the hypotheses reveal that, for each feedback level, women accepted more influence from their partner than did men from theirs. Results from these tests were as follows: Conditions (1) and (2): \( F \) (1, 123) = 5.84, \( p = .017 \); Conditions (3) and (4): \( F \) (1, 123) = 3.75, \( p = .055 \); Conditions (5) and (6): \( F \) (1, 123) = 29.68, \( p = .000 \). Findings from the two control groups indicate that, in the absence of performance information, there was a clear sex difference in expectations: men felt superior to women in task ability, whereas

<table>
<thead>
<tr>
<th>Condition (as per Table 1)</th>
<th>( N )</th>
<th>Sex and Role of Referent</th>
<th>( M )</th>
<th>( SD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Male subject scoring worse than female partner</td>
<td>22</td>
<td>F</td>
<td>Other</td>
<td>72.50</td>
</tr>
<tr>
<td>(2) Female subject scoring worse than male partner</td>
<td>23</td>
<td>M</td>
<td>Other</td>
<td>66.39</td>
</tr>
<tr>
<td>(3) Male subject scoring better than female partner</td>
<td>21</td>
<td>M</td>
<td>Self</td>
<td>69.86</td>
</tr>
<tr>
<td>(4) Female subject scoring better than male partner</td>
<td>20</td>
<td>F</td>
<td>Self</td>
<td>74.50</td>
</tr>
</tbody>
</table>
women held the opposite view. Because the subjects perceived the task as gender-neutral rather than masculine, what these results show is the single effect of sex composition of dyad. Thus they establish that gender was a diffuse status characteristic for the subjects.

Data from the other two comparisons show, as anticipated, that this sex effect was tempered but not eliminated by equality in the levels of performance.

Results on the auxiliary measure show the same pattern as those on rejection of influence, again as expected. On relative ability of self and partner, ANOVA indicates significant effects from type of dyad ($M$ (male subject) = 2.94, $M$ (female subject) = 2.41, $F$ (1, 123) = 28.54, $p = .000$), from feedback condition ($M$ (subject worse than partner) = 2.35, $M$ (subject better than partner) = 3.05, $M$ (no scores) = 2.66, $F$ (2, 123) = 18.11, $p = .000$), and from the interaction ($F$ (2, 123) = 7.35, $p = .001$). Results from internal analyses are as follows: Conditions (1) and (2): $F$ (1, 123) = 7.46, $p = .007$; Conditions (3) and (4): $F$ (1, 123) = 0.32, $p = .572$; Conditions (5) and (6): $F$ (1, 123) = 36.10, $p = .000$.

In sum, results from the manipulation checks show that overall, subjects’ perceiving them the same scores. Table 3, however, shows that subjects formed different expectations in Conditions (1) and (3), and in Conditions (2) and (4). Thus it is possible to view each of these four conditions as either confirming or disconfirming (albeit not definitely) the expectations formed in the corresponding control groups. One then can compare these findings with the predictions that the original status characteristic formulation makes for such a situation, and with results such as those obtained by Wagner et al. (1986). For the current experiment, where relative success and relative failure are ambiguous, that theory would predict that differences between control and disconfirmation conditions will be significant, whereas those between control and confirmation conditions will not. The results from contrasts on rejection of influence support all four of these predictions, as follows: Conditions (3) and (5) (male subjects, confirmation): $F$ (1, 123) = 0.84, $p = .362$; Conditions (2) and (6) (female subjects, confirmation): $F$ (1, 123) = 1.20, $p = .276$; Conditions (1) and (5) (male subjects, disconfirmation): $F$ (1, 123) = 17.27, $p = .000$; Conditions (4) and (6) (female subjects, disconfirmation): $F$ (1, 123) = 6.28, $p = .014$. Contrasts on reported relative ability yield highly similar results. In addition, although on both measures of perceived competence one would have expected the results from Condition (3) to be higher than those from Condition (5), the former are still high enough to support all hypotheses involving them. Perhaps, given the difficulty of the task, ambiguous failure relative to the partner communicates inferiority much more definitely than the corresponding ambiguous success communicates superiority.

---

9 This experiment was designed to equate the two actors at an average level of performance without actually...
tions of the task and of their involvement in it had been created as expected. Findings on standards indicate that the predicted sex-of-referent effect is more pronounced for other than for self, and data on perceived competence are consistent with this difference in every respect.

DISCUSSION

In my view, two aspects of the results merit special attention, namely those on (1) sex linkage of task and (2) standards for self and for other.

Although the instructions identified the task as masculine, subjects' responses indicate that instead, on average, they viewed it as gender-neutral. Clearly, the sex linkage manipulation was not strong enough. In future work it would be useful to reword the instructions to place more emphasis on that linkage, and perhaps even to display "evidence" of it on graphs and tables. Notice, however, that the experiment thus unintendi‐dly constitutes a test of the hypotheses in the burden of proof case, and therefore results in a stricter test of the double standards formulation.\(^{10}\)

Regarding standards, the significant main effect from sex of referent supports Hypothesis 1, while the nonsignificant effects from role of referent and from the interaction support Hypothesis 2. Contrasts add more information on how these variables are related; the results from three of the four contrasts (Conditions (1) and (2), (1) and (4), and (2) and (3)) are clearly as expected, while the contrast between Conditions (3) and (4) is of borderline significance. All things consid‐ered, then, the hypotheses receive substantial support.

The difference in the support received by Hypothesis 1 in the "self" and the "other" conditions may be due to the following: although status processes may be equally strong for self and for other, they could have been tempered by additional processes in the case of self. When subjects had to set standards for the partner (Conditions (1) and (2)), gender was the only item of information differentiating the two referents. Thus the distinction between the male and the female partner was a sharp one. On the other hand, when subjects had to set standards for themselves (Conditions (3) and (4)), other information in addition to sex of self may have become salient. For example, subjects may have recalled their own performances in the past on similar tasks—something they could not do regarding the partner. As a result, the additional information should have made the standards for self less differentiated than those for other.

Although no data were collected in the experiment that could be used to test this interpretation, the following arguments make it highly plausible. In the first place, manipulation checks had revealed no further differences between "self" and "other" conditions. A thorough examination of all questionnaire items, including fillers, yielded the same result. Second, work on attribution theory reports a tendency to overestimate the importance of situational factors in assessments of self, and a corresponding overestima‐tion of dispositional factors in assessing others (Jones and Nisbett 1972; Monson and Snyder 1977; L. Ross 1977; M. Ross and Fletcher 1985). Empirical findings support the existence of this tendency ("the fundamental attribution error") under a variety of conditions. Accordingly it would be more likely that other, rather than self, would be judged in stereotypical terms in this study. Moreover, here the tendency would have been enhanced by the fact that subjects indeed were constrained by the available information to assess the partner in those terms. Comments (both written and oral) volunteered by several subjects render informal support to this view. In future research, it would be relatively simple to add questions on the factors that subjects considered in making their assessments.

Finally, the following implication of the results is worth noting. Findings on perceived competence shown on the top four rows of Table 3 reflect, condition per condition, those obtained on standards. When standards are

\(^{10}\) Alternatively, one could argue that the subjects believed the task to be masculine, as instructed, but reported it to be gender-neutral because this would have enhanced them in the eyes of the experimenter. This argument is highly speculative, however, because in doing so, the subjects also would have been contradicting the instructions of the person they were trying to impress. Moreover, I found no indications that any subjects believed the experiment concerned whether they accepted the stated sex linkage. Finally, results from the rest of the manipulation checks do not support the view that impressing the experimenter was paramount in the subjects' minds. Therefore I see no reason not to accept their responses on perceived sex linkage.
differentiated clearly by sex of referent (as in the “other” conditions), there is a corresponding difference in the two measures of competence. When the difference between standards decreases (as in the “self” conditions), the perceived competence measures become correspondingly closer. Thus both sets of findings support the double standards formulation.

EXPERIMENT 2

OBJECTIVES

A large number of studies on status characteristics and expectation states have used the standardized experimental setting, as I do here. This has served to maintain comparability across studies and thus has increased the cumulativeness of the research. One common feature of most of those studies has been that (as in Experiment 1) subjects have been prevented from seeing each other. This has served to control for the formation of expectations on the basis of characteristics other than those intentionally manipulated. In addition, subjects either have been led to believe or have been told explicitly that they would have no face-to-face interaction with the partner. Such features of the situation make the subject less accountable for his or her assessments of self and other; this factor, in turn, should foster status generalization. By “accountability” I mean the extent to which self anticipates having to justify his or her actions. To my knowledge, variations in this factor have not received adequate attention in expectation states research. In most of this work, accountability has been treated as a variable that is useful to keep constant—a decision which is based on design rather than on theoretical considerations. Foschi et al. (1994) and Foschi et al. (1995) discuss it explicitly as a factor in status generalization, but again, both studies keep the variable constant. Self-awareness, a related topic, is investigated by Martin and Sell (1983).

In general, expectation states theory makes no assumptions about self’s awareness of the principles used in processing status information—that is, the theory proposes that their operation may or may not be conscious (Berger et al. 1985:37). This position also extends to the use of double standards (Foschi 1989:63). Increased accountability, on the other hand, should make people more aware of these principles—that is, it should make them think more about what they say and do. There is some evidence from expectation states research itself that such awareness decreases status effects. Thus Martin and Sell (1983) created either higher or lower performance expectations for self than for other in a setting that either included or did not include self-focusing stimuli (mirrors, camera, television monitor). Results indicate that subjects in the former groups showed the effects of status information less than those in the latter.

Outside the status characteristics literature, several studies have examined the effects of accountability on information processing. The results most directly relevant to the research reported here show that subjects give greater thought to received information when they expect to be accountable for their responses than when they do not (Tetlock 1983, 1985; Tetlock and Kim 1987). Thus, these findings reveal that accountable subjects consider the different aspects of an issue more carefully than do nonaccountable subjects, pay more attention to inconsistent items of information, and attempt to integrate them. Accountability also has been found to increase the use of strategies intended to create favorable impressions in others. (See, for example, Miller and Schlenker 1985 for a study showing that participants’ attributions are less egotistical in public than in private conditions.) In expectation states research, both of these types of processes (namely cognitive and impression management) should translate into a decrease in the use of stereotypical responses as accountability increases.

The first objective of this study is to incorporate “low accountability for one’s assessments” as a scope condition of the theory on activation of double standards. The study thus serves as a partial replication of the first experiment by retesting Hypothesis 1 with the explicit addition of that scope condition. I focus on those groups from Experiment 1 in which a double standard emerged most clearly, namely the “other” groups. In this way Experiment 2 also assesses the reliability of that finding.

The second objective is to explore the limits of that scope condition by investigating whether it can be extended to include a medium level. (Note that the intention is not to study the effects of clearly contrasting values of this variable, such as “low” and “high” levels, for example.) The increase in
accountability should narrow the gap between the two standards: the standard applied to the female performer should decrease, while the standard set for the male performer should increase. Perceptions of relative competence should reflect this change in standards.

In this study I thus examine whether a medium level of accountability eliminates the double standard or merely reduces its magnitude. Results also would serve as indirect evidence of the extent to which the subjects treat the use of a double standard as legitimate under the circumstances. That is, medium accountability should not be sufficient to deter the activation of double standards if these are deemed to be a proper (i.e., culturally acceptable) way of processing information. (For analyses of the role of legitimation processes in the formation of expectations, see Ridgeway and Berger 1986; Ridgeway and Walker 1995.)

**METHOD**

*Subjects and Experimenters*

Subjects were 48 men and 48 women. Average ages and standard deviations were 18.65 [1.14] for the men and 18.81 [1.23] for the women. The subject pool was the same as that used for Experiment 1; Experiment 2 was conducted shortly after the earlier study.

Each participant was teamed with another of the opposite sex, and teams were assigned at random to one of two conditions, either low or medium accountability. The study therefore was a 2 (sex of subject and partner) x 2 (level of accountability) design, with 24 subjects per cell. Each session was conducted in the same way as in the previous experiment. For this reason and because the same subject pool was used, I did not consider it necessary to replicate the control groups of Experiment 1. An overview of the design appears in Table 4.

**Procedures and Materials**

The only differences in the procedures used in the two experiments concern the accountability manipulation. Thus, in Experiment 2, some paragraphs were incorporated into the instructions as follows. Subjects in the low-accountability groups were instructed "For the sake of confidentiality, please do not write your name on these forms" and were told "You can be assured that no one, including your partner, will be seeing your responses to the questionnaires, with the exception of the research team. In addition, at no time will you be required to justify your responses to these questionnaires or your choices on the tasks, nor will you be meeting with your partner during the course of the study." In the medium-accountability conditions, the instructions stated "At the conclusion of the study you will be meeting your partner to compare and discuss your answers on these questionnaires and on the two tasks you will be performing. And please remember to write your full name legibly on the front page of each questionnaire." (The instructions used in Experiment 1 implied that no interaction with the partner would take place, thus making all the conditions of that study highly comparable to the low-accountability groups of the present experiment. Note also that, as stated earlier, the medium accountability conditions were not designed to represent a maximum increase in this factor. That could have been achieved, for example, by telling the subjects that they were expected to justify all of their answers to both the partner and the research team.)

**Table 4. Overview of Experiment 2**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sex of Subject</th>
<th>Sex of Partner</th>
<th>Scores Received by Subject (Maximum: 20)</th>
<th>Level of Accountability</th>
<th>Standards for Ability in Higher Scoring Person</th>
<th>Perceived Competence in Subject and in Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>M</td>
<td>F</td>
<td>11</td>
<td>Low</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(2)</td>
<td>F</td>
<td>M</td>
<td>11</td>
<td>Low</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(3)</td>
<td>M</td>
<td>F</td>
<td>11</td>
<td>Medium</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(4)</td>
<td>F</td>
<td>M</td>
<td>11</td>
<td>Medium</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Manipulation checks on this variable were included in the second questionnaire and the interview.

RESULTS

An analysis of the information obtained through the postexperimental instruments resulted in the exclusion of 11 subjects (five men and six women), or 11.5% of the total number of participants. The same rejection rules as in Experiment 1 were applied, and the resulting exclusion rate is similar to that of the earlier study. The reasons for exclusion were as follows: suspicion (three subjects), misunderstanding the instructions (four subjects), and lack of task orientation and/or collective orientation (five subjects). (One person was classified in two of these categories.) The following analysis includes only the 85 retained subjects.

Manipulation Checks

As in Experiment 1, the postexperimental questionnaire yielded information about the subjects' perceptions of the task. Means and standard deviations on these items were as follows: creative (1)—routine (5): 2.85 [1.15]; important (1)—unimportant (5): 2.87 [1.06]; easy (1)—difficult (5): 3.51 [1.18]; intuitive (1)—learned (5): 2.31 [1.13]; and masculine (1)—feminine (5): 2.85 [0.76]. Subjects' dispositions toward the task were also assessed as in the earlier study. Averages and standard deviations on these measurements were: interested (1)—uninterested (5): 2.40 [1.10]; motivated (1)—unmotivated (5): 2.45 [1.06]; and involved (1)—uninvolved (5): 2.23 [1.00]. As expected, ANOVA for each of these eight variables showed no statistically significant effects from either type of dyad or level of accountability. In addition, a comparison with Experiment 1 in these respects shows a close overall correspondence between the two sets of findings. (The only noticeable difference occurs on perceived difficulty; subjects in this study assessed the task as less difficult than those in Experiment 1. I do not attach any special importance to this difference, however, given (1) the high level of similarity between the two studies in all other respects and (2) the fact that in both experiments, subjects nonetheless perceived the task to be difficult.)

The postexperimental results also show that all subjects recalled exactly the scores received by themselves and by their partner, and correctly identified the sex of the latter. Finally, one 5-point bipolar scale served as a manipulation check of accountability, assessed in terms of perceived level of privacy of the situation. This item read as follows: “Considering who will have access to your responses (e.g., partner, research assistant, other researchers), rate how private (1)—public (5) you perceived these responses to be.” Means and standard deviations for the low-accountability groups were as follows: Condition (1) (male subjects): 2.17 [1.10]; Condition (2) (female subjects): 1.98 [0.76]. Values for the medium-accountability groups were: Condition (3) (male subjects): 2.88 [1.20]; Condition (4) (female subjects): 3.13 [1.14]. As expected, ANOVA results show a significant main effect from this manipulation (M (low accountability) = 2.08, M (medium accountability) = 3.00, F (1, 81) = 24.46, p = .000), whereas neither the main effect from sex of subject nor the interaction is significant. Simple contrasts indicate that accountability level made a significant difference for men (F (1, 81) = 7.21, p = .009) as well as for women (F (1, 81) = 18.50, p = .000). In other words, subjects in the low-accountability groups perceived the context of the study to be significantly more private than did those in the medium-accountability groups.

Dependent Variables

As in Study 1, I begin by examining the results on standards for the higher-scoring person, shown in Table 5. As expected, ANOVA indicates a significant main effect from sex of referent of standards (M (male referent) = 66.77, M (female referent) = 71.98, F (1, 81) = 7.19, p = .009), whereas neither accountability nor the interaction is significant. Sex of referent had a significant effect only when accountability was low (F (1, 81) = 5.16, p = .026).

Perceived competence in self and in partner was measured primarily through rejection of influence. As Table 6 shows, men rejected more influence from the partner than did women. ANOVA indicates, as expected, that the main effect from sex of subject was significant (M (male subject) = .543, M (female subject) = .452, F (1,81)) = 9.68, p = .003), whereas the effects from either
accountability or the interaction were not. Again, internal analyses show that sex of subject resulted in a significant difference in only the low-accountability groups \((F(1, 81) = 8.66, p = .004)\). Perceptions of relative ability in self and in other were assessed as well; as Table 6 also shows, these varied by sex of subject and level of accountability. ANOVA indicates a significant main effect from both of these variables, while the interaction was not significant. Results for sex of subject were: \(M\) (male subject) = 2.39, \(M\) (female subject) = 2.12, \(F(1, 81) = 5.73, p = .019\). For the second factor the analysis shows the following: \(M\) (low accountability) = 2.38, \(M\) (medium accountability) = 2.14, \(F(1, 81) = 4.40, p = .039\). Simple contrasts reveal borderline significance from sex of subject (Conditions (1) and (2): \(F(1, 81) = 2.97, p = .089\); Conditions (3) and (4): \(F(1, 81) = 2.77, p = .10\)) and nonsignificant effects from accountability.

In sum, all manipulation checks (except, again, for sex linkage of task) are as expected. So are the results on standards and rejection of influence: the wider the gap in standards, the larger the difference in influence rejection. Moreover, data from the low-accountability conditions closely replicate those from Conditions (1) and (2) of Study 1, as anticipated. Findings on perceived relative ability include some unexpected results, and these are discussed below.

Table 6. Experiment 2: Perceived Competence in Self and in Partner

<table>
<thead>
<tr>
<th>Condition (as per Table 4)</th>
<th>(N)</th>
<th>(M)</th>
<th>(SD)</th>
<th>(M)</th>
<th>(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Male subject scoring worse than female partner, low accountability</td>
<td>21</td>
<td>.552</td>
<td>.156</td>
<td>2.52</td>
<td>.602</td>
</tr>
<tr>
<td>(2) Female subject scoring worse than male partner, low accountability</td>
<td>21</td>
<td>.429</td>
<td>.146</td>
<td>2.24</td>
<td>.436</td>
</tr>
<tr>
<td>(3) Male subject scoring worse than female partner, medium accountability</td>
<td>22</td>
<td>.534</td>
<td>.108</td>
<td>2.27</td>
<td>.550</td>
</tr>
<tr>
<td>(4) Female subject scoring worse than male partner, medium accountability</td>
<td>21</td>
<td>.474</td>
<td>.131</td>
<td>2.00</td>
<td>5.48</td>
</tr>
</tbody>
</table>
DISCUSSION

Two sets of findings from this experiment are of special interest. First, results indicate that the gender-based double standard, although still present, loses its magnitude when accountability is increased. Thus, although a significant sex-of-subject main effect is found in standards as well as in rejection of influence, in both cases the significance is traced to the low-accountability conditions only. These findings thus indicate that gender-based double standards are activated and used most definitely at that level of accountability. It also may be that low accountability is a scope condition for the operation of the other status processes studied in expectation states theory. Furthermore, the reduction in the gap in standards following an increase in accountability suggests that the legitimacy of using gender as an indication of competence is somewhat limited. This should not be surprising in view of the major changes in the status value of gender differences that have occurred in the US and Canada in recent years.

Second, results on relative ability, on the other hand, are somewhat different from those on rejection of influence: whereas the former is significantly lower when accountability increases, a comparable finding is not observed in rejection of influence. In my view, this is related to differences between the two measures. In line with other expectation states research, I consider rejection of influence to be the primary indicator of perceived competence. The measure has the advantages of being both an unobtrusive behavioral response (and therefore more likely to reflect a person's true assessments) and highly reliable (because it is the result of observations made over several trials). Perhaps a significant main effect from accountability was observed in relative ability but not in influence rejection because of the subjects' heightened awareness, in Conditions (3) and (4), of what was being assessed through the auxiliary measure. As a result, they may have considered it desirable to lower the values they reported. In view of this difference, in the next section I base my conclusions about perceived competence mainly on the results for rejection of influence.

GENERAL DISCUSSION AND CONCLUSIONS

In this article I present the results from two studies on the activation and use of gender-based double standards in assessing competence. The theoretical ideas both elaborate and extend aspects of the expectation states formulation, while the results represent experimental demonstration of the existence and direction of such double standards and their consequences. Thus, overall, subjects in opposite-sex dyads report a stricter standard for a female than for a male partner, even though both persons performed at the same level and the difference in scores between self and other was also equal. The gap between the standards is reflected in differences in perceived competence, and findings on these two variables show a high level of consistency both within and across the two experiments.

Considered together, the studies identify two conditions under which the double standard is most pronounced: (1) when the referent is an "other" who differs from self only with respect to gender, and (2) when subjects feel that accountability for their assessments is low. The double standard emerges even though the subjects report (against the experimenter's instructions) that they consider the task to be gender-neutral. It is also worth noting that the gender information was presented with minimal cues: subjects knew that the partner was of the opposite sex, but no items reinforcing this communication (such as hobbies and first names) were added.

The design of these studies included a computerized version of the experimental setting commonly used in expectation states research. As a result, findings from the present experiments can easily be compared with other work on expectation states, as I have done throughout this article. The question about the standards for competence that I used represents an addition to the questions commonly asked in that work. Standards were measured directly by requesting subjects to state the percentages of correct responses that they required for a definite inference of ability. The question was understood readily by the participants and also proved to be an unobtrusive instrument: not even any of those who were subsequently rejected because of suspicion were aware that this question was central to the study.

These two experiments investigate opposite-sex dyads. Subjects were informed specifically that the partner was of the opposite sex, whereas on other factors they either were
told that they were equal to the partner or were given no information. These features of the situation create optimum conditions for gender to become a salient variable. In the future, a more thorough investigation of sex effects could be achieved by studying contexts in which self is paired with a partner of either the same or the opposite sex. Useful information also could be obtained from conditions in which sex of partner is unknown because these conditions would serve to address questions about the source of the double standard: do men set strict standards for women, do women set lenient standards for men, or do both practices occur?

These studies provide key information on the operation of double standards. There is, however, still much work to do to understand this practice. For example, it would be worthwhile to study its occurrence under other values of sex linkage of task (particularly the “feminine” task) and other levels of performance. It would also be of interest to test whether gender-based double standards for lack of ability constitute another status-maintenance practice. Finally, I should emphasize that the theory behind this work concerns not only gender but any status attribute. Thus it would be important to test the theory in other cases, involving a single status characteristic as well as several. After all, in many everyday situations it is common to be affected by more than one status variable. Women, for example, experience lower status and its consequences to different degrees depending on the combined effects of gender and other attributes, especially ethnicity, age, social class, and level of formal education.

Here I have examined aspects of the use of gender-based double standards for competence at the interpersonal level. As discussed earlier, competence standards are basically requirements imposed on a person’s performance. When different standards are applied to different categories of people, that practice contributes to the maintenance of the interactional status quo. It would be surprising, however, if the consequences of this practice were limited to its direct effect on interpersonal assignment of competence. I suspect that double standards also are at the root of several other important phenomena. Two possible effects of differences in requirements are the gap in earnings between men and women with equal qualifications, and the gender differences in the amount and intensity of work-related stress they experience. In future research, it would be useful to investigate links between interpersonal double standards and processes conceptualized at other levels of analysis, from individual to organizational and macrostructural.

Understanding double standards is also important for reasons other than their theoretical significance. For example, suppose it becomes widely recognized that members of a social category are, or have been, commonly disadvantaged in settings involving evaluation, particularly through the use of a stricter ability standard. How can this situation be reversed? This is exactly the question that affirmative action programs address. Simply introducing an explicitly more lenient standard for one category in order to make up for past wrongs is not the answer. Such an approach would lead to an inference of lesser ability among members of that category; this, in turn, would perpetuate the inequality. Well-designed affirmative action programs are not based on the implementation of such a “reverse double standard.” Instead they include demonstrations of ability that leave no doubts about the superior quality of the chosen applicants.

REFERENCES


Epstein, Cynthia F. 1970. “Encountering the Male Establishment: Sex-Status Limits on Women’s
Careers in the Professions.” American Journal of Sociology 75:965–82.


**Martha Foschi** is Professor of Sociology at the University of British Columbia. Her research interests focus on status processes, particularly on gender and ethnicity as sources of biases in ability evaluation. The two studies reported here are part of a series of experiments she has recently carried out on gender-based double standards for competence.