Gender stereotypes regarding task competence may lead perceivers to set different standards for diagnosing competence in women versus men. Specifically, stereotypes may prompt lower minimum standards (or initial screening criteria) but higher confirmatory standards for women than men (Biernat & Kobrynowicz, 1997). In two studies simulating hiring decisions, predictions were that women would be (1) more likely than men to make a short list for a job but (2) less likely than men to be hired for the same job. Results were generally consistent with predictions only among female participants (Studies 1 and 2), among those exposed to a female experimenter (Study 1), and among those held accountable for their decisions (Study 2). The role of motivational factors in the setting of standards is discussed.

An important aspect of leadership is the perception of task competence, and a considerable amount of research documents that men are generally perceived as more competent than women. Indeed, the very essence of gender stereotypes defines men as instrumentally competent and agentic compared to women (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Kite & Deaux, 1987; Smith & Midlarksy, 1985; Spence & Sawin, 1985). This stereotype also affects evaluations of individual women. For example, work products from “masculine” or neutral settings tend to be judged higher in quality when attributed to
men than to when attributed to women (Swim, Borgida, Maruyama, & Myers, 1989; see also Biernat & Manis, 1994), and a number of studies have demonstrated that perceivers assume a man is more likely than a woman to have ability at a novel task (Balkwell & Berger, 1996; Gerdes & Garber, 1983; Hansen & O’Leary, 1985; Heilman & Guzzo, 1978; Pugh & Wahrman, 1983). Recent research also suggests that when women attempt to display a competent impression through self-promotion, they are liked and hired less often than comparable males (Rudman, 1998; Rudman & Glick, this issue).

But gender stereotypes regarding task competence do not uniformly lead to the perception of an individual woman as less competent than comparable men. Some research has found, for example, that assertive women are viewed more positively than assertive men (Linehan & Seifert, 1983), and other research has reported equivalent responses to competent men and competent women (e.g., Foddy & Smithson, 1999). Some insight into these conflicting findings may be gained from considering the shifting standards model (Biernat & Manis, 1994; Biernat, Manis, & Nelson, 1991), which provides the theoretical basis for the research reported here.

**Shifting Standards and Social Judgment**

The shifting standards model suggests that when we judge individual members of stereotyped groups on stereotyped dimensions, we compare them to within-category judgment standards. For example, given stereotypes that men are better leaders than women (Eagly & Johannesen-Schmidt, this issue; Eagly, Karau, & Makhijani, 1995), we are likely to judge the leadership competence of a particular woman relative to (lower) standards of competence for women and the leadership competence of a particular man relative to (higher) standards of competence for men. The result is that evaluations of men and women may not be directly comparable: “Good” for a woman does not mean the same thing as “good” for a man.

The operation of stereotype-based standard shifts can be seen when one compares judgments that are made on subjective rating scales to those made on more objective rating scales. A “subjective scale” refers to any rating system in which the units of judgment have no ties to external reality. For example, Likert-type scales, semantic differentials, and many instruments that display trait judgments as continua (e.g., very incompetent to very competent) are subjective or slippery in that their rating points can be differentially defined and adjusted (e.g., across time, across perceivers, and across targets). On the other hand, an “objective scale” refers to an externally anchored, “common rule” scale consisting of judgment units that maintain a constant meaning across contexts and/or types of targets. For example, judging the heaviness of an object in pounds would be an objective judgment, whereas rating the object on a light to heavy scale would be a subjective judgment. In evaluating people, objective scales may include estimates of standardized test
scores or grades (to assess competence stereotypes), monetary or time judgments (to assess salary, effort, or worth), or rank orderings (which place a target individual in a readily interpretable position relative to a given frame of reference; see Biernat, Crandall, Young, Kobrynowicz, & Halpin, 1998; Biernat & Manis, 1994). In short, objectivity refers to a judgment whose meaning can be readily interpreted because of the external frame of reference.

A key prediction of the shifting standards model is that objective or externally anchored judgments are more likely than subjective judgments to reveal the influence of stereotypes. That is, judgments of individual group members will generally be assimilated to group stereotypes when objective response scales are used, but (because of within-category standard shifts) will show less strong, null, or contrastive influences of stereotypes when subjective scales are used.

For example, in a study of U.S. Army captains’ perceptions of one another’s leadership competence, participants judged their male peers to be decisively better leaders than their female peers in objective units (rankings), but not in subjective evaluations (outstanding to needs improvement ratings; Biernat et al., 1998). This discrepancy between objective and subjective evaluations occurred because the rankings forced judges to array targets on a single evaluative dimension and were therefore sensitive to available stereotyped perceptions, but subjective evaluations concealed these perceptions because they were based on the use of within-rather than across-category judgment standards: A male and female officer could each be labeled a “very good” leader, but because of sex-specific standards, the label translated into a higher rank when the target was male.

The signature shifting standards pattern (stronger stereotyping effects on externally anchored than on subjective response scales) has been documented in a variety of judgment domains for both sex and racial groups. Judges seem to shift their standards in rating women versus men on the dimensions of height, weight, income, verbal ability, writing competence, aggression, parenting involvement, leadership, and job-related competence; standards for Blacks versus Whites similarly shift in the domains of verbal ability, athleticism, and job-related competence (Biernat et al., 1991; Biernat et al., 1998; Biernat & Kobrynowicz, 1997; Biernat & Manis, 1994; Kobrynowicz & Biernat, 1997). The phenomenon of shifting standards can account for the fact that an individual woman may not be perceived as (subjectively) less competent than a man. Because of gender stereotypes, women and men are judged relative to sex-specific standards.

**Setting Judgment Standards**

Thus far, we have described how the tendency to shift judgment standards may result in an apparent masking of stereotyping effects on subjective judgment scales. In another line of research that has emerged from the shifting standards model, we have directly examined the setting of judgment standards. Do judges
actually show evidence of holding males and females, or members of any two contrasting stereotyped groups, to different standards?

With respect to judging task competence, the shifting standards model suggests that women are held to lower standards than men. In our first study testing this hypothesis, we asked participants to review the résumé of a male or female job applicant and to indicate the number of job-relevant skills they would require of this candidate before they would feel that he or she “meets the minimum standard” to perform the skill (Biernat & Kobrynowicz, 1997). Both male and female judges set lower minimum requirements for women than for men. In other words, participants conveyed their expectations that women would not be as competent as men.

Despite these lower minimum standards for women, the same study documented that women were held to higher confirmatory standards than men. When participants were asked what skill level they would require to be certain that the applicant “had the ability to perform the job,” more evidence of skill was required of women than men. These data suggest that whereas it may be relatively easy for a woman to meet minimum group-based standards, she must work harder to confirm that her performance is ability-based. Similar results emerged for Black versus White male applicants: Blacks were held to lower minimum standards but higher confirmatory standards than Whites (Biernat & Kobrynowicz, 1997, Study 2).

This finding of stricter confirmatory standards for women and Black men rings true to anecdotal accounts suggesting that women and Blacks may need to “work twice as hard to be perceived as half as good” as White men (see Carter, 1993) and with research on double standards for competence derived from expectation states theory (Foddy & Smithson, 1989; Foschi, 1989, 2000; Foschi & Foddy, 1988). Stereotypes about competence lead perceivers to have low expectations for women, as evidenced in the setting of lower minimum standards. But because of these low standards, it takes more evidence to confirm a woman’s task competence.

Standards in Hiring Decisions

The present studies were designed to extend our work on gender-based minimum and confirmatory competence standards by examining whether these standards map onto important steps in the hiring process. Specifically, minimum standards may correspond to the point in a decision process at which one determines which candidates are still in the running for a position, as in the creation of a short list, and confirmatory standards may correspond to the decision of which person from the short list should, in fact, be hired. If this is the case, our prior work suggests that female applicants will be more likely than male applicants to make a short list, but male applicants will nonetheless be more likely than females to be hired. That is, lower minimum standards for women should make it easier to pass an initial screening process (the short-listing procedure), but higher confirmatory
standards should make it more difficult for women to pass the scrutiny required to be hired.

We present two studies designed to test these hypotheses, each involving a simulation of the hiring process. In the first, participants are asked to consider a single applicant (male or female) and to set standards for short listing versus hiring. In the second study, participants are asked to consider a pool of applicants from which they must first produce a short list, and, ultimately, a hire. These studies also examine potential moderators of the predicted judgment patterns, including participant sex, experimenter sex, and accountability instructions.

Study 1

Method

Participants and procedure. One hundred seventy-five introductory psychology students (88 females, 87 males) from the University of Kansas participated in exchange for course credit and completed the procedure under the guidance of one of two female or one of three male experimenters. Two participants (one female, one male) were later dropped for not following instructions. Participants were asked to imagine that they were in charge of evaluating applicants for a job and were given a folder containing instructions, a job description, a resumé, and an evaluation form. Participants in the short list standard condition read that a short list is designed to “narrow down a large pool of applicants into a small group who are still in the running for a position. To be placed on the short list, an applicant must pass an initial screening for the position.” Furthermore, “only one person from the short list is ultimately hired for the job.” Participants in the hiring standard condition were simply told that their task was to decide whether to hire the applicant whose resumé they would be reviewing; they were provided with no information on short listing.

After reading the instructions, all participants reviewed the same job description. The job title, however, varied, so that the applicant was applying for either an executive secretary position (feminine) or an executive chief of staff position (masculine). Following the job description, all participants reviewed the same resumé, except that half received a resumé with a man’s name (Kenneth Michael Anderson) and the other half a woman’s name (Katherine Marie Anderson). The names Kenneth and Katherine were chosen because they convey a nonspecific age (Kasof, 1993) and roughly equal intellectual competence (Mehrabian, 1990; J. Kasof, personal communication, January 19, 2001). The resumé was designed to be of moderate caliber: the applicant had a 2.9 grade point average (out of a possible 4.0) and some work experience.
Dependent measures. To assess short-listing/hiring standards, participants were asked “what level of performance would you require of this applicant in order for you to place him/her on the short list (or hire him/her) for this position?” These instructions (short list vs. hire) were repeated for each item. The standards index was an average of 11 items, each assessed in percentile or percentage form: the percentile ranking the applicant would need on a general standardized ability test, the percentile ranking the applicant would need from the writers of his/her three letters of recommendation, and the estimated scores (in percentage correct) the applicant would need on standardized tests of decision making, interpersonal relations, leadership, monitoring, motivation, oral communication, problem solving, planning, and seeking/accepting advice. These estimates represent objective scales and should thus reveal gender-based differences in standards of competence. Coefficient alpha for this index was .87. Lastly, regardless of condition, all participants indicated whether they would short-list and whether they would hire the applicant.

Results and Discussion

Setting standards. What skill level did participants require of the applicants? The standards index was submitted to an Applicant Gender × Position × Standard (short list, hire) × Participant Sex between-subjects analysis of variance (ANOVA). First, standards were set higher for the chief of staff ($M = 85.27$) than for the secretarial position ($M = 83.11$), $F(1, 156) = 6.73$, $p < .05$, and they were also set higher when the standard was to hire ($M = 85.64$) rather than short list ($M = 82.76$) the applicant, $F(1, 156) = 11.39$, $p < .001$. The only other reliable effect was the Applicant Gender × Standard × Participant Sex interaction, $F(1, 156) = 3.76$, $p = .05$.

Figure 1 depicts this interaction by displaying the Applicant Gender × Standard means separately for female (top panel) and male (bottom panel) participants. As can be seen in the top panel, female participants demonstrated the predicted pattern of effects: Katherine was held to lower standards for short listing but higher standards for hiring relative to Kenneth. A contrast testing the Applicant Gender × Standard interaction for female participants only was significant, $F(1, 156) = 3.88$, $p = .05$. Additional contrasts indicated that the Kenneth-Katherine difference under short-listing standards was not reliable ($p > .25$), though that difference approached significance under hiring standards ($p < .06$, one-tailed). Among male participants, the interaction was not reliable, $F < 1$. Interestingly, standards were reliably higher for hiring than for short listing in all cases ($ps < .05$) except when Kenneth was judged by female participants.

Decision to short-list and hire. Regardless of whether participants were asked to consider short-listing or hiring standards, they all made both decisions. Proportions of short-listing and hiring decisions were submitted to an Applicant Gender × Position × Standard × Participant Sex × Experimenter Sex × Decision
mixed-model ANOVA. Experimenter sex was included as it became clear this factor importantly influenced judgments (though it did not moderate the standard-setting effects described above). Not surprisingly, there was a main effect for Decision, $F(1, 140) = 76.03, p < .0001$, such that applicants were short-listed 73.8% of the time but hired at a rate of 37.8%. In addition, the Decision × Applicant Gender × Experimenter Sex interaction was significant, $F(1, 140) = 6.20, p < .05$, and is depicted in Figure 2.

As can be seen in the top panel of the figure, the pattern of results was consistent with predictions only when the experimenter was female: Katherine was more likely to be short-listed but less likely to be hired than Kenneth. A contrast testing this Decision × Applicant Gender interaction when the experimenter was female was reliable, $F(1, 140) = 5.95, p < .05$, though the Kenneth-Katherine difference
approached significance only in the case of hiring decisions, $p < .06$ (one-tailed). When the experimenter was male, the interaction was not significant, $F < 1$, nor were the simple effects of applicant gender on either short-listing or hiring decisions ($ps > .08$). Further analyses indicated that the Applicant Gender × Experimenter Sex interaction was significant only for the hiring decision, $F(1, 140) = 5.50, p < .05$; applicant gender did not reliably affect short-listing decisions for either sex of experimenter.

**Summary.** These data suggest that on reports of standards, only female participants showed some evidence of the predicted pattern of effects—more lenient standards for short listing but harsher standards for hiring a female (though the leniency toward females in standards for short listing was not reliable). In terms of short-listing and hiring decisions themselves, only those exposed to a female
experimenter (regardless of their own sex) showed the predicted pattern—more short listing but less hiring of the female relative to the male applicant. Once again, however, only the tendency for harshness toward females in hiring was reliable in this case. Male participants did not differentially set standards based on applicant sex, and those exposed to male experimenters showed a nonsignificant trend to favor the female in both short listing and hiring.

What are we to make of these participant and experimenter sex effects? Based on our prior work (Biernat & Kobrynowicz, 1997), we predicted no participant sex effects (or had we been forced to choose, our money would have been on men showing the shifting standards effect more strongly than women). There are, however, some precedents in the literature for females’ relative harshness toward other females (Broder, 1993; Goldberg, 1968; Miller & McReynolds, 1973; Rudman, 1998). We will return to this point in the General Discussion.

The unpredicted experimenter sex effect also forces a post hoc explanation. We suggest one possibility: Participants may have felt more “accountable” for their decisions when faced with a female experimenter, perhaps because women are assumed to scrutinize hiring decisions more carefully than men. Accountability refers to “the implicit or explicit expectation that one may be called on to justify one’s beliefs, feelings, and actions to others” (Lerner & Tetlock, 1999, p. 255) and motivates decision makers to reach justifiable and optimal solutions (Lerner & Tetlock, 1999; Tetlock, 1992). Some evidence consistent with this post hoc account can be gleaned from correlations between the standards participants set and the decisions they made. In general, one would expect that higher standards are associated with less favorable short-listing and hiring decisions. When the experimenters were female, this was the case: In the short-listing condition, standards and short-listing decisions were negatively related, $r (n = 39) = –.33, p < .05$, and in the hiring condition, a similar though weaker trend emerged, $r (n = 29) = –.20, ns$. When the experimenters were male, standards and decisions were unrelated in both the short-listing condition, $r (n = 50) = –.05$ and in the hiring condition, $r (n = 53) = –.01$, both ns. Although these sets of correlations were not reliably different, the pattern of greater consistency for female experimenters is at least suggestive of an accountability explanation. In Study 2, we explicitly manipulated accountability to better explore this possibility.

**Study 2**

**Method**

This study aimed to model short-listing and hiring decisions more explicitly by presenting participants with a pool of applicants to consider rather than just one resumé. Also in this study, the position was designed to be clearly masculine in nature (mechanical engineer).
Pretesting. Sixteen resumés were created and pretested for quality (without names attached). From this set, we created seven pairs; in each pair, the resumés did not differ in perceived quality. For the main study, we assigned one resumé from each pair a female name and the other a male name, then switched the genders in a second set of resumés (see below). Three letters of recommendation were also pretested to be of equal quality, all \( ps > .49 \).

Participants and procedure. Sixty-four introductory psychology students (25 females, 39 males) from the University of Kansas participated in exchange for course credit. Participants were told that they would follow the steps of a hiring process by considering applicants for a mechanical engineering internship program. They first viewed the resumés of 14 applicants and from that pool selected 3 applicants to short-list. The experimenter noted the short-listed candidates and returned with three letters of recommendation ostensibly written for each of the candidates. These letters had been prepared such that each version of the letter was rotated among the short-listed applicants according to a random-number sequence. After reviewing the letters for each of the short-listed applicants, participants selected one applicant to hire.

The resumés were presented in two different orders, and for each matched pair, the gender associated with one applicant was switched in a second presentation set. Thus, within each pair, each resumé was paired with a female name in one set and a male name in the other. As will be seen below, this manipulation had no effect on judgment patterns.

Because we did not anticipate the experimenter sex effect in Study 1, we did not attend to this factor as we conducted Study 2: All but 5 of the participants completed the procedure with male experimenters. This study did, however, include an explicit manipulation of accountability, the factor we believe may be responsible for the earlier findings. All participants were told that the Engineering Physics department on campus was interested in students’ perspectives on applying for jobs. To the “accountable” participants we added that the department would make recommendations based on their evaluations and that they “should be able to justify any decision” they made. In contrast, participants in the “no accountability” condition were told that their evaluations would simply be compiled with those of others. Participants in the accountable condition also signed their decision sheet and placed it in an envelope addressed to the head of the engineering department; participants in the no-accountability condition placed the unsigned sheet in a blank envelope.

Results and Discussion

Proportions of women short-listed and hired were first arcsine-transformed, then submitted to a Participant Sex × Accountability × Presentation Set × Decision
(short-list/hire) repeated-measures ANOVA. Two effects emerged: A Decision × Participant Sex interaction, $F(1, 57) = 8.71, p < .01$, and a Decision × Accountability interaction, $F(1, 57) = 3.64, p = .06$.

The former interaction is depicted in Figure 3 (means are untransformed proportions). Male and female participants did not differ in their proportions of short-listing female applicants, $F < 1$, and in neither case did the proportions significantly differ from chance (.50). A rather dramatic pattern appeared, however, on hiring decisions, where the participant sex effect was significant, $F(1, 57) = 9.32, p < .01$. Men were much more likely than women to hire a female applicant. The overhiring of women by men was not significantly greater than chance, $t < 1$, but the underhiring of women by women (at a rate of 28%) was reliably lower than chance, $t(24) = 2.40, p < .03$. Further analyses indicated that the difference between short listing and hiring of women was reliable for both female and male participants, $p$s $< .05$.

The other effect of interest in the overall analysis was the marginally significant Decision × Accountability interaction. Accountability had no effect on short-listing decisions, $F < 1$ ($Ms = .48$ and .49) but did tend to affect hiring decisions, $F(1, 57) = 3.79, p < .06$. Accountable participants were somewhat less likely than nonaccountable participants to hire a female for the job ($Ms = 41\%$ and 58\%, respectively). Thus accountability, the phenomenon we believe to be responsible for harshness toward the female applicant among female experimenters in Study 1, decreased the likelihood that a woman would be hired, relative to nonaccountability.

![Fig. 3. Proportion of female applicants short-listed and hired, by participant sex, Study 2.](image-url)
Summary. We found mixed support for our hypothesis that women would be more likely than men to be short-listed for jobs but less likely to be hired. In this multiapplicant study, short-listing decisions were not affected by applicant gender. Instead, the proportion of females on the short list hovered near 50%. This suggests that evaluators may be oriented toward a roughly equal sex representation on the short list (perhaps especially when a pool of applicants contains equal numbers of women and men) rather than toward favoring those held to a lower standard. Study 2 was consistent with Study 1, however, in demonstrating that only female participants showed evidence of antifemale hiring selections. Because most of the participants in this study worked with a male experimenter, we could not test for experimenter sex effects as we did in Study 1. The accountability manipulation did reveal, however, that it was under accountability instructions that female applicants were particularly unlikely to be hired.

General Discussion

The goal of these studies was to examine how stereotypes of women as less competent than men influence the standards judges set in the hiring process. We predicted that because of expected gender differences in task competence, women would be held to lower standards than men and therefore would be more likely to make a short list for a position. At the same time, because more evidence of ability is required to overcome low expectations, women would be less likely than men to meet hiring standards.

The two studies presented here offered partial support for our predictions. First, although all participants tended to be more lenient toward women than men in terms of short listing, this effect was never statistically reliable. Thus, the one potential benefit to women of being negatively stereotyped (i.e., being held to lower standards for short listing) was not realized. Second, there was evidence that judges set harsher standards for hiring a female applicant in Study 1, but this was only the case among female participants. Hiring decisions were also less favorable for female applicants relative to male applicants in Study 1, but only when the experimenter was female. Evidence that a female applicant was less likely to be hired was also found in Study 2, but again, this pattern was evident only among female participants.

Participant Sex Effects

As indicated earlier, our expectation was that no sex differences would emerge or that men, because they tend to endorse gender stereotypes regarding agentic competence more strongly than women (see Rudman & Glick, this issue), would show stronger antiwoman hiring bias. Indeed, there are a number of studies and meta-analyses documenting that when sex of judge matters, men tend to be harsher
than women toward women (Carli, this issue; Eagly, Makhijani, & Klonsky, 1992; Schein, this issue; Schein & Mueller, 1992; Schein, Mueller, Lituchy, & Liu, 1996). Our unexpected findings here, however, led us back to the literature on gender-based judgment, where we also found evidence that women are not immune to bias against women (Broder, 1993; Costrich, Feinstein, Kidder, Marecek, & Pascale, 1975; Goldberg, 1968 [cf. Swim et al., 1989]; Hagen & Kahn, 1975; Rudman, 1998).

For example, Rudman (1998, Study 2) examined perceivers’ impressions of self-promoting males and females. Among a number of complex results, she found that when participants were outcome-dependent on the target being judged, self-promoting men were judged more hireable than self-promoting women, particularly by women. Furthermore, in a decision similar to hiring, participants were asked to select a partner for a “Jeopardy!” game (Rudman, 1998, Study 3). Given the choice between a male or female self-promoter, men selected evenly across gender, whereas women always chose the male. Thus, at least under conditions of outcome dependence, women (but not men) found the self-promoting woman less desirable and hireable than the self-promoting man. Other studies have documented that females are more likely to exclude a competent female than a competent male from their group (Hagen & Kahn, 1975) and that females dislike an aggressive woman more than an aggressive man (Costrich et al., 1975). Still other research suggests that these effects are not limited to undergraduates. Analyzing reviews of grant proposals to the National Science Foundation (in economics), Broder (1993) found that female reviewers gave significantly lower scores than male reviewers to female-authored proposals.

The gist of these studies is that asserting one’s skills and qualifications may still be problematic for women, even if their audience is other women. A number of explanations have been put forth in an attempt to explain this phenomenon. The “Queen Bee Syndrome” (Staines, Travis, & Jayaratne, 1974) suggests that successful women may tend to view other women as intruders or competitors given the zero-sum nature of women’s positions in organizations. In a different vein, Mathison (1986) suggested that women may devalue the competence of assertive women as a means of reducing the dissonance they feel over not being assertive themselves. Still another explanation is that women may believe that their own credibility will be questioned unless they judge other women harshly: If an unqualified woman gets ahead, it may reflect badly on all women (Broder, 1993). A related account is that women hold other women to higher standards because they know that women need to be exceptional to succeed.

It is worth remembering that it was the male participants in our studies whose judgments were inconsistent with hypotheses. In this sense, their behavior deserves explanation as well. Male participants consistently showed a profemale bias in hiring, except when they were accountable. We suggest that male participants may have been motivated to appear nonsexist or to be fair. These motives may have
translated into (over)correcting for sexist bias and overhiring females. Interestingly, fairness concerns may also explain female participants’ greater harshness toward females: Being fair may have meant not favoring women, and in their attempts to avoid appearing biased toward their ingroup, they, too, may have overcompensated by underhiring women. Thus, rather than male participants supporting competent women, or female participants punishing them, all participants may have been eschewing ingroup bias or trying to be fair.

Future work in which fairness motives are explicitly manipulated is needed to test these ideas. Furthermore, it will be important to demonstrate whether participant sex effects of this sort are limited to the sample and context studied here (undergraduates in a laboratory setting). Although debriefing and suspicion checks indicated that participants in these studies were neither aware of our hypotheses nor particularly attuned to gender, studying decision makers in the real world may provide additional insight into how hiring standards may differ based on both applicant and evaluator sex.

**Experimenter Sex Effects: Accountability Differences?**

The other unexpected finding in our research was the experimenter sex effect in Study 1 (see Figure 2). Female applicants were slightly more likely than male applicants to make the short list but less likely to be hired, but only when the experimenter was female. Why was this the case? We must admit that we have no firm answer to this question, and because we were unable to test for a replication in Study 2, we offer only limited speculation on the issue.

We suggested in Study 1 that participants may have felt more accountable toward a female than male experimenter. Indeed, the consistency between standards and short-listing/hiring decisions in the female experimenter conditions suggests that female experimenters may have made participants feel more identifiable, more subject to close scrutiny, and more in need of justifying their decision—important aspects of being accountable (Lerner & Tetlock, 1999). More generally, a female experimenter may have heightened perceivers’ sensitivity to gender as a cue in short-listing or hiring decisions. The research literature suggests that accountability may amplify bias by encouraging “indiscriminate use” of what may be available, yet irrelevant, information (Lerner & Tetlock, 1999, p. 264). In our studies, the available but irrelevant information was applicant gender, and to the extent that female experimenters made participants feel accountable, participants were more likely to use gender stereotypes in their decisions. In Study 2, in which accountability was explicitly manipulated, we found similar use of gender stereotypes: Accountable participants tended to underhire women.

The results of both studies also highlight one additional point: Accountability affected hiring but not short-listing decisions. In Study 1, the short-listing patterns were identical for female and male experimenters (see left panel of Figure 2), and
in Study 2, they were identical for accountable and nonaccountable participants. This suggests that participants were willing to consider females favorably when the decision was of little consequence (when a short list was being created; when accountability was low) but that they tended to shift toward a stereotypical, promale bias when the consequences were high (hiring under accountable conditions).

Conclusions

This line of research extends the shifting standards model by directly examining how judgment standards affect hiring decisions. Previous research has shown that judges set lower minimum but higher confirmatory standards for women relative to men (Biernat & Kobrynowicz, 1997), and the present data show that men and women are, in fact, held to different standards in the hiring process. In general, it was somewhat easier for women to make a short list, a fact that we attribute to the low standards or expectations perceivers hold about a woman’s task competence. When it came to hiring, however, women were held to higher standards, though only by female participants, those exposed to a female experimenter (Study 1), and those held accountable for their decisions (Study 2).

The fact that our predictions did not hold for all participants under all conditions points to the inherent complexity in carrying any judgment model into a real-world context. The shifting standards model has been largely cognitive in its emphasis, but simply holding the stereotype that men are more task competent than women is not sufficient to explain the judgmental patterns we observed here: Motivational factors must also be taken into account. For example, the desire to be fair, concerns about appearing sexist, or egalitarian commitment to a gender-balanced workforce may affect how competence stereotypes are translated into consequential decisions (e.g., hiring, promoting, firing). Extensions of this research in which the gender makeup of the initial applicant pool is varied could address some of these issues. In Study 2, equal numbers of women and men applied for the “masculine” position, and this may have heightened participants’ sensitivity to gender and gender bias. With a skewed pool that better reflects the gender composition of an occupation (i.e., in which the large majority of applicants are male), male participants in particular may feel less pressure to place women on the short list and, consequently, hire them.

Furthermore, this research need not be limited to how minimum and confirmatory standards map onto short-listing and hiring decisions. For example, in many settings, the standards for hiring an individual may be quite different than those for promotion (e.g., to associate professor in academe or to partner in a law firm). Presumably, standards are lower to hire than to promote, and these standards may differ further based on an applicant’s social category memberships. Our prediction is that, ceteris paribus, promotion standards are harsher for women or racial
minorities, even when initial hiring standards may be less strict. Data regarding women's actual representation in academic fields, for example, are consistent with the notion that associate professorships are less likely to accrue to women, relative to their presence in the assistant professor pool (National Science Foundation, 1999).

In conclusion, these studies highlight the theoretical breadth of the shifting standards model and the importance of considering how motives may influence the application of group stereotypes to individual actors. To better understand the motivational forces affecting evaluators, future research should move beyond the use of undergraduates in role play to decision makers in the real worlds of industry and academe. We hope that this work leads to a firmer understanding of the subtle and not-so-subtle ways in which gender influences assumptions about competence and decisions involving hierarchy and leadership.

References


MONICA BIERNAT is Associate Professor of Psychology at the University of Kansas. She earned her PhD in social psychology at the University of Michigan in
1989 and was on the faculty at the University of Florida before moving to Kansas in 1992. Her research examines the processes of stereotyping and prejudice and focuses specifically on how stereotypes guide (and distort) judgments of individual members of stereotyped groups. She is currently associate editor of the Journal of Personality and Social Psychology and was the 1998/1999 winner of the American Psychological Association Distinguished Scientific Award for Early Career Contribution to Psychology in the area of Social Psychology.

KATHLEEN FUEGEN is a doctoral student in the Social Psychology program at the University of Kansas. She received her BA in psychology from the University of Northern Iowa in 1994 and her MA from the University of Kansas in 1998. Her research interests include stereotyping, prejudice, and the perception of discrimination.