Journal of Experimental Social Psychology 48 (2012) 271–276

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Reports

Interpreting and reacting to feedback in stereotype-relevant performance domains ☆

Monica Biernat *, Kelly Danaher

University of Kansas, USA

ARTICLE INFO

Article history:
Received 21 September 2010
Revised 12 August 2011
Available online 26 August 2011

Keywords:
Stereotypes
Feedback
Performance
Standards

ABSTRACT

Feedback on task performance is often phrased in subjective language (e.g., "not bad!"), but how do recipients understand or translate that feedback into a clear, objective, performance metric? We suggest that when feedback is provided in a stereotype-relevant domain, translation is made with reference to stereotyped expectations for one's group. In Study 1, women and men were exposed to negative subjective feedback about their performance on a leadership task; in Study 2, Black and White participants were provided subjective negative feedback, or no feedback, on an academic writing task. Women relative to men, and Black students relative to White students, translated their feedback to indicate *objectively worse performance*. Furthermore, this translation mediated a drop in the importance placed on the domain among women and Blacks. This research extends the literature on gender- and race-based reactions to feedback by noting the importance of the immediate interpretation of the feedback received.

When we receive feedback from others about a performance or work product, it often takes the form of subjective language—"great job" or "not bad" or "that was awful!" The question in this research is how we interpret or *translate* this feedback, and whether the social groups we belong to affect this translation.

We suggest that when feedback is provided in a domain where stereotypes about one's group exist, translation is made with reference to those stereotyped expectations. For example, in the domain of leadership, stereotypes about women's lesser competence relative to men may lead women and men to interpret identical feedback differently: Women may assume "good" indicates an objectively worse performance than do men, because that evaluation references a lower standard. Similarly, in the domain of academics, stereotypes about Blacks' lesser academic competence relative to Whites may mean that Black students interpret "good" to mean a worse objective performance than do Whites.

These predictions are derived from the shifting standards model, which suggests that we judge individual members of stereotyped groups on stereotype relevant dimensions relative to *within-category* standards or expectations (Biernat, Manis, & Nelson, 1991). Such standard shifts are possible because of the subjectivity of language. Adjectives such as "smart" or "emotional" have no fixed meaning (Huttenlocher & Higgins, 1971); thus, we might label others equivalently despite having very different underlying representations of them. A man and a woman might both be described as "aggressive" but that adjective may mean something different for each target. A key finding in shifting standards research is that *subjective judgments* may reveal contrast effects (e.g., a woman judged a better chief of staff than a man) at the same time that *common rule* or *objective judgments* reveal assimilation to stereotypes (a man judged to score objectively higher on job skill tests than a woman; Biernat & Kobrynowicz, 1997). Subjective judgments are commonly assessed in stereotyping research and in the real world (e.g., in performance appraisals; Murphy & Cleveland, 1995). Research on shifting standards suggests that the objective meaning of these judgments may differ depending on who is being described.

We have some evidence that *perceivers* "translate" subjective descriptions with reference to targets' social category memberships and associated stereotypes. For example, men described as "tall" were estimated to be almost 8 in. taller than "tall" women (Roberts & Herman, 1986), men described as "good parents" were assumed to be *objectively less involved* in child care than a similarly described woman (Kobrynowicz & Biernat, 1997), and favorable letters of recommendation were judged to indicate *lower levels* of objective achievement for women than men, particularly when the letter writer was known to be sexist (Biernat & Eidelman, 2007). Although perceivers' judgments of others are important, how *recipients* construe and respond to subjective feedback—particularly negative feedback—is important as well. This is the focus of the research reported here, in which we examine whether subjective feedback in a stereotype-relevant domain is translated differently depending on gender (Study 1) and race (Study 2), and whether other downstream consequences of such translation differences accrue.

A considerable body of research has addressed how feedback affects the self-esteem of members of negatively stereotyped groups. For example, in a study comparing Black and White students'

☆ This research was supported by a grant from the National Science Foundation to Monica Biernat (NSF 51720).

* Corresponding author at: Department of Psychology, University of Kansas, Lawrence, KS 66045-7556, USA. Fax: +1 785 864 5696.
E-mail address: biernat@ku.edu (M. Biernat).

0022-1031/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2011.08.010

reactions to failure or success feedback from a White evaluator, Major, Spencer, Schmader, Wolfe, and Crocker (1998) found that objective feedback had no effect on Black students' self-esteem but did affect the self-esteem of Whites. Another study found that Black students' self-esteem was harmed by negative feedback from a Black evaluator but not a White evaluator, largely because they questioned the White evaluator's objectivity (Banks, Stitt, Curtis, & McQuater, 1977; see also Coleman, Jussim, & Isaac, 1991). Women's self-evaluations also tend to be more negatively affected by negative feedback than men's (Roberts, 1991; Roberts & Nolen-Hoeksema, 1989).

Others have highlighted that the feedback from outgroup members produces attributional ambiguity, as one cannot be sure whether the feedback reflects the merits of one's work product or discrimination (e.g. Cohen & Steele, 2002; Cohen, Steele, & Ross, 1999; Crocker, Voelkl, Testa, & Major, 1991; Hoyt, Aguilar, Kaiser, Blascovich, & Lee, 2007; Major & O'Brien, 2005). Minority group members in particular may be more vigilant than others to contextual indications that their social group is devalued (Purdie-Vaughns, Steele, Davies, Diltman, & Crosby, 2008). For example, Blacks and Latinos were more likely to discount feedback (both negative and positive) from Whites (e.g., Crocker et al., 1991; Hoyt et al., 2007), and in a recent study, Blacks who were praised by a White female evaluator viewed her as less polite, and tended to assume she had lower expectations of them, than did Whites who received such praise (Lawrence, Crocker, & Blanton, 2011). Assurances from evaluators that high standards are being used may reduce the tendency to discount feedback (Cohen et al., 1999).

But none of this earlier research has addressed the direct *translation* or understanding of feedback received, the focus of this research. Our main prediction is that members of groups stereotyped as deficient in a performance domain (e.g., women in leadership, Study 1; Blacks in academic writing, Study 2) will interpret subjective feedback about their performance to indicate an objectively *less good* performance than their positively stereotyped counterparts (e.g., men, Whites). In Study 2, we include a no feedback control condition to document that feedback is necessary to produce differential translation; that the effect is based on more than Black participants' expectations of lower scores.

We also examine potential downstream consequences of feedback and its translation, including devaluing of the performance domain (Crocker & Major, 1989; Eccleston, Smyth, & Lopoo, 2010; Ogbu, 1991; Schmader & Major, 1999), as well as decrements in subsequent performance. We expect that women and men in Study 1 will enter the lab placing equal importance on leadership skills, and that Black and White participants in Study 2 will initially place equal importance on writing ability, the domains in which we offer feedback. But to the extent that women and Black students interpret the feedback they receive to be more objectively negative than do men and Whites, domain importance and subsequent motivation to perform well may be reduced. By integrating the operation of shifting standards with other perspectives on the consequences of feedback, we hope to offer a better understanding of how and why feedback settings may be fraught for members of negatively stereotyped groups.

Study 1

Method

Participants and procedure

Participants were 46 White University of Kansas (KU) undergraduates (26 women, 20 men) who participated in a study on "leadership ability" in exchange for course credit. Each participant was run individually, and all instructions were delivered via computers that were supposedly networked to other labs on campus. Participants were asked to role play being the President of a student organization on campus (KU Lead), whose job was to begin preparing for the upcoming semester by sending an email to all members. Participants were told that this was an opportunity to show their leadership skills, and that another participant "will be reading your message from the point of view of a member of the KU Lead organization."

Participants next completed a measure of perceived importance of leadership, then composed an email (which conveyed their name and gender) and "sent" it. The program simulated wait time as the "other participant," whose gender was unspecified, supposedly read the email and prepared an evaluation. After several minutes, feedback appeared in what looked like a returned email message. The feedback was constant across all participants and was designed to be subjective, and moderately negative, in tone:

> "Hey, nice try, [participant's name], but I'm not for sure your e-mail was the best. If I werea member of KU Lead and receieved your e-mail, I would think your leadership skills are not good (but not super bad either). Your email mesasge was not the best or the worst email I've seen, but it definetely was weak in authority . . . While the writing was ok I'm not convincd that your leadership skills are "all there."KULead Member" *(misspellings intended)*.

Participants then "translated" this feedback into objective scores, answered several other questions, and completed a (supposed) test of leadership ability, as described below. All participants were probed for suspicion and carefully debriefed and reassured about the bogus nature of the feedback prior to dismissal.

Dependent measures

Leadership importance was assessed with 3-items: "To what extent is being a good leader important to you?," "To what extent is the current leadership e-mail writing task important to you?," and "To what extent are your leadership skills for writing the current email important to you?," all answered on 1 (*not at all important*) to 7 (*very important*) scales. Participants completed this measure before they wrote their emails and again near the end of the study (αs = .83 and .77 at Times 1 and 2, respectively).[1]

The key dependent variable, objective translation of the feedback, was assessed with two items: "based on the feedback you received . . . how much money would the evaluator award you for your leadership ability (offered options = $25, $50, $75, $100, or $125) and ". . . what letter grade (A, B, C, D, F) in leadership do you think the group member would have given you?" Both items were treated as 1–5 scales and were averaged to produce the objective translation index (α = .61).[2]

As an indicator of perceived *subjective* positivity, participants were asked to what extent "the feedback I received was positive" (1 = *strongly disagree* to 7 = *strongly agree*). To assess perceived standards, participants indicated their agreement with two statements: "I think my 'group member' held me to low/high standards for leadership aptitude/ability" (1 = *strongly disagree* to 7 = *strongly agree*). The "low standards" item was reverse scored and an average calculated (α = .48).

Finally, participants took a bogus 9-item "Leadership Skills Test" that included six math/logic items from the GMAT interspersed with "management reaction" questions we created to further the cover story that the test measured leadership ability. Only the 6 GMAT

[1] In both studies, participants also completed a measure of mood (the PANAS; Watson, Clark, & Tellegen, 1988), and Heatherton and Polivy's (1991) social and performance state self-esteem scales. We found no effects of participant sex in Study 1, or of participant race, feedback, or their interaction in Study 2 on these measures, on post-test measures of the same constructs, or on change over time, and therefore these variables will not be further discussed.

[2] Although instructors may recognize the subjectivity of grade assignment, grades nonetheless offer a *common metric* of judgment, in that an A is an A regardless of who receives it (see Biernat & Kobrynowicz, 1997).

items could actually be scored for accuracy; the total number of correct out of 6 served as an indicator of performance.

Results

Participants rated the positivity of the feedback at 3.24 (on a 1–7 scale), and there was no sex difference, $t<1$, $p>.90$. Thus, men and women perceived the feedback as equally *subjectively* negative. Nonetheless, as predicted, they differentially translated this feedback into *objective*, *common rule* units: Women assumed the feedback indicated lower objective standing ($M=1.96$, $SD=.96$) than did men ($M=2.58$, $SD=.92$), $t(44)=2.19$, $p<.05$. Also consistent with predictions, women ($M=3.98$, $SD=1.32$) assumed they were held to lower standards than did men ($M=4.90$, $SD=.97$), $t(44)=2.62$, $p<.05$.

To assess downstream consequences of the translation process, we examined change in perceived importance of leadership ability from the beginning to end of the study. A Participant Sex×Time of measurement ANOVA indicated significant main effects of time, $F(1,44)=9.26$ as well as the Sex×Time interaction, $F(1,44)=4.00$, $ps=.05$. The effect of sex was not significant at either point in time ($ps>.55$), but women showed a significant drop in leadership importance from the beginning ($M=5.11$, $SD=1.12$) to end of the study, ($M=4.68$, $SD=1.11$) $p<.01$, whereas men did not ($M_{T1}=4.90$, $SD=1.22$; $M_{T2}=4.81$, $SD=1.33$), $t<1$.

Performance on the "leadership test" was also worse for women ($M=3.62$, $SD=1.33$) than men ($M=4.60$, $SD=1.19$), $t(44)=2.61$, $p<.05$.

Did objective translations matter for the change in importance of leadership or for performance? To the extent that participants translated the feedback negatively, leadership importance dropped from the beginning to the end of the study ($r=.43$, $N=46$, $p<.01$), though test performance was unaffected ($r=.10$, *ns*). We next examined whether objective translations mediated the effect of participant sex on changes in leadership importance over time. We used Preacher and Hayes (2008) bootstrapping method to test the meditational model depicted in Fig. 1, in which post-feedback importance was the dependent variable, and pre-feedback importance was a covariate. This method computes 5000 bootstrapped samples to estimate bias corrected confidence intervals of mediated effects; in this case, the effect of sex on change in leadership importance via objective translations of feedback. Results indicated significant mediation; mediated effect = .14, $SE=.10$, 95% CI = .01–.47.

Discussion

Women and men in Study 1 entered our lab placing equally strong importance on leadership, and after receiving negative feedback, they judged it to be equally subjectively negative. But consistent with predictions, women "translated" this feedback to indicate a worse objective performance than did men; they also assumed they were held to lower standards than did men. These findings suggest that members of groups stereotyped as deficient on a performance dimension may translate feedback with the stereotype in mind. "Bad" at leadership is perceived to be worse if one is a woman than a man. Additionally, the importance placed on leadership declined over the course of the study in women, but not men, and this effect was mediated by objective translations. We therefore suggest that translations are an important piece of the puzzle of how feedback matters for negatively stereotyped group members.

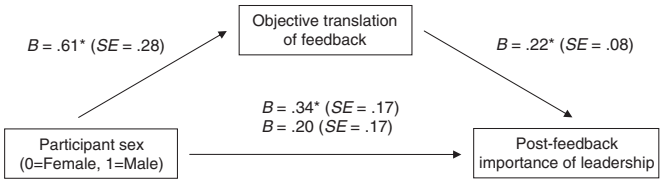

Fig. 1. Mediational model: Objective translation mediates the effect of participant sex on post-feedback importance of leadership, Study 1. Notes: All coefficients are unstandardized regression coefficients, with pre-feedback writing importance controlled. *$p<.05$.

An important caveat regarding these findings is in order. Study 1 used a non-experimental design; all participants received the same subjective feedback and thus we cannot be sure whether the participant sex effects were triggered by the feedback per se. Our performance findings in particular are problematic; perhaps women and men would have performed differently in the absence of any feedback or its translation. Of course the fact that we measured change in importance over time argues against any basic sex difference, as does the finding that only objective translations, but not subjective views of the positivity of the feedback differed by sex. Nonetheless, to address this issue, Study 2 includes a comparison between a negative feedback and no feedback condition, and includes a pre- and post-feedback measure of performance.

Additionally, Study 2 focused on a different social category (race) and a different stereotyped domain (academic writing) to address the same key prediction: That negatively stereotyped group members (African Americans in this case) will translate subjective feedback to indicate a worse objective performance than will more positively stereotyped targets (White Americans).

Study 2

Method

Participants

Participants were 26 Black (31% male) and 33 White (36% male) KU undergraduates. Forty-six participants were paid $15 for their participation, and the remaining 13 were recruited through the departmental participant pool. Of the Black participants, 73.1% were paid; of the White participants, 81.8% were paid, $\chi^2(1, N=59)<1$, *ns*. Paid versus unpaid ratios were also identical across feedback condition (79.3% paid in the negative feedback condition and 76.7% paid in the no feedback condition, $\chi^2(1, N=59)<1$, *ns*).

Procedure and materials

The study used a between-subjects Feedback (negative, none)× Participant Race (Black, White) design. All participants were run individually by a White female experimenter, who told participants that the University's administration had been thinking about instantiating a comprehensive exam for graduating seniors and that their task was to write an essay expressing their opinion on the issue. Participants were also either told the essay would be evaluated by the experimenter, or were given no information on this point. Those in the negative feedback conditions were also shown a blank evaluation sheet prior to writing their essay, which indicated seven evaluative criteria, including clarity of thought, critical thinking, etc. A 7-point scale ranging from 1 (*inadequate*) to 7 (*outstanding*) accompanied each item.

After learning about the writing task, participants completed a demographics form along with a 3-item measure of writing importance, similar to the importance items used in Study 1: "To what extent is having good writing skills important to you?," "To what extent is the current essay writing task important to you?," and "To what extent are your writing skills for this essay task important to you?" (1 = *not at all important* to 7 = *very important*) scales ($\alpha=.67$).

The researcher then left the room while participants composed their essays on a laptop computer. After 15 min, the researcher retrieved the computer and left to ostensibly read and evaluate the essay (those in the no feedback condition were told that the researcher needed to "take care of something"). After 4 min, she returned with a completed feedback sheet for participants in the negative feedback

condition; the sheet indicated that the participants received two 2s, three 3s, one 4, and a 3 on "overall assessment" on the essay.

All participants then completed measures of objective feedback translation, subjective positivity, perceived standards, and a second measure of writing importance. Those in the no feedback condition were asked to answer the translation questions "as if the researchers had evaluated your essay."

To assess objective translation of the feedback they received (or imagined they had received), participants were asked "what letter grade (A+ to F) do you think the researcher would have given you for your essay?" and "what percent score do you think the researcher would have given you for your essay (0 to 100%)?" Letter grade was converted to a continuous numeric scale (F=0, A+=12), then the two items were standardized and averaged ($\alpha = .92$).

Participants were also asked "How negative or positive was the assessment you received (or would have received) from the researcher?" (1=*very negative* to 7=*very positive*), and they indicated their agreement with the statement "the researcher held you [or would hold you] to high standards for writing skill." All participants then answered the 3 writing importance items again ($\alpha = .71$).

Finally, participants were given 10 min to compose another opinion essay on the topic of "a possible tuition increase at the university." Afterwards, they were carefully screened for suspicion and extensively debriefed about the purpose of the study and the nature of the false feedback. Two judges, blind to race and feedback condition, later independently coded all essays for quality, using a 1–5 rating scale. The correlation between raters was modest, $r = .51$; the two raters' judgments were averaged to create an index of writing quality for each essay.

Results and discussion

All dependent variables were analyzed using Feedback (negative, none)×Participant Race (Black, White)×Participant Type (paid, unpaid) analyses of variance (ANOVA). There were occasional significant effects involving participant type (as well as participant sex), but because these were never theoretically interesting and never qualified the results reported below, they will not be discussed further.

Understanding of feedback

Participants in the negative feedback condition perceived the assessment as significantly more negative ($M = 2.48$, $SD = .83$) than those in the no feedback condition ($M = 5.22$, $SD = .85$), $F(1,51) = 78.39$, $p < .0001$. No other effects were significant, $ps > .16$. Thus, as in Study 1, there were no effects of social category membership on *subjective* perceptions of feedback valence.

But as can be seen in Fig. 2, *objective* translations varied by race and condition, as predicted: The main effects of feedback, $F(1,51) = 48.40$, $p < .0001$, and participant race, $F(1,51) = 8.49$, $p < .01$, were significant, as was the Feedback×Participant Race interaction, $F(1,51) = 9.44$, $p < .01$. Simple effects tests indicated that Black participants translated the feedback more negatively than did Whites only following negative feedback, $p < .001$; the race difference was nonsignificant in the no feedback condition, $p > .90$. These data demonstrate that the actual receipt of feedback was important to produce a race-based difference in objective translation. That is, Black students (in the absence of feedback) did not merely assume that they would receive worse grades than did Whites.

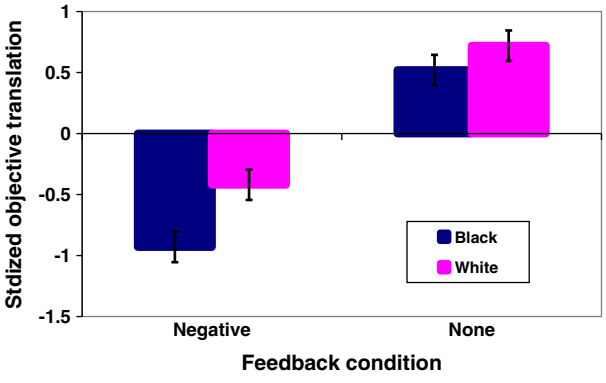

Fig. 2. Objective translations of feedback, by race and feedback, Study 2.

Analysis of the single item assessing standards indicated a main effect of race, $F(1,52) = 7.50$, $p < .01$: Black participants ($M = 3.96$, $SD = 1.54$) assumed they were held (or would be held) to lower standards than did Whites ($M = 4.88$, $SD = 1.11$).

Writing importance and performance

Mean levels of writing importance by time, race, and condition appear in Fig. 3. A mixed-design ANOVA with time of measurement as a repeated factor indicated a main effect of time, $F(1,51) = 20.40$, $p < .0001$, interactions between Time and Feedback, $F(1,51) = 23.27$, $p < .0001$, and Time and Race, $F(1,51) = 18.73$, $p < .0001$, and a Time×Feedback×Race interaction, $F(1,51) = 4.77$, $p < .05$. The importance placed on writing significantly dropped over time for Blacks in the negative feedback condition, $p < .01$, but not for Blacks in the no feedback condition, $p > .30$. Among Whites, importance significantly *increased* over time in the no feedback condition, $p < .05$, and marginally significantly dropped in the negative feedback condition, $p < .07$. The drop in importance was significantly greater for Blacks than Whites in the negative feedback condition, $p < .001$, whereas change in importance was not reliably different for Blacks and Whites in the no feedback condition, $p > .13$.

Writing quality was similarly analyzed, with essay performance (essay 1, essay 2) as a repeated factor. The main effect of race, $F(1,51) = 12.40$, $p < .0001$,[3] was qualified by the 3-way interaction, $F(1,51) = 4.07$, $p < .05$. As Fig. 4 indicates, performance significantly dropped from pre- to post-feedback for Blacks in the negative feedback condition, $p = .05$; in each of the other conditions, there was no change, $ps > .25$. Furthermore, this drop in performance from the first to second essay was significantly greater for Blacks than Whites in the negative feedback condition, and for Blacks in the negative compared to no feedback condition, $ps < .05$. Change scores did not differ for Blacks and Whites following no feedback, or for Whites in the negative versus no feedback condition, $ps > .50$.

In short, Blacks who received negative feedback were uniquely negatively affected: They suffered from drops in the perceived importance in writing, and in writing performance, from pre- to post-feedback.

Correlations and mediation

We next examined whether objective translation predicted the change in importance and performance over time. As in Study 1, objective translations were uncorrelated with change in essay quality over time ($r = .10$ overall, $N = 59$, and *ns* within each race/feedback condition). But also as in Study 1, translation predicted change in writing importance in the negative feedback condition ($r = .65$, $n = 29$, $p < .0001$), though not the no feedback condition ($r = .02$, $n = 30$, *ns*). To the extent that subjective feedback was translated negatively, the importance of writing fell from the beginning to end of the study.

Given this correlation, we again used Preacher and Hayes (2008) bootstrapping method to test the meditational model (Fig. 5). As in Study 1, objective translations of negative feedback mediated the effect of race on change in domain importance, mediated effect = .37, $SE = .15$, 95% CI = .16–1.36.

[3] Given the main effect of race on essay quality, we examined whether any of the previously reported results were altered by controlling for quality of writing on essay 1. They were not; all significant effects remained so with performance quality controlled.

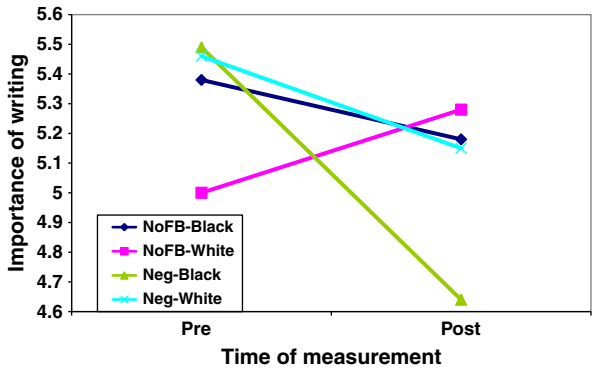

Fig. 3. Importance placed on writing skill by race, feedback condition, and time of measurement, Study 2.

General discussion

These studies were designed to examine whether members of negatively stereotyped groups, when exposed to subjective feedback in the stereotyped domain, translate that feedback to indicate a *lower* objective level of performance than do members of more positively stereotyped groups. We predicted that given gender stereotypes about leadership (Study 1), and racial stereotypes about academic performance (Study 2), female and Black students would translate the feedback to indicate an objectively *worse* performance (than men and Whites, respectively). This is precisely what we found.

Importantly, the race difference in translation in Study 2 did not emerge in the no feedback condition, in which participants were simply asked to indicate what objective score they "would have" received from an evaluator. Thus, the effect was not based on heuristic assumptions that evaluators would give lower scores to Blacks than Whites; instead, negative subjective feedback had to be received and translated. Furthermore, in both studies, *subjective* perceptions of feedback valence were unaffected by sex or race. Women and men in Study 1, and Blacks and Whites in Study 2, judged the negative feedback to be equally subjectively negative. Instead, it was the translation of negative subjective language into objective performance ratings that provided evidence for gender- and race-based shifting standards.

In both studies, we also found that participants made assumptions about the standards to which they were held that were consistent with race and gender stereotypes: Women assumed that they were held to lower leadership standards than did men, and Blacks assumed they were held to lower academic writing standards than did Whites. However, in data not reported earlier, we found that perceived standards were only modestly related to objective translation (in Study 1, $r = .23$, and in Study 2, $r = .19$, *ns*), and thus we have no evidence that

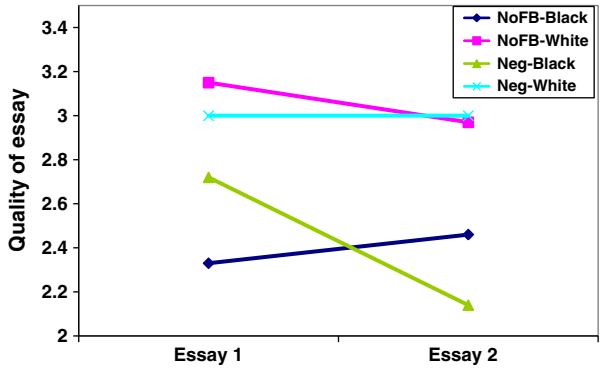

Fig. 4. Quality of essays, by race, condition, and time of measurement, Study 2.

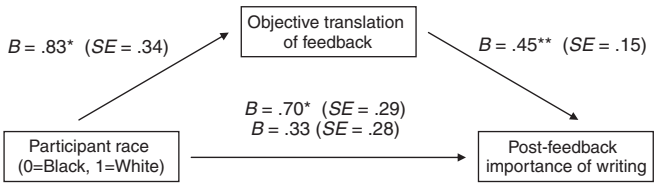

Fig. 5. Mediational model: Objective translation mediates the effect of race on post-feedback importance of writing, negative feedback condition, Study 2. Notes: All coefficients are unstandardized regression coefficients, with pre-feedback writing importance and participant type (paid, unpaid) as controls. $^{*}p<.05$, $^{**}p<.01$.

assumed standards played a direct role in producing translations. Thus, we cannot definitively report a story of the *process* by which our key outcome—lower objective translations by negatively stereotyped group members—emerged in the negative feedback condition.

However, we did document some important downstream consequences of feedback and its translation: Following negative feedback, women in Study 1 and Black participants in Study 2 uniquely showed a) reductions in the importance they placed on the performance domain and b) lower quality of a post-feedback performance (for Blacks in Study 2, a significant drop in writing quality from the beginning to end of the study session). Participants began our studies with equal investment in the performance domain. But following negative feedback, women and Blacks showed evidence of domain disidentification (Steele, 1997). We further documented that objective translations—how participants interpreted the negative feedback they received—played a mediating role in this process. That is, female and Black participants translated the negative feedback to indicate a worse objective score, which, in turn, predicted a drop in domain importance over time. Performance quality also dropped over time in these groups, though objective translations did not mediate this effect.

We suggest that the consequences of receiving negative feedback and interpreting it negatively were motivational in nature—female and Black participants were less motivated to invest in the performance domain following the process of translating negative feedback. Still, additional research will be necessary to determine whether this reduction in motivation was due to internalization of the objective evaluation of their work (e.g., "I believe I deserved a D") or to "dissing" of the evaluator ("I won't work hard for her"). Based on data not reported here, we do know that likeability of the evaluator was unaffected by sex or race, but further investigation of this issue is needed.

In both studies, the feedback to which we exposed participants was moderately negative, and thus we cannot say whether the differential translation of feedback occurs regardless of its valence. The shifting standards perspective predicts a main effect of social category membership (e.g., given stereotypes about academic performance, Black students may always translate feedback with reference to lower standards). But we recognize that positive feedback may be accepted as consistent with self-perceived performance, and therefore may reduce the tendency to refer to group-based standards. Further research is necessary to examine how feedback valence affects the translation of feedback, especially given that even positive feedback may produce attributional ambiguity for the negatively stereotyped (Cohen et al., 1999; Crocker et al., 1991; Major & Sawyer, 2009).

Our findings are also limited by the fact that our evaluators were either of unspecified gender and race (Study 1) or were White females (Study 2). Thus, we cannot speak to whether and how evaluator race or gender is figured into translations. From the shifting standards perspective, we might predict that to the extent that stereotypes are culturally shared, expectations about groups and the use of within-category standards are applied regardless of evaluator category. But an attributional ambiguity account would make different predictions, in that feedback from an outgroup member might be particularly suspect (see Major & Sawyer, 2009; Mendes, Major,

McCoy, & Blascovich, 2008). Research assessing the translation of feedback from both ingroup and outgroup members is needed.

This research extends the literature on shifting standards by documenting that gender and race stereotypes affect interpretation of negative feedback about the *self*, and the literature on reactions to feedback by highlighting the importance of the *immediate understanding* or *interpretation* of the feedback received. Our analysis is consistent with the view that encoding processes play an important role in stereotyping effects (von Hippel, Sekaquaptewa, & Vargas, 1995). If members of negatively stereotyped groups encode subjective feedback more negatively than do positively stereotyped group members, this may set in motion a process whereby stereotypes are confirmed and maintained. We hope that this paper triggers additional work on the complex processes by which gender- and race-based stereotypes affect feedback, its interpretation, and its consequences.

References

Banks, W. C., Stitt, K. R., Curtis, H. A., & McQuater, G. V. (1977). Perceived objectivity and the effects of evaluative reinforcement upon compliance and self-evaluation in Blacks. *Journal of Experimental Social Psychology*, *13*, 452–463.

Biernat, M., & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology*, *37*, 1149–1175.

Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*, 544–557.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*, 485–499.

Cohen, G. L., & Steele, C. M. (2002). A barrier of mistrust: How negative stereotypes affect crossrace mentoring. In J. Aaronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 303–327). San Diego, CA: Academic Press.

Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, *25*, 1302–1318.

Coleman, L. M., Jussim, L., & Isaac, J. L. (1991). Black students' reactions to feedback conveyed by White and Black teachers. *Journal of Applied Social Psychology*, *21*, 460–481.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*, 608–630.

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*, 218–228.

Eccleston, C. P., Smyth, J. M., & Lopoo, L. M. (2010). Unraveling the race paradox of achievement and self-views. *Social Psychology of Education*, *13*, 1–18.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality & Social Psychology*, *60*, 895–910.

Hoyt, C. L., Aguilar, L., Kaiser, C. R., Blascovich, J., & Lee, K. (2007). The self-protective and undermining effects of attributional ambiguity. *Journal of Experimental Social Psychology*, *43*, 884–893.

Huttenlocher, J., & Higgins, E. T. (1971). Adjectives, comparatives, and syllogisms. *Psychological Review*, *78*, 487–504.

Kobrynowicz, D., & Biernat, M. (1997). Decoding subjective evaluations: How stereotypes provide shifting standards. *Journal of Experimental Social Psychology*, *33*, 579–601.

Lawrence, J. S., Crocker, J., & Blanton, H. (2011). Stigmatized and dominant cultural groups differently interpret positive feedback. *Journal of Cross-Cultural Psychology*, *42*, 165–169.

Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. *Annual Review of Psychology*, *56*, 393–421.

Major, B., & Sawyer, P. J. (2009). Attributions to discrimination: Antecedents and consequences. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 89–110). New York: Psychology Press.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin*, *24*, 34–50.

Mendes, W. B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology*, *94*, 278–291.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage Publications.

Ogbu, J. U. (1991). Minority coping responses and school experience. *Journal of Psychohistory*, *18*(4), 433–456.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Diltman, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*, 615–630.

Roberts, T. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin*, *109*, 297–308.

Roberts, J. V., & Herman, C. P. (1986). The psychology of height: An empirical review. In C. P. Herman, M. P. Zanna, & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario symposium, Vol. 3*. (pp. 113–140)Hillsdale, NJ: Erlbaum.

Roberts, T., & Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles*, *21*, 725–747.

Schmader, T., & Major, B. (1999). The impact of ingroup vs. outgroup performance on personal values. *Journal of Experimental Social Psychology*, *35*, 47–67.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. *Advances in Experimental Social Psychology*, *27*, 177–254.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality & Social Psychology*, *54*, 1063–1070.