

Gender- and Race-Based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups

Monica Biernat and Diane Kobrynowicz
University of Kansas

Stereotypes may influence judgment via assimilation, such that individual group members are evaluated consistently with stereotypes, or via contrast, such that targets are displaced from the overall group expectation. Two models of judgment—the shifting standards model and status characteristics theory—provide some insight into predicting and interpreting these apparently contradictory effects. In 2 studies involving a simulated applicant-evaluation setting, we predicted and found that participants set lower minimum-competency standards, but higher ability standards, for female than for male and for Black than for White applicants. Thus, although it may be easier for low- than high-status group members to meet (low) standards, these same people must work harder to prove that their performance is ability based.

Social stereotypes may affect judgments of individual group members in one of two ways. In an *assimilative* mode, stereotypes may lead individuals to judge group members consistently with the group expectation. For example, ambiguously aggressive behavior is more likely to be considered hostile if committed by a Black than a White target (Allport & Postman, 1947; Duncan, 1976; Sagar & Schofield, 1980; see also Devine, 1989). On the other hand, stereotypes may also produce *contrastive* effects. For example, because men are expected to earn more money than women, a man is less likely to be called financially successful than a woman who objectively earns the same income (Biernat, Manis, & Nelson, 1991).

At an intuitive level, these phenomena are both readily comprehensible and plausible, yet they are also, undeniably, contradictory. This apparent paradox is effectively articulated in Carter's (1993) recent book, *Reflections of an Affirmative Action Baby*. Carter (1993) wrote the following about the difficulties faced by Black Americans as they seek to prove their worth in school and the workplace: "Our parents' advice was true: We really do have to work twice as hard to be considered half as good [as Whites]" (p. 58). In this statement, he acknowledges the assimilative effects that stereotypes can have: Because of low expectations for Blacks, equal performance from a Black and White target is not sufficient to garner Blacks a positive evaluation; more is needed from a Black (see, e.g., Davis &

Watson, 1982; Pettigrew & Martin, 1987; Yarkin, Town, & Wallston, 1982).

But just a few pages earlier in the book, in a statement that is more consistent with the contrastive effects of stereotypes, Carter wrote the following: "[I am] somewhat skeptical . . . that Black people of intellectual talent have a harder time than others in proving their worth. My own experience suggests quite the contrary, that like a flower blooming in winter, intellect is more readily noticed where it is not expected to be found" (p. 54). In this case, low expectations are presumed to lead to very positive evaluations of competent Blacks (see also Jackson, Sullivan, & Hodge, 1993; Jussim, Coleman, & Lerch, 1987).

How can these apparently contradictory phenomena be understood? In the broader judgment literature, a number of models have been introduced to account for when assimilative as opposed to contrastive effects are obtained. For example, Herr (1986) suggested that feature overlap between a primed category or context and a target concept is critical to the judgment outcome: Extreme primes lead ambiguous targets to be contrasted from the primed category, but moderate primes lead to assimilation, presumably because the prime can incorporate the target in the latter, but not the former, case (see also Goldstone, 1995; Herr, Sherman, & Fazio, 1983; Higgins, 1989; Hilton & von Hippel, 1990; Manis, Nelson, & Shedler, 1988). Similarly, Schwarz and Bless's (1992) inclusion-exclusion model suggests that when a context or category is included in one's representation of a target stimulus, assimilation to the context is the result; when this information is excluded from the representation, contrast results. Martin's set/reset model (Martin, 1986; Martin, Seta, & Crelia, 1990) suggests that assimilation is the default outcome when concepts are used to interpret target information, whereas contrast results from a more effortful resetting process, in which contextual information is partialled out of a judgment if deemed irrelevant. Petty and Wegener (1993; Wegener & Petty, 1995), on the other hand, suggested that either assimilation or contrast may be the natural default mode of a context and that perceivers may flexibly correct for these judgment biases. In still another variation on this theme, Manis and

Monica Biernat and Diane Kobrynowicz, Department of Psychology, University of Kansas.

This research was supported by National Institute of Mental Health Grant R29MH48844. We thank Martha Foschi, John Skowronski, and Chris Crandall for their helpful comments on a draft of this article, and Denise Epstein, Shabnam Hashemy, Harun Hazim, Shawna Hillary, Jaren Higginbotham, and Claire Noller for their assistance with data collection.

Correspondence concerning this article should be addressed to Monica Biernat, Department of Psychology, 426 Fraser Hall, University of Kansas, Lawrence, Kansas 66045-2160. Electronic mail may be sent via the Internet to biernat@stat1.cc.ukans.edu.

Paskewitz (1984) argued that contrast and assimilation effects are the result of independent processes that operate in parallel: Contextual cues give rise to expectations, which produce assimilation, but also serve as reference or comparison points, which produce contrast (see also Manis, Biernat, & Nelson, 1991; Newman & Benassi, 1989).

Each of these perspectives suggests that whether assimilation or contrast occurs is dependent on characteristics of both the context and the target stimuli to be judged. For example, in the domain of stereotypes, an incidental prime of a concept category (e.g., *dependent*) leads to assimilative effects, but only when the target to be judged is a member of a category to which the primed concept applies (e.g., a woman; Banaji, Hardin, & Rothman, 1993). These models do not, however, shed any clear light on the phenomenon that Carter (1993) described: the *same* "competent" Black individual in a single context may in one case be assimilated to, and in another case contrasted from, the negative stereotype of Blacks. In other words, holding target attributes and context constant, it is still possible that either assimilation or contrast might result. What is missing from the analysis, and what can illuminate the issue, is a consideration of the type or nature of the judgment that a perceiver is asked to make. More specifically, we make a distinction between judgment questions that lead the respondent to make a broad-based inference about a target versus those that require an assessment of the target's minimum standing on an attribute. This distinction will become more clear through a consideration of two models of social judgment that provide the basis for the present work: the *shifting standards model* (Biernat, 1995; Biernat & Manis, 1994; Biernat et al., 1991), and *status characteristics theory* (Berger, Fisek, Norman, & Zelditch, 1977; Berger, Rosenholtz, & Zelditch, 1980; Berger, Wagner, & Zelditch, 1985; Foddy & Smithson, 1989; Foschi, 1992; Foschi, Lai, & Sigerson, 1994; Webster & Foschi, 1988).

Inferring Minimal and Broad-Based Competence

The shifting standards model focuses on the role stereotypes play in creating judgment standards against which individual members of stereotyped groups are evaluated. This model suggests that when judging group members on stereotype-relevant dimensions, perceivers use within-category reference points to make these estimates. Assuming, for example, that perceivers hold the stereotype that "women are less competent than men," they will judge the competence of women against (low) standards for women, and the competence of men against (high) standards for men. The result is that it may not be appropriate to directly compare the subjective evaluations given to a male and a female target: Because each is judged relative to his or her sex, the meaning of an evaluation is category specific (e.g., "she's good for a woman" may mean something quite different than "he's good for a man"). However, before elaborating on the implications of this model for target evaluations of this sort, we highlight one of the approach's more basic assumptions: Stereotypes activate judgment standards. Thus, in the example above, the stereotype that "men are more competent than women" will lead perceivers to call to mind their expected distributions of men and women on the dimension of *compe-*

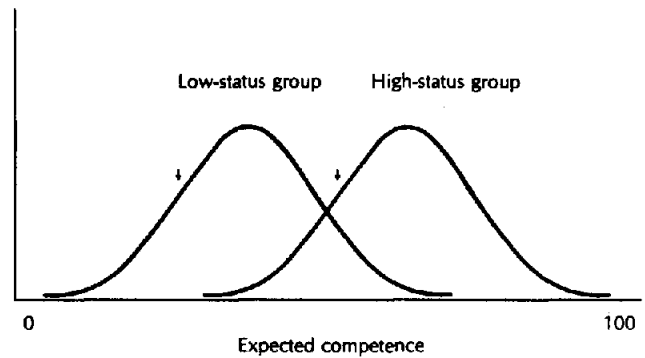


Figure 1. Schematic depiction of stereotyped representation of competence and minimum-standard levels for low- and high-status groups.

tence (see Figure 1). When a judge is asked to determine whether a man versus a woman performs reasonably well, the shifting standards model suggests that the judge may decide that good performance is anything that falls at, say, the 30th percentile of a sex-specific distribution. As can be seen in Figure 1, this decision rule results in the activation or setting of lower standards for diagnosing competence in women than in men (and, if the dimension and sexes were reversed, lower standards for diagnosing lack of competence in men than in women).

Results consistent with this prediction were reported by Biernat and Manis (1994) in the domain of *inferences about aggression and passivity*. In keeping with relevant gender stereotypes, lower standards were used to diagnose aggressiveness, but higher standards were used to diagnose passivity in a female than a male target. Assuming that a similar pattern holds in the domain of *competence*, stereotype-based standards may make it easier for a woman than a man to surpass the (low) standard for her group and be seen as impressively competent (e.g., "for a woman, she's very good"); at the same time, it may be easier for a man than a woman to surpass the (low) incompetence standard for his group and be seen as particularly incompetent (e.g., "for a man, he's very incompetent"). Thus, by virtue of shifting standards, women may find it easier than men to be recognized as competent, and men may find it easier than women to be recognized as incompetent. Substituting gender for race, this model echoes Carter's (1993) "flower blooming in winter" (stereotype-contrastive) metaphor.

Status characteristics theory (a part of the broader expectation-states theory) focuses on double standards with regard to status-based (in practice, most often gender-based) judgments of competence. In this model, standards are defined as "rules providing performance requirements for the inference of either ability or lack of ability" (Foddy & Smithson, 1989, p. 74). Although many aspects of this model are quite consistent with the shifting standards model, status characteristics theory makes one specific prediction that is at odds with the shifting standards approach. Specifically, status characteristics theory predicts that standards of ability are higher (i.e., require more evidence of ability) for low-status than high-status group members and, conversely, that standards for lack of ability are higher (require more evidence of lack of ability) for high-status than low-status individuals (Foschi, 1992). That is, a (high-status) man has to

do less than a (low-status) woman to prove his ability, and he is allowed more latitude (more demonstrations of low ability) than a woman before lack of ability is inferred (for evidence supportive of the use of gender-based double standards, see Foddy & Smithson, 1989; Foschi & Foddy, 1988; Pugh & Wahrman, 1983; Ridgeway, 1982; Wagner, Ford, & Ford, 1986). Substituting gender for race, this model echoes the "twice as hard to be considered half as good" sentiment expressed by Carter (1993) and suggests that stereotypes generally lead to assimilation: Individuals are judged in accordance with group expectations.

To summarize, although both models describe how status- or stereotype-based expectations influence judgment standards, the shifting standards model suggests that, for example, standards of competence will be lower for women than for men, whereas the status characteristics perspective suggests the converse: that standards of competence will be higher for women than for men. A resolution of these conflicting predictions can be found by noting an important distinction between the two models. In the shifting standards framework, standards refer to the *minimal level* of an attribute that is expected from a group: In the case of competence, minimal criteria to qualify as competent are lower for women than for men. Status characteristics theory, on the other hand, is concerned with standards for making broad-based inferences (i.e., correspondent inferences; Jones & Davis, 1965) of ability.

In the status characteristics approach, some person external to the judgment situation initially labels a performance a *success*, and the key question is whether the success is "interpreted as either conclusive or inconclusive evidence of ability" (Foschi, 1992, p. 186). According to the theory, "unexpected performance elicits a stricter standard, because the judge requires stronger evidence that this performance was due to an ability" (Foddy & Smithson, 1989, p. 76; see also Jones, Davis, & Gergen, 1961). For a member of a high-status group (e.g., a man), successful performance on a relevant task is sufficient, though perhaps not necessary, to infer high ability, but for a member of a low-status group (e.g., a woman), successful performance is necessary, but may not be sufficient, to infer high ability. Framing this in natural language terms, males may need to do well to infer high ability, but females must do very well to infer that same level of high ability (Foddy & Smithson, 1989; Ridgeway, 1982). This phenomenon is schematically presented in Figure 2. The status characteristic of gender (or race) affects the psychological mapping of performance onto ability, so that a higher performance score is required of a woman to infer the same ability level as is granted to a man.

In short, whereas the standards relevant to the shifting standards model concern the level of evidence that is necessary for members of one versus another group to meet minimum criteria on an attribute, standards in the status characteristics perspective refer to a higher level of proof—a level that allows for the broad-based inference that the attribute is ability based. This suggests that the shifting standards model may be more applicable than status characteristics theory to judgments concerning minimum standards or criteria to qualify as having an attribute (e.g., displayed competence in a domain), whereas status characteristics theory may be more applicable than the shifting standards model to broad-based inferences that this attribute is indicative of underlying ability. Thus, when perceivers are asked to make judgments that focus on *minimum standards*, we should

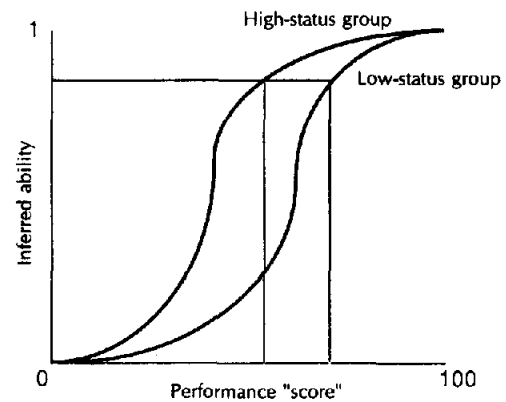


Figure 2. Schematic representation of psychological mapping of performance onto ability for low-status and high-status group members. From "Fuzzy Sets and Double Standards: Modeling the Process of Ability Inference," M. Foddy and M. Smithson, in *Sociological Theories in Progress: New Formulations* (p. 84), edited by J. Berger, M. Zelditch, Jr., and B. Anderson, 1989, London: Sage. Copyright 1989 by J. Berger. Adapted with permission.

find evidence that they set lower standards for women than men, and when they are asked to make broad-based ability inferences, they should more readily label the same level of performance by a man and a woman as indicative of ability in the man (see also Deaux, 1976, 1984; Deaux & Emswiler, 1974; Hansen & O'Leary, 1985; Swim & Sanna, 1996; Yarkin et al., 1982). The two studies presented here involve explicit tests of these predictions (with the substitution of Black-White in Study 2 for female-male in Study 1).

We should reiterate that the use of gender- and race-based competence stereotypes underlies assessments of both minimum standards and ability standards. Both types of judgments invite a direct manifestation of these stereotypes: The minimum-standards judgment elicits disclosure of the low expectations that are held for low-status group members, and the ability-inference judgment allows perceivers to indicate that because of these low expectations, a more stringent view is taken of competence evidence. This latter pattern is labeled *assimilative* in that evidence is viewed consistently with stereotyped expectations. Minimum-standards judgments, however, set the stage for a *contrastive* pattern of target evaluations (given low standards, a job candidate can be perceived as quite impressive, relative to other members of the low-status group). Other models of assimilation and contrast effects make no attempt to account for the stereotype-based decision rules that are used in the setting of differential performance standards, as outlined here.

Objective and Subjective Response Scales

In addition to the effects of judgment type (i.e., are minimal vs. broad-based inferences being made?) on the tendency for judgments to be assimilated to versus contrasted from stereotypes, the shifting standards model suggests that another variable—*response scale format*—is critical in predicting these patterns as well. Specifically, the shifting standards model has focused on the distinction between judgments that are made on subjective versus objective rating scales (Biernat, 1995).

Subjective scales refer to rating formats that allow the judge to impose his or her own meaning on the scale points. That is, the meaning of a unit of measurement can be adjusted to accommodate either the object being judged or the perceiver's idiosyncratic specifications. Scales that use Likert-type or semantic-differential formats most readily fit this description. For example, endorsing a 5 on a Likert-type scale and checking a response option labeled *very competent* are subjective judgments: The meaning of a 5 or of a *very competent* can shift, depending on the nature of the object being so described.

On the other hand, an *objective* response scale is one in which the measurement units have the same meaning regardless of the category to which the rated object belongs. Furthermore, objective judgment scales have some link to external reality; they are *externally anchored* or *common rule* scales. For example, one could evaluate a target's competence by judging how much money (in dollars) he or she should earn. Because a dollar is a dollar, regardless of who is earning it, we describe such a measurement unit as objective in nature. In previous work, we have included, under the rubric of *objective* scales, such units as feet and inches (for judging height), pounds (for judging weight), dollars, letter grades, and SAT scores (for assessing competence), explicit behaviors, and rank orderings of targets on dimensions of interest (see Biernat, 1995, for a more complete discussion of such scales). Again, what these scales have in common is the constant meaning of their measurement units, regardless of social category membership of the targets being judged. These scales differ from subjective scales in that they do not allow for category-based shifts in the meaning of response options.

What is the relevance of this distinction between subjective and objective judgment? The shifting standards model has suggested that social stereotypes lead individuals to differentially adjust the meaning of subjective response scale units to fit the expected range of variability in the target group being judged. For example, when judging the heights of male and female targets on a 1 (*very short*) to 7 (*very tall*) response scale, perceivers may use a 7 to describe any woman who appears to be over 5 ft. 9 in., but may reserve that same 7 only for men who appear to be over 6 ft. 2 in. The result is that subjective response scales of this sort may mask stereotyping effects: Men and women may appear to be judged equal in height, not because participants ignore the gender and height stereotype, but rather because this stereotype has led them to set different standards and to define rating scale units in a within-category (within-gender) fashion. Thus, when Locksley and her colleagues (Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982) found that male and female targets who engaged in an aggressive behavior were judged to be equally aggressive, they neglected to consider that this equivalence might be more apparent than real. A rating of 5 for a woman may mean something quite different than that same rating applied to a man. This is so because subjective judgments (e.g., "very aggressive" or "very competent") are implicitly linked with qualifiers such as "for a woman" or "for a man."

This problem of rating scale shifts diminishes if judges make their evaluations on objective response scales. We have argued that objective ratings are more likely than subjective ratings to reflect perceivers' "true" mental representations of targets.

Thus, when judging male and female heights in feet and inches, we more clearly see evidence of the gender-height stereotype (men are judged taller than women); subjective height judgments tend to mask this effect (see Biernat et al., 1991). In short, because objective scales prevent category-based meaning shifts, we anticipate that the predictions outlined above regarding the shifting standards and status characteristics models will be supported when judgments are made in objective rather than subjective units.¹ Thus, when we ask study participants to set minimum standards to determine whether a male or female job applicant has demonstrated competence, we anticipate that those who make this judgment by indicating an explicit number of examples of skills they require the candidate to supply (an objective response) will be more likely than those who indicate whether they'd like to see "a few" versus "many" skills (a subjective response) to show evidence of setting lower standards for women than men. Similarly, when we ask respondents to set criteria for making ability inferences, those responding in objective rather than subjective units should show more striking evidence of setting higher ability standards for women than for men. In general, and in keeping with our previous research, subjective responses should produce evidence of reductions or reversals of such effects (Biernat & Manis, 1994; Biernat et al., 1991).

Overall Evaluations

To this point, we have described predictions regarding the setting of judgment standards to infer minimal standing or broad-based ability in the domain of competence. This emphasis is important, because virtually all of the research to date on the shifting standards model has not focused directly on the setting of standards but rather on stereotype-based judgments or evaluations of targets. For example, in the research on gender and height described above, participants were not asked to indicate their standards for deciding what *tall* means for men and women, but rather estimated the heights of a series of individual male and female targets (Biernat et al., 1991). Similarly, in our research on gender stereotypes regarding verbal ability and racial stereotypes regarding verbal ability and athleticism, participants made global judgments of male and female and Black and White targets, respectively, on these attributes, but did not specifically indicate their group-based standards or expectations (Biernat & Manis, 1994). Whereas the setting of different standards was inferred in these prior studies, the main goal of the present research is to clearly and explicitly demonstrate the setting of differential performance standards to make inferences about minimal standing or broad-based ability on the dimension of *competence*. In addition, in an attempt to replicate previous findings in a different judgment domain, the studies will also include questions that assess perceivers' overall evaluations of male and female (Study 1) and Black and White (Study 2) job applicants.

Our earlier research on target evaluations has consistently

¹ Status characteristics researchers have never, to our knowledge, explicitly addressed the distinction between subjective and objective judgments. We doubt, however, that they would have any difficulties with or objections to our comments on this point.

demonstrated that objective response scales reveal evidence of stereotype operation (assimilation), whereas subjective scales produce reductions or reversals (contrast) of these effects. The precise pattern that appears on subjective response scales (reduction or reversal of stereotyping effects) can be predicted by assessing the fit of targets to stereotyped expectations (see Biernat & Kobrynowicz, 1997; Biernat, Vescio, & Manis, in press). Specifically, subjective judgments should reveal reductions of stereotyping effects when targets display attributes that are stereotype consistent or ambiguous, whereas reversals of stereotyping effects (contrast) should occur when targets display stereotype-inconsistent attributes. This latter prediction is consistent with expectancy violation theory (Jussim et al., 1987): A stereotype-inconsistent target may produce a "wow" effect (e.g., "she's really competent compared to most women!"), which leads to an extremized response on subjective scales. Objective scales, because their meaning does not shift, should nonetheless consistently reveal evidence of assimilation to stereotypes.

Most relevant to the present studies are data from Biernat and Manis (1994, Study 1), in which participants were asked to evaluate a male or female author of a masculine (e.g., bass fishing) or feminine (e.g., meal planning) magazine article. We assumed that a woman who wrote about feminine topics and a man who wrote about masculine topics were stereotype consistent in their behaviors, but that individuals who wrote on topics presumably relevant to the opposite sex were stereotype inconsistent. Following our usual paradigm, half of the participants judged the quality of the magazine article on objective response scales (e.g., monetary worth in dollars) and half on subjective scales (e.g., *very little to lots* of money). Results indicated that stereotype-inconsistent targets (i.e., female authors of masculine articles and male authors of feminine articles) were judged more negatively than stereotype-consistent targets, among the judges who used objective response scales. However, these same stereotype-inconsistent targets were judged more positively than stereotype-consistent targets when the judges used subjective response scales. That is, on subjective judgment scales, stereotype-inconsistent targets were contrasted from the relevant stereotype (e.g., men were thought to be subjectively better at writing feminine articles than women), but on objective scales, the authors were assimilated to stereotyped expectations (e.g., male authors were thought to have produced feminine articles that were worth less money than those of the female authors).

We anticipate a similar pattern to hold in the present research, in which we examine judgments of male and female applicants for "masculine" versus "feminine" jobs (Study 1). Not only will this provide a conceptual replication of the magazine article study described above but also it will provide an additional link between status characteristics theory and the shifting standards model. Both models emphasize the importance of stereotypes or expectations for the perception of individual targets, and Foschi (1992) explicitly argued that sex of the actor will interact with "gender of the task" at hand, so that, for example, "a man's performance at a masculine task will be assessed as better than that same performance by a woman" (p. 185). Furthermore, the status characteristics perspective suggests that ability attributions will be stronger when a target performs in a stereotype-consistent (i.e., "gender appropriate") domain. Thus, sex

typing of the job may have an impact on overall evaluations of male and female targets, as Biernat and Manis's (1994) research suggests, but may additionally affect the setting of standards to infer minimal competence versus broad-based ability: Lower expectations should produce low standards to infer minimal competence and high standards to attribute that performance to broad-based ability.

Overview of the Studies

In both studies presented here, participants were asked to evaluate a low-status or high-status person's suitability for a job. In Study 1, the low-status person was a White woman and the high-status person was a White man. In Study 2, the low-status person was a Black man, and the high-status person was a White man. Given our discussion that gender stereotypes regarding competence may be dependent on the sex typing of tasks, Study 1 also included a manipulation of job type: It was either feminine (i.e., executive secretary) or masculine (i.e., chief of staff) in nature. The other between-participants manipulations in both studies consisted of the two variables that we believed were key in determining whether assimilative or contrast tendencies would emerge: the type of judgment participants were asked to make, either a minimum-standard judgment or an ability inference, and the response scale on which these judgments were made, subjective or objective in nature (see descriptions below).

We were concerned with two general types of dependent measures. Of central importance were the *standards* questions; these concerned the performance criteria that judges set to determine whether an applicant should be considered for a position. The second type were general *evaluation* questions; these allowed us to meet our subsidiary goal of replicating earlier research findings (Biernat & Manis, 1994). On the basis of the shifting standards and status characteristics models, we made the following predictions:

1. On standards questions, respondents making minimum-standard judgments will show evidence of setting lower standards for the low-status than the high-status target person when these judgments are made in objective units. Subjective responses will mask this pattern.²
2. On standards questions, respondents making ability inferences will show evidence of setting higher standards for the low-status than the high-status target person when these judgments are made in objective units. Again, subjective judgments will mask this pattern.
3. Because sex typing of the job might affect the extent to which a strong performance is expected, the patterns predicted in Hypotheses 1 and 2 will be particularly likely to hold when the target is being evaluated for a "masculine" job (Study 1 only). When the job is "feminine," an opposite pattern of ef-

² We made no specific predictions regarding the precise form of this masking on the standards questions, as it was unclear, for example, how a high objective standard was likely to translate into subjective language. However, as indicated in the Overall Evaluations discussion above and in Prediction 5, we did make specific predictions about the expected pattern of effects on subjective *evaluation* questions.

fects will emerge, because it is the woman, not the man, who is positively stereotyped or imbued with high status.

4. On evaluations of the target applicant, objective scales will be more likely than subjective scales to reveal stereotype operation: the high-status target will be judged more positively than the low-status target.

5. Prediction 4 will be moderated by the sex typing of job (Study 1 only): Objective scales will reveal that the woman is considered to be better than the man in the feminine job, but the man is considered to be better than the woman in the masculine job. Subjective scales will reveal reversals (stereotype contrast) of these effects, because the male applicant for the feminine job and the female applicant for the masculine job are engaging in stereotype-inconsistent endeavors.

Study 1

Method

Participants and Procedure

Participants were 313 undergraduates at the University of Kansas (196 women, 115 men, 2 unknown). Because no reliable sex-of-participant effects emerged in these data, this variable will not be discussed further. Participants were tested individually or in groups of 2 to 6 but completed all materials independently. Two experimenters were used, one male and one female; as will later be noted, experimenter sex resulted in a main effect on one of the critical dependent variables and, for that reason, was retained as a factor in all analyses.

After being told that the study involved employment interviewing, each participant received a folder that contained a job description, an applicant's resume, and an evaluation form. Participants were asked to first read the front of the folder, which contained the instruction set manipulation (described below), and then to read the description of the position and the applicant's resume. Once they had reviewed this information, they were asked to complete the applicant evaluation form, referring back to the job description and the applicant's resume if they chose.

All participants received the identical job description; however, the job title varied, so that the applicant applied to be either an executive secretary (feminine position) or an executive chief of staff (masculine position). Pretesting indicated that the job description paired with the latter title was rated as more masculine, less feminine, having more status, having a significantly higher starting salary, and as having significantly fewer women employed in the position than the identical job description paired with the executive secretary title (all $ps < .05$). In addition, all participants reviewed the identical resume, with applicant gender varied. Half received a resume with a man's name (Kenneth Anderson), the other half with a woman's name (Katherine Anderson). The applicant was always college educated and had qualifications (e.g., job experience) that were designed to be of moderate caliber.

To test the predictions from the shifting standards model, the response scale on which participants made their judgments was varied. Half of the participants responded to the items with an objective response scale (e.g., number of examples of skills). The other half responded with Likert-type rating scales (e.g., *few* to *many* examples of skills).

Finally, the instructions accompanying the applicant evaluation items varied. Half of the participants were instructed to respond to the evaluation form by determining how they would decide if the applicant had the ability required to be successful at the position. The other half were instructed to determine if the applicant met the minimum standard to be successful at the position. We believed that use of these different instructions might provide an avenue for integrating the shifting stan-

dards and status characteristics models. To summarize, the study was based on a 2 (sex of applicant) \times 2 (sex typing of job) \times 2 (instructional set: minimum standard vs. ability inference) \times 2 (response scale: subjective vs. objective) factorial design.

Dependent Measures

Two types of dependent variables were assessed: One type was designed to measure the standards or criteria that were set for applicant performance, and the other was designed to assess evaluations of the candidate in his or her position. For all of the standards items, instructions (minimum standard vs. ability) were repeated before each question, but for the evaluation items, they were not. As indicated above, half the sample responded on objective and half on subjective response scales. For each item, the question stem was identical across scales, but response options varied, as described below. Each of the indexes was standardized within scale type, so as to allow cross-scale comparisons. For each index, judgments greater than 4 standard deviations from the mean were deleted from analysis. This resulted in the deletion of a total of 15 data points (across all individual items included in the indexes) from 5 participants; each of these 5 individuals was in a different experimental condition.

Standards—criteria assessment. Critical to testing the shifting standards and status characteristics predictions was an assessment of the standards or thresholds that perceivers set to determine the competence of the target applicant. For example, at what point in an interview or information-gathering procedure would they feel comfortable that the applicant possessed the required skills? How much evidence would be satisfactory? How many "hoops" would the applicant be required to jump through before meeting criterion? We assessed these standards in two ways; in each case, half the participants answered on subjective and half on objective response scales.

First, to assess levels of performance required before the applicant should be recommended for a second interview, we asked participants two questions: (a) What general level of performance was required of the applicant (objective response: applicant's rank relative to other applicants; subjective response: 1 [*low*] to 7 [*high*] rating), and (b) what level of performance on a "standardized ability test" did the applicant need to meet (objective response: applicant's percentile score; subjective response: 1 [*low*] to 11 [*high*] rating). Coefficient alpha on this two-item index, which is referred to as the *levels* measure, was .67.

Second, participants were provided with a list of nine skills described as "important to success in this position." The skills included decision making, interpersonal relations, leadership, monitoring, motivation, oral communication, problem solving, planning, and willingness to seek and accept assistance. Participants were asked to indicate "how many examples of each skill you would require of this applicant before feeling confident that he or she (has the ability/meets the minimum standard) to perform the skill." In the objective condition, participants indicated a number for each skill; in the subjective condition, they rated each skill on a 1 (*very few*) to 7 (*many*) response scale. Coefficient alpha on this 9-item "skill examples" index was .94.

Evaluations. In addition to the standards questions, we were interested in perceivers' global evaluations of the applicant. Participants were asked to predict the likely job evaluations that the applicant would receive if he or she were eventually hired. Six questions were included in this index ($\alpha = .82$): (a) What is your overall impression for how well this individual might perform in the position for which he or she is applying (objective: rank order of target relative to others in the same position; subjective: 1 [*very poorly*] to 7 [*very well*]); (b) if you were to review nine applicants in addition to this one, how would you guess this applicant might compare (objective: rank order; subjective: 1 [*very poorly*] to 7 [*very well*]); (c) if this applicant were hired, how competent would you expect him or her to be when performing the responsibili-

ties for this position (objective: percentage of work responsibilities at which the applicant would be competent; subjective: 1 [*not at all competent*] to 7 [*very competent*]); (d) how likely is it that this applicant is overqualified for this position (objective: percentage indicating likelihood of overqualification; subjective: 1 [*not at all likely*] to 10 [*very likely*]); (e) what overall letter grade would you assign to this applicant (objective: actual letter grade from A+ to F; subjective: 1 [*low*] to 7 [*high*]); and (f) how likely is it that this individual will eventually be promoted to vice president (objective: percentage indicating likelihood of promotion; subjective: 1 [*not at all likely*] to 5 [*very likely*]).

Manipulation Checks

To ensure that participants noted the applicant's sex and job title, they were asked to write the candidate's name and the position applied for at the top of the evaluation sheet. All participants correctly answered these questions. After completing all dependent variables, respondents also indicated the status, masculinity, and femininity of the position that the applicant applied for on a 1 to 7 scale, as well as the presumed starting salary of the position. Compared with the executive secretary position, the executive chief of staff position was rated as higher in status, $F(1, 295) = 29.96, p < .0001, M_{\text{chief}} = 5.70, M_{\text{sec}} = 4.96$; higher in starting salary, $F(1, 285) = 61.90, p < .0001, M_{\text{chief}} = \$44,111, M_{\text{sec}} = \$29,108$; and higher in masculinity, $F(1, 295) = 5.16, p < .02, M_{\text{chief}} = 3.78, M_{\text{sec}} = 3.37$; but lower in femininity, $F(1, 295) = 74.17, p < .0001, M_{\text{chief}} = 3.34, M_{\text{sec}} = 4.78$.³

Results

Overview

Predictions 1–3 concerned expected judgment patterns on the standards, or performance criteria, questions, and Predictions 4–5 focused on applicant evaluations. For each hypothesis, our key concern was with the predicted pattern of means on the objective judgment scales. For Hypotheses 1–3, we suggested only that the subjective scales would mask these patterns (see Footnote 2). Technically, then, it was appropriate to conduct separate analyses of the objective and subjective scale conditions. However, we took the more conservative path of first conducting omnibus analyses that included all of the data and later partitioning the sample into the objective and subjective conditions. All analyses took the form of an Applicant Gender \times Job Type \times Instructions (minimum standards vs. ability standards) \times Response Scale (objective vs. subjective) \times Sex of Experimenter analysis of variance (ANOVA).

Standards for Applicant Performance

We asked two sets of questions designed to assess whether performance standards and expectations varied on the basis of applicant gender, job type, instruction, and response scale on which judgments were made. On the index tapping performance levels the participant required from an applicant before recommending him or her for a second interview, we found only a main effect of instructional set, $F(1, 269) = 6.19, p < .02$. Instructions to assess ability led participants to require a higher level of performance ($M = .12$) than did instructions to assess minimum standards ($M = -.13$). These data provided us with general evidence that respondents saw the inference of ability as a higher standard than the inference of minimum competence.

To further assess judgment standards, participants were asked

to indicate the number of examples of a series of nine important skills they would require of the applicant "before feeling confident that he or she (has the ability)/(meets the minimum standard) to perform the skill." The five-way ANOVA on this index revealed two significant effects. First, an unexpected (and theoretically irrelevant) main effect of experimenter gender, $F(1, 269) = 4.33, p < .04$, indicated that participants exposed to the male experimenter ($M = .15$) set higher standards overall than did participants exposed to the female experimenter ($M = -.08$). Of more theoretical interest was the three-way interaction among applicant gender, instructions, and response scale, $F(1, 269) = 6.18, p < .02$, which is relevant to Hypotheses 1 and 2. This interaction is displayed in Figure 3.

The top panel depicts judgments in objective units, and the lower panel depicts subjective judgments. Because predictions differed for the objective and subjective response scale conditions, we conducted separate analyses of them. The objective judgment data provide striking evidence supportive of both the shifting standards and status characteristics perspectives. The Applicant Gender \times Instructions interaction was significant, $F(1, 136) = 7.01, p < .01$. Planned contrasts (using the overall error term) indicated that when the skill examples question was framed in terms of meeting minimum standards—the judgment of most interest from the shifting standards perspective—participants indicated a lower standard for the female than for the male applicant, contrast $t(74) = 1.81, p < .08$.⁴ In keeping with Hypothesis 1, these data indicated that participants had lower expectations for female than male applicants, which may set the stage for the "flower blooming in winter" phenomenon described by Carter (1993). When the skills question was framed in terms of making an ability inference—the critical judgment in status characteristics research—participants indicated that they required significantly more evidence of skill from a female than a male applicant, $t(74) = 2.26, p < .05$. This is the "twice as hard to be considered half as good" effect, which supported Hypothesis 2. This interaction was also driven by the fact that for female applicants, ability standards were significantly higher than minimum standards, $t(74) = 2.84, p < .01$; these standards were not reliably different for male applicants, $t(74) = 1.14, p > .25$.

Also consistent with the shifting standards model was the finding that subjective judgments revealed a very different pattern of effects (bottom panel of Figure 3). The Applicant Gender \times Instructions interaction was not significant for subjective judgments ($F < 1$), and none of the means reliably differed from each other.

Finally, the standards data provided no evidence supportive of Hypothesis 3, which concerned the influence of job type (masculine or feminine) on the setting of standards. The relevant Applicant Gender \times Job Type \times Instruction \times Response Scale

³ Given these and our earlier reported pretest findings regarding perceptions of the job description, the job type manipulation could be considered either a sex-typing or status manipulation. Locating the precise determinant of any effects of this factor is not of particular interest to us here, because status and masculinity of occupations are certainly confounded in the real world (Basow, 1992).

⁴ Although we had directional predictions for each male–female contrast examined here, all reported tests were two-tailed.

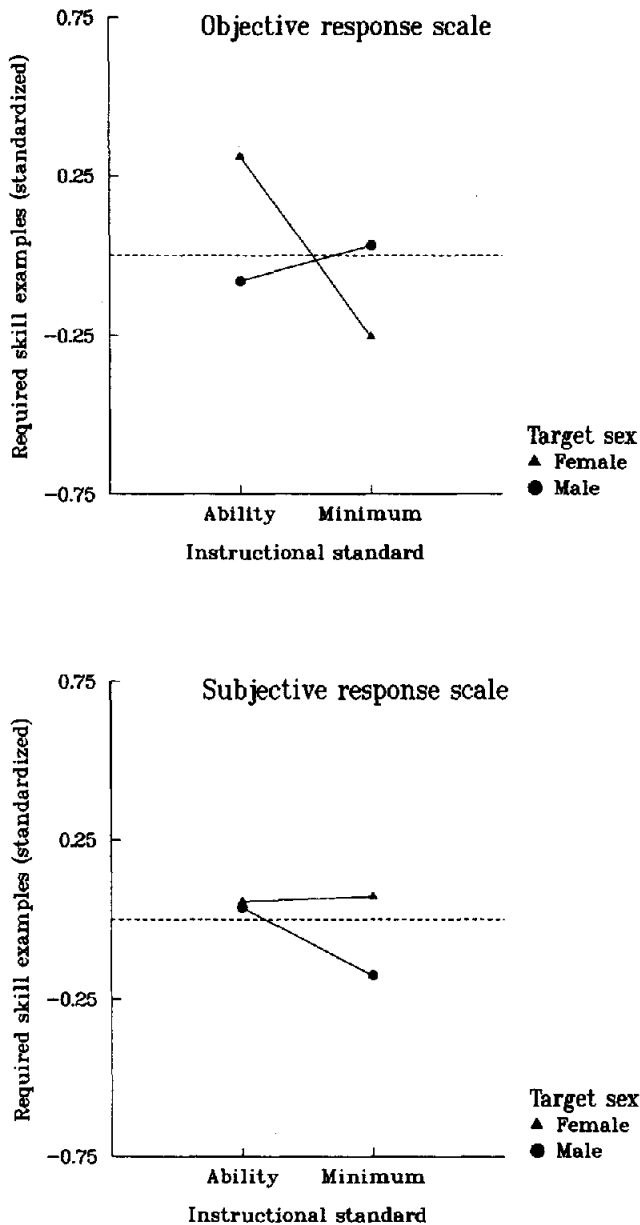


Figure 3. Number of skill examples required to document ability or minimum standards, by applicant sex and response scale, Study 1.

interaction was not significant, $F(1, 269) = 2.79, p < .10$, nor was the Gender \times Job Type \times Instructions interaction significant when the objective judgment conditions alone were considered, $F(1, 136) = 1.40, p > .20$. Performance standards did not differ depending on job type.

Evaluation of the Applicant

Predictions 4 and 5 suggest that objective judgment scales will reveal stereotype operation: Men should be evaluated more positively than women when the job is "masculine," and women should be judged more positively than men when the job is

"feminine," whereas subjective scales should reveal reversals of these effects. Operationally, support for these predictions would take the form of a significant Applicant Gender \times Job Type \times Response Scale interaction and a significant Gender \times Job Type interaction when the objective and subjective conditions were considered separately. The overall ANOVA revealed a main effect of job type, $F(1, 270) = 10.62, p < .002$, such that applicants for the feminine (secretary) job were judged more positively ($M = 0.11$) than applicants for the masculine (chief of staff) job ($M = -.12$), in addition to the predicted Applicant Gender \times Job Type \times Response Scale interaction, $F(1, 280) = 11.91, p < .001$.

This interaction is depicted in Figure 4; the objective judgment conditions appear in the top panel, and the subjective judgment conditions appear in the lower panel. Looking first at the objective conditions, considerable support for Predictions 4 and 5 can be noted. The Gender \times Job Type interaction was significant, $F(1, 139) = 7.08, p < .01$, and planned contrasts indicated that for the feminine (secretary) job, the female applicant was judged significantly more positively than the male applicant, contrast $t(77) = 1.99, p < .05$, whereas the converse was true for the masculine (chief of staff) job—the male applicant was judged more positively than the female applicant. This latter effect was not, however, reliable, $t(74) = 1.35, p < .18$.

Also consistent with shifting standards predictions was the finding that when judgments were made in subjective units (lower panel of Figure 4), reversals of these patterns appeared. Subjectively, the female chief of staff applicant was evaluated more positively than the male chief of staff applicant, and the male secretarial applicant was evaluated more positively than the female secretarial applicant. This Gender \times Job Type interaction was significant, $F(1, 141) = 5.25, p < .05$, and the individual contrasts were marginally significant, $t_s(77) = 1.72$ and 1.84 , respectively, $p_s < .09$. To summarize the evaluation findings, objective judgments revealed evaluations consistent with stereotypes regarding fit between gender and job: Women were evaluated more positively for the feminine job; men were evaluated more positively for the masculine job. However, subjective judgments indicated the opposite, stereotype-contrastive patterns—as though the male applicant for the feminine job was perceived as very good, for a man, and the female applicant for the masculine job was perceived as very good, for a woman (see also Biernat & Manis, 1994, Study 1).

Discussion

These data provide strong support for the operation of both assimilative and contrastive outcomes in gender-based, employment-relevant judgment. Which of these phenomena occurred depended on the form and type of the judgment question, as predicted by the guiding theoretical perspectives of this work: status characteristics theory and the shifting standards model. We more specifically evaluate the quality of the data by revisiting each of the five hypotheses outlined in the introduction.

The first three hypotheses concerned predicted patterns of judgment on questions tapping the standards or criteria that applicants were required to meet on their road to a second interview. Specifically, we expected that minimum standards for passing an applicant on to a second interview would be lower

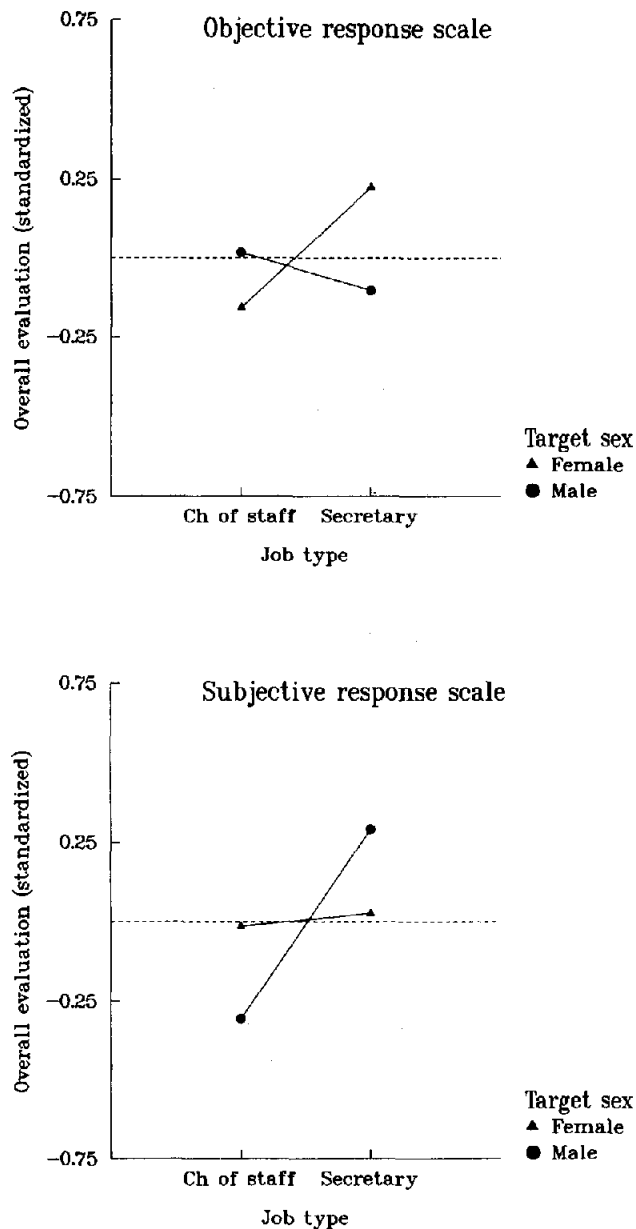


Figure 4. Overall applicant evaluations by applicant sex, job type, and response scale, Study 1. Ch = chief.

for female than for male applicants. Essentially, such a pattern would document that perceivers expect less from a woman than a man, which could lead to the "flower blooming in winter" experience described by Carter (1993). Judgments in objective units clearly supported this pattern; also consistent with Hypothesis 1 was the finding that subjective judgments masked this effect. In Hypothesis 2, we predicted that when making ability inferences, judges would set higher requirements for a female than a male applicant. This is the judgment pattern consistent with a status characteristics perspective: A (low-status) female applicant must do more to prove her ability than does a (high-status) male. As documented in Figure 3, this is precisely the

pattern we found among participants who made their judgments in objective units: More examples of skills were required for a woman than a man, to document ability. Also in keeping with Hypothesis 2 and the shifting standards model, subjective judgments produced a reduction of this effect. Because subjective response scales are susceptible to within-category meaning shifts, the subjective label *many skill examples* was apparently translated into objectively more examples when applicants were female rather than male.

In Hypothesis 3, we suggested that the patterns of effect discussed in the previous paragraph would be most readily apparent when participants were considering applicants for the masculine (chief of staff) rather than the feminine (executive secretary) position and that opposite patterns would appear when the secretarial position was considered. This was based on both status characteristics and shifting standards predictions that an applicant for a "sex-inappropriate" job both would be held to a lower minimum standard and would be required to do more to prove ability than an applicant for a "sex-appropriate" job. This hypothesis was not supported in analyses of either the levels or skills indexes. We did, however, find that several of the individual items that made up these indexes conformed to the predicted pattern. Although this leaves us with some faith in the hypothesis, further work is clearly needed before any strong conclusions can be reached.

Hypotheses 4 and 5 centered on predictions regarding applicant evaluations: How well or poorly would the applicant perform in the position? Once again, as predicted by the shifting standards model, objective judgments revealed stereotype-consistent judgment patterns: When applicants were evaluated for the feminine (executive secretary) position, Katherine was viewed more favorably than Kenneth, but when applicants were evaluated for the masculine (chief of staff) position, Kenneth was rated more favorably than Katherine. Respondents assigned higher evaluations to individuals in sex-appropriate jobs: Katherine would make a better executive secretary; Kenneth would make a better chief of staff. For those making subjective judgments, however, these patterns were, as predicted, reversed. The shifting standards model suggests that stereotype-inconsistent targets (e.g., a male secretary or a female chief of staff) are contrasted from the expectations for their genders, and subjective response language readily captures the "wow" effect that such stereotype inconsistency inspires ("he would be very good for a male secretary"). Because of standard shifts, we assume that what perceivers label as *very good* for a man in a secretarial position is objectively less than what they label as *very good* for a woman in the same position (see also Biernat & Manis, 1994).

Because research on status characteristics theory has focused primarily on gender as a status variable (Foddy & Smithson, 1989; Foschi, 1992; Foschi et al., 1994; Webster & Foschi, 1988), Study 1's focus on gender provided a good starting point for comparing that theory with the shifting standards perspective. However, as noted in the introduction, we were intrigued by the paradox of assimilation and contrast as exemplified in Carter's (1993) comments about *race*. For that reason, we designed Study 2 as a conceptual replication of Study 1 in which race of applicant was manipulated rather than gender. In Study 2, we forewent the inclusion of a job-type manipulation, as race

typing of jobs is a less straightforward matter than is sex typing, particularly among college-educated applicants, as used here. Thus, although Hypotheses 3 and 5 are not relevant in Study 2, we anticipate replication of all other patterns noted in Study 1.

Study 2

Method

Participants were 277 undergraduates at the University of Kansas (161 women, 115 men, 1 unspecified), who received course credit for their involvement in the study. Participants arrived to the lab individually or in groups of 2 to 4 but always worked independently at the task. The procedure was identical to that of the prior study except that no job-type manipulation was used (the applicant always applied for the executive chief of staff position), and the status feature we focused on was race, rather than gender. The applicant was always male, but he was identified as either Black or White via (a) ethnicity cues in his name (Marcus Washington versus Mark Washburn; see Bodenhausen, 1990), and (b) the leadership experience listed on his application (fund-raising chair for the Black Student Union vs. the Student Alumni Association). A manipulation check at the end of the evaluation questionnaire revealed that all participants correctly inferred the applicant's race, though 2 individuals in the White condition did not answer the question. Because we assume the "default" race was White, these 2 participants were left in the data set. All other procedures, manipulations, and dependent variables were identical to those used in Study 1. As in Study 1, we also deleted data points that deviated more than 4 standard deviations from an item or index mean, resulting in a total loss of 18 data points from 6 participants, who were distributed across different conditions.

Results

Standards for Applicant Performance

Analyses were based on a series of Applicant Race \times Instructions (minimum standards vs. ability) \times Response Scale (objective vs. subjective) ANOVAs. We used the same measures as described in the gender study to assess race-based performance standards and to test Hypotheses 1 and 2. Again, we expected that judgments in objective units would reveal that when ability inferences were made, higher standards would be set for Black than for White applicants, but that when minimum-standards inferences were made, lower standards would be set for Blacks than for Whites. In this study, the index measuring required level of performance to recommend a second interview revealed a main effect of instructional set, $F(1, 268) = 27.37, p < .0001$, so that a higher level of performance was required if ability ($M = .26$) rather than minimum standards ($M = -.26$) was being assessed. Furthermore, the critical Applicant Race \times Instructions \times Response Scale interaction was significant in this analysis, $F(1, 268) = 3.99, p < .05$.

Depicted in Figure 5, and in keeping with Hypotheses 1 and 2, this interaction indicated that when judges made objective estimates (top panel), they required a higher level of performance to infer ability in a Black than a White applicant but a lesser level of performance to meet minimum standards for a Black than a White applicant. The Race \times Instructions interaction on these objective data was significant, $F(1, 132) = 4.56, p < .04$, and planned contrasts indicated that whereas the Black-White difference for minimum-standards inferences was reliable, $t(65) = 2.07, p < .04$, the comparable effect for ability

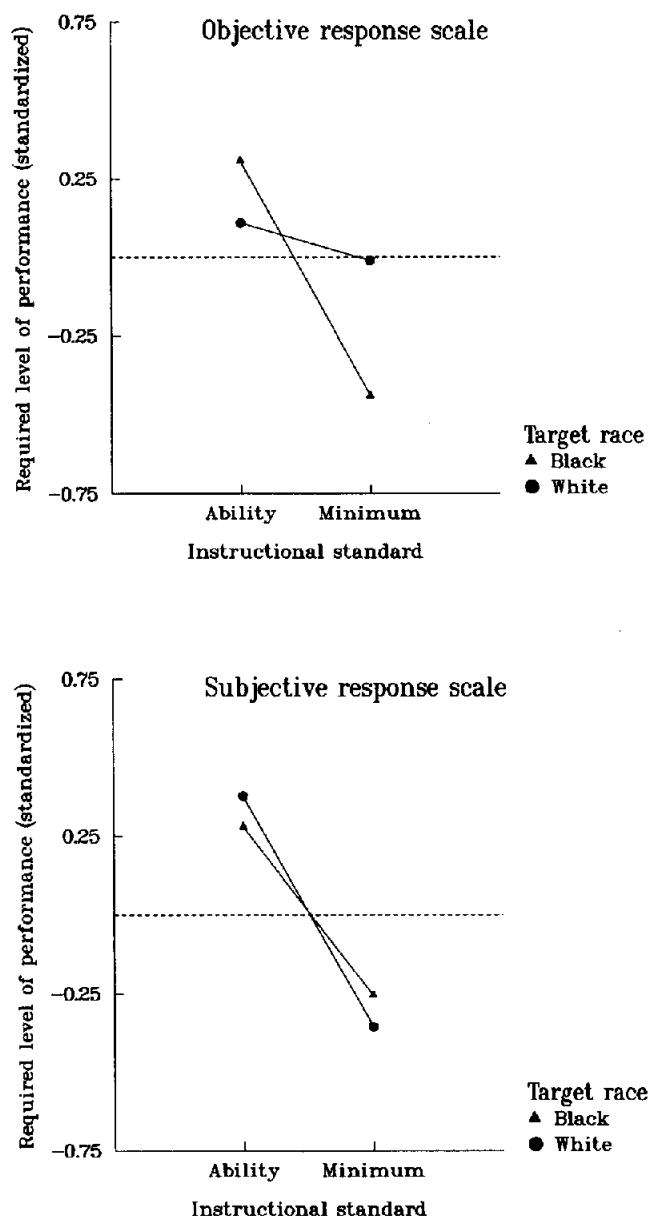


Figure 5. Required level of performance to document ability or minimum standards, by applicant race and response scale, Study 2.

inferences was not, $t(67) = 1.01, p > .25$. This interaction was also driven by the significantly higher level of performance required to make ability, as compared with minimum-standards inferences when the applicant was Black, $t(65) = 3.60, p < .001$; standards did not reliably differ when the applicant was White, $t(67) = 0.58, n.s.$

Also in keeping with predictions, subjective response scales (lower panel) revealed a rather different pattern of judgments. Analysis of these data alone revealed only a significant main effect of instructional set, $F(1, 136) = 18.24, p < .0001$; the Race \times Instructions interaction was not significant ($F < 1$), although the general pattern of means was the reverse of that

reported for the objective scales. For both Black and White applicants, ability standards were higher than minimum standards when subjective judgments were made, contrast $t(72) = 2.78$ for Blacks, $t(65) = 3.19$ for Whites, $ps < .01$.

The comparable analysis on the skills index revealed no significant effects (all $ps > .10$). Thus, whereas the skills index but not the levels index revealed a significant Gender \times Instructions \times Scale interaction in Study 1, it was the levels but not the skills index that revealed the comparable Race \times Instructions \times Scale interaction in Study 2. Note, however, that although the three-way interaction on skills was not significant in this study ($p > .25$), the means were in the predicted direction (e.g., objectively, the Black ability standard [$M = .11$] was higher than the White ability standard [$M = -.18$], whereas the Black minimum standard [$M = .08$] was lower than the White minimum standard [$M = .17$]; subjectively, these patterns were reduced or reversed [respective $Ms = .13, .03, -.01, -.16$]).

Evaluation of Applicant

Hypothesis 4 suggests that evaluations of the applicant in objective units will be more likely than those in subjective units to reveal evidence of stereotype operation. Overall, we found evidence for a reverse stereotyping effect: The Black applicant ($M = .09$) was evaluated more positively than the White applicant ($M = -.09$), $F(1, 269) = 3.87$, $p = .05$. However, the Race \times Scale interaction, $F(1, 269) = 3.50$, $p < .07$, indicated that this was only true when judgments were made on subjective response scales (Ms for Black and White were .17 and $-.18$, respectively), $t(138) = 2.72$, $p < .05$. Objective response scales revealed no Black-White difference ($Ms = .01, -.01$). Thus, although we expected that Black applicants would be judged more negatively than White applicants in objective units, we found no evidence for that here; rather, the race effect was reliable, and showed the predicted evidence of reverse stereotyping, only among participants who made subjective judgments.

Discussion

These data provide a strong conceptual replication of the Study 1 findings and do so by substituting race for gender as a status cue. In assessments of hiring criteria, participants indicated that they set lower minimum standards for Black than White applicants but required more evidence to document ability in Black than White applicants. These patterns are in keeping with the shifting standards and status characteristics models, respectively. Also, as the shifting standards model predicts, these findings were apparent only among individuals who made objective judgments but were slightly reversed among those who made subjective judgments (see Figure 5). Once again, the subjective judgments tell a substantially different story than do the objective judgments, which, by virtue of their connection to external anchors, provide a more accurate reflection of perceivers' impressions.

It was the case that of our two indexes measuring standard assessments (levels and skill examples), the above effect was significant only on the levels variable in the race study but on the skill examples variable in the gender study. We have no satisfactory explanation for why this was the case; nonetheless,

the consistent patterns of means across studies attest to the general reliability of the effect. Minimum standards are set lower for low- than for high-status applicants, but ability inferences are easier to obtain if one is of high rather than low status.

With regard to applicant evaluations, we continued to find that instructional set did not affect these judgments but that Black and White applicants were evaluated differently depending on the form that evaluations took (subjective or objective). Specifically, the Black and White applicants were judged equivalently in objective units, but the Black was judged more favorably than the White in subjective units. This pattern is somewhat different from that predicted in Hypothesis 4: Objective judgments revealed no racial difference rather than a stereotype-consistent judgment pattern. Technically, it is true that the objective condition revealed more anti-Black bias than did the subjective condition, but this bias was not as strong as we expected. Perhaps social desirability and self-regulation concerns are more operative in studies of race than gender, leading participants to suppress stereotypic responding (see also Biernat & Vescio, 1993; Gaertner & Dovidio, 1986; Monteith, 1993). What is critical from the shifting standards perspective, however, is that subjective responses continued to show a substantially different pattern than objective responses; one that we assume is based on standard shifts. Also, because Study 2 focused on the chief of staff position only, the most direct comparison with Study 1 lies in the gender-based evaluation data that appear in the bottom panel of Figure 4. As in the present study, it was also the case in Study 1 that the status (gender) effect was stronger (in a stereotype-contrastive direction) when subjective rather than objective judgments were made.

General Discussion

Two studies that involved simulation of an employment decision-making setting provided substantial support for both status characteristics theory (Berger et al., 1977, 1980, 1985; Foddy & Smithson, 1989; Foschi, 1992; Foschi et al., 1994; Webster & Foschi, 1988), and the shifting standards model (Biernat, 1995; Biernat & Manis, 1994; Biernat et al., 1991). Status characteristics theory, and particularly its focus on gender-based double standards of competence, led to the prediction that it would be more difficult for women than men, and by extension for Blacks than Whites, to document their ability in a competence-related domain. In both studies, we found that participants required women and Blacks, relative to men and Whites, respectively, to "jump through more hoops" to prove that they had the ability to fill a position. This is consistent with Figure 2's schematic depiction of double standards in the mapping of performance onto ability and echoes Carter's (1993) lament that Blacks "really do have to work twice as hard to be considered half as good" (p. 58) as Whites.

The shifting standards model, and more specifically its focus on stereotype-based judgment standards, makes predictions that appear, at first glance, to contradict the theory of status characteristics. Assuming that perceivers expect Blacks relative to Whites, and women relative to men, to be less competent in employment settings, they will set lower performance standards for these individuals. Both studies presented here documented that in fact, judges do set lower minimum performance standards for women

and Blacks than for men and Whites. We suggest that the process involved is a straightforward one: The instructional set to compute minimum standards led judges to access their stereotyped expectations, which resemble the distributions depicted in Figure 1. The pattern of data therefore provides direct evidence that judges hold gender and racial stereotypes, and hence double standards, regarding competence. That this pattern appeared in conjunction with the ability inference data described above suggests that the shifting standards model and status characteristics theory are not contradictory, but rather complementary, judgment models. Low-status individuals, held to lower minimum standards, may more easily meet the within-group criteria that identify them as reasonable or successful job candidates. But this success is implicitly interpreted relative to the (lower) category standard. This leaves low-status individuals with a long road to pave before their ability is documented (see also Pettigrew & Martin, 1987). Furthermore, these studies make an important methodological point: Minimum standards and ability standards are higher for devalued groups only when these standards are measured in objective response scale units.

Both studies also indicated that the difference between the minimum-standard and ability-standard instructional sets was most marked when perceivers considered devalued group members (women and Blacks) but was not reliable when "valued" group members were judged (White men). That is, for the female and Black applicants, minimum standards were significantly lower than ability standards, but these standards did not differ for White male applicants (see top panels of Figures 3 and 5). Although it seems reasonable to suppose that an ability standard is always a stricter criterion than a minimum standard, these data suggest that this norm does not hold when one has the benefit of membership in a valued group.

The setting of lower minimum standards for stereotypically low than high status groups is also likely to produce the "flower blooming in winter" effect. An unexpectedly strong showing on the part of a low-status person may indeed be evaluated more positively than the same (expected) showing by a high-status target (Jackson et al., 1993; Jussim et al., 1987). Although we did not directly assess it, we assume that because the "job applicants" in our studies were educated and had moderately strong credentials, the low-status target quite likely surpassed and the high-status target (barely) met minimum-standards criteria. The shifting standards model again was valuable in predicting the resulting patterns of employee evaluation. In keeping with the "flower blooming in winter" effect, subjective evaluations revealed that the unexpected targets—women and Black men applying for the chief of staff position and White men applying for the executive secretary position—were rated more positively than the expected targets with the same qualifications.

Yet again, the present data point to another level of complexity in the judgment process. Despite the subjectively favorable evaluations of "flowers blooming in winter" objective judgments—those more directly tied to external reality and not subject to category-based shifts in meaning—indicated more favorable perceptions of the expectancy-consistent applicants (the "flowers blooming in spring"). Participants evaluated female secretaries and White male chiefs of staff relatively favorably on the externally anchored, objective indexes.

What we suggest, then, is that individuals from negatively

stereotyped, low-status groups face a rather different set of judgment outcomes than do individuals from positively stereotyped, high-status groups. Although some points in the judgment process may appear to be favorable to these low-status individuals (e.g., the fact that they are held to lower standards), these low standards are themselves patronizing in nature (Foschi, 1992). Furthermore, the ultimate outcome for a low-status person is a longer, more difficult trek to document ability and evaluations that are objectively less positive than those awarded to similarly credentialed individuals from high-status groups. The fact that evaluators are subjectively positive toward low-status targets provides little consolation here, because these subjective evaluations involve the implicit comparison with the (lower) within-category standard: "She's very good for a woman, but that is quite different from very good for a man."

We reported two studies to document these mechanisms and effects, but much further work is needed to replicate, extend, and test the limits of these findings. We suggest that it would be valuable in future work to (a) continue to examine the effects of job type and other contextual cues to applicant appropriateness, (b) manipulate applicant quality, (c) measure the extent to which applicants do or do not meet standard criteria, (d) focus more explicitly on the processes involved in the setting of judgment standards, and (e) test the effects of judges' own category membership on judgment. With regard to this latter point, research in the status characteristics tradition has suggested that gender-based double standards for competence are more likely to be held by men than by women (Foschi et al., 1994). We found no meaningful participant gender effects in Study 1, but it is unclear to what extent this was due to a lack of statistical power. Furthermore, our participant population did not allow for a test of participant race effects in Study 2.

We also described the *simultaneous* setting of minimum and ability standards and discussed the *translation* of subjective into objective ratings. But our research relied on between-participants designs that do not allow for an explicit test of these operations. For example, we were unable to test whether low minimum standards were directly correlated with high ability standards or high subjective evaluations with lower objective evaluations. A within-participants design would provide a more rigorous test of our model, but we suspect that it would take a highly inventive researcher to develop measures of minimum standards and ability criteria, in both subjective and objective units, that have construct equivalence yet do not tire or alert participants to the research hypotheses.

We hope that we have provided some resolution to the seemingly paradoxical statements that run through Carter's (1993) book—a flower blooming in winter may still have to work hard to prove its worth!—and that we have raised a number of questions that others deem worthy of further research. One of the more general lessons that can be taken from our studies is that the form a question takes, and the specific type of judgment respondents are asked to make, may have large effects on the pattern of data obtained. By incorporating different forms and types of judgment in a single study, we may gain a more complex yet more complete picture of the judgment process.

References

- Allport, G. W., & Postman, L. (1947). *The psychology of rumor*. New York: Holt.

- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, 65, 272-281.
- Basow, S. A. (1992). *Gender: Stereotypes and roles* (3rd ed.). Belmont, CA: Brooks/Cole.
- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M., Jr. (1977). *Status characteristics and social interaction: An expectation-states approach*. New York: Elsevier.
- Berger, J., Rosenholtz, S. J., & Zelditch, M., Jr. (1980). Status organizing processes. *Annual Review of Sociology*, 6, 479-508.
- Berger, J., Wagner, D. G., & Zelditch, M., Jr. (1985). Introduction—Expectation states theory: Review and assessment. In J. Berger & M. Zelditch, Jr. (Eds.), *Status, rewards, and influence: How expectations organize behavior* (pp. 1-72). San Francisco: Jossey-Bass.
- Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In Y.-T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 87-114). Washington, DC: American Psychological Association.
- Biernat, M., & Kobrynowicz, D. (1997). *Shifting standards and the assimilation and contrast of judgments to group stereotypes*. Manuscript submitted for publication.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66, 5-20.
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60, 485-499.
- Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology*, 29, 166-202.
- Biernat, M., Vescio, T. K., & Manis, M. (in press). Judging and behaving toward members of stereotyped groups: A shifting standards perspective. In C. Sedikides, J. Schopler, & C. Insko (Eds.), *Intergroup cognition and intergroup behavior*. Hillsdale, NJ: Erlbaum.
- Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, 1, 319-322.
- Carter, S. L. (1993). *Reflections of an affirmative action baby*. New York: Basic Books.
- Davis, G., & Watson, G. (1982). *Black life in corporate America: Swimming in the mainstream*. Garden City, NY: Anchor Press/Doubleday.
- Deaux, K. (1976). Sex: A perspective on the attribution process. In J. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 335-352). Hillsdale, NJ: Erlbaum.
- Deaux, K. (1984). From individual differences to social categories: Analysis of a decade's research on gender. *American Psychologist*, 39, 105-116.
- Deaux, K., & Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29, 80-85.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology*, 34, 590-598.
- Foddy, M., & Smithson, M. (1989). Fuzzy sets and double standards: Modeling the process of ability inference. In J. Berger, M. Zelditch, Jr., & B. Anderson (Eds.), *Sociological theories in progress: New formulations* (pp. 73-99). London: Sage.
- Foschi, M. (1992). Gender and double standards for competence. In C. L. Ridgeway (Ed.), *Gender, interaction, and inequality* (pp. 181-207). New York: Springer-Verlag.
- Foschi, M., & Foddy, M. (1988). Standards, performances, and the formation of self-other expectations. In M. Webster, Jr., & M. Foschi (Eds.), *Status generalization: New theory and research*. Stanford, CA: Stanford University Press.
- Foschi, M., Lai, L., & Sigerson, K. (1994). Gender and double standards in the assessment of job applicants. *Social Psychology Quarterly*, 57, 326-339.
- Gaertner, S. L., & Dovidio, J. (1986). The aversive form of racism. In J. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). New York: Academic Press.
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, 6, 298-304.
- Hansen, R. D., & O'Leary, V. E. (1985). Sex-determined attributions. In V. E. O'Leary, R. K. Unger, & B. S. Wallston (Eds.), *Women, gender, and social psychology* (pp. 67-99). Hillsdale, NJ: Erlbaum.
- Herr, P. M. (1986). Consequences of priming: Judgments and behavior. *Journal of Personality and Social Psychology*, 51, 1106-1115.
- Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, 19, 323-340.
- Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75-123). New York: Guilford Press.
- Hilton, J. L., & von Hippel, W. (1990). The role of consistency in the judgment of stereotype-relevant behaviors. *Personality and Social Psychology Bulletin*, 16, 430-448.
- Jackson, L. A., Sullivan, L. A., & Hodge, C. (1993). Stereotype effects on attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology*, 65, 69-84.
- Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219-266). New York: Academic Press.
- Jones, E. E., Davis, K. E., & Gergen, K. (1961). Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology*, 63, 302-310.
- Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology*, 52, 536-546.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39, 821-831.
- Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals. *Journal of Experimental Social Psychology*, 18, 23-42.
- Manis, M., Biernat, M., & Nelson, T. (1991). Comparison and expectancy processes in human judgment. *Journal of Personality and Social Psychology*, 61, 203-211.
- Manis, M., Nelson, T. E., & Shedler, J. (1988). Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology*, 55, 28-36.
- Manis, M., & Paskewitz, J. (1984). Judging psychopathology: Expectation and contrast. *Journal of Experimental Social Psychology*, 20, 363-381.
- Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, 51, 493-504.
- Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's ability to expend effort in forming an

- impression. *Journal of Personality and Social Psychology*, 59, 27–37.
- Monteith, M. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice reduction efforts. *Journal of Personality and Social Psychology*, 65, 469–485.
- Newman, S. E., & Benassi, V. A. (1989). Putting judgments into context: Contrast effects. *Journal of Personality and Social Psychology*, 56, 876–889.
- Pettigrew, T., & Martin, J. (1987). Shaping the organizational context for Black American inclusion. *Journal of Social Issues*, 43, 41–78.
- Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Corrections for context-induced contrast. *Journal of Experimental Social Psychology*, 29, 137–165.
- Pugh, M. D., & Wahrman, R. (1983). Neutralizing sexism in mixed-sex groups: Do women have to be better than men? *American Journal of Sociology*, 88, 746–762.
- Ridgeway, C. L. (1982). Status in groups: The importance of motivation. *American Sociological Review*, 47, 76–88.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39, 590–598.
- Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 217–245). Hillsdale, NJ: Erlbaum.
- Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, 22, 507–519.
- Wagner, D. G., Ford, R. S., & Ford, T. W. (1986). Can gender inequalities be reduced? *American Sociological Review*, 51, 47–61.
- Webster, M., Jr., & Foschi, M. (1988). Overview of status generalization. In M. Webster, Jr., & M. Foschi (Eds.), *Status generalization: New theory and research* (pp. 1–20, 477–478). Stanford, CA: Stanford University Press.
- Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68, 36–51.
- Yarkin, K. L., Town, J. P., & Wallston, B. S. (1982). Blacks and women must try harder: Stimulus person's race and sex attributions of causality. *Personality and Social Psychology Bulletin*, 8, 21–24.

Received June 5, 1995

Revision received May 30, 1996

Accepted July 2, 1996 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.