Brief report

The positive feedback bias as a response to self-image threat

Kent D. Harber1*, Reshma Stafford2 and Kathleen A. Kennedy3
1Rutgers University, Newark, New Jersey, USA
2Rutgers University, New Brunswick, New Jersey, USA
3Princeton University, Princeton, New Jersey, USA

This research examined whether Whites favourably bias their feedback to minorities in order to see themselves as egalitarian. White teacher trainees first had their egalitarian self-images affirmed, left unchanged, or threatened. Then they provided feedback on a poorly written essay supposedly authored by either a Black or a White student. As predicted, trainees in the Black writer/self-image threat condition selectively rated essay content more favourably, recommended less time for skill development, provided more favourable copy-editing comments, and generated more equivocating 'buffers'. In contrast, trainees in the Black writer/self-image boost condition supplied feedback indistinguishable from feedback provided by trainees in the White writer conditions, which was unaffected by the self-image conditions. The implications for minority education and intergroup communication are discussed.

Feedback from Whites to Blacks can be positively biased, such that Whites sometimes give more praise and less criticism to Blacks than to Whites for work of equal merit. This positive feedback bias occurs in both written and face-to-face feedback, when feedback is delivered anonymously as well as publicly, and when recipients are peers or students (Harber, 1998, 2004; Harber, Gorman, & Gengaro, 2009).

Understanding the cause of the feedback bias addresses an issue gaining increased attention in intergroup relations research. This is the disruptive anxiety that Whites often experience when interacting with minorities (Richeson & Shelton, 2007; Stephan & Stephan, 1985), and the heightened out-group deference (i.e. exaggerated friendliness, agreement, and solicitousness) that often follows. For example, Whites display a classic physiological threat response to Black but not to White interaction partners, but then report greater levels of liking for and agreement with Black partners than White partners (Mendes, Blascovich, Lickel, & Hunter, 2002). Also, the more

*Correspondence should be addressed to Dr Kent D. Harber, Department of Psychology, Rutgers University, Smith Hall, 101 Warren Street, Newark, NJ 07102, USA (e-mail: kharber@psychology.rutgers.edu).

DOI: 10.1348/014466609X473956
emotional discomfort Whites experience with minorities, the more positively they rate minorities (Littleford, Wright, & Sayoc-Parial, 2005).

There are differing explanations for the link between intergroup anxiety and out-group deference. One is that it reflects Whites' concerns that they will be perceived as bigoted by minorities or others (Littleford et al., 2005; Stephan & Stephan, 1985; Vorauer & Kumhyr, 2001). In this case, out-group deference advertises egalitarianism outwardly, providing a shield from external social censure (Hastorf, Northcraft, & Picciotto, 1979).

Another explanation is that intergroup anxiety is solipsistic – that it arises from Whites' concerns about seeing themselves as racist. In this case, out-group deference serves to display egalitarian values inwardly to oneself (Dutton & Lake, 1973; Gaertner & Dovidio, 1986). If self-image maintenance motives (e.g. Tesser, 1988) motivate out-group deference, then Whites' interactions with minorities will be affected by how Whites regard themselves – even when there is no danger of external social censure.

The self-image maintenance and social desirability accounts of out-group deference have different implications for intergroup contact. When anxiety arises from the anticipated reactions of others, both Whites and minorities may have more direct control of how interactions unfold (e.g. by testing and confirming the limits of candour). But when Whites use out-group deference to allay their own self-image concerns, minorities are confronted with behaviour that can seem perplexing and even demoralizing. For example, receiving positive feedback from a White individual whose true motives are unclear can depress Blacks' self-esteem (Crocker, Voelkl, Testa, & Major, 1991).

However, these implications are relevant only if self-image maintenance motives actually induce out-group deference. Most studies on self-image threat do not directly relate self-image concerns to intergroup communication, but instead to vigilance, physiological arousal, social withdrawal, guilt, disrupted cognition, and other reactions (see Crocker & Garcia, 2009, for a review). In the few studies that tested the connection between self-image concerns and deferential communication to out-groups (e.g. Dutton & Lake, 1973), the communications were not anonymous and may, therefore, have reflected external pressures rather than internal motives.

**Direct test of self-image concerns as moderators of the positive feedback bias**

To demonstrate that self-image motives cause out-group deference, it is necessary that the audience to which the Whites play is solely, exclusively themselves. The present research attempted to produce these conditions within the context of instructional feedback. White teacher trainees first displayed – to only themselves – a sympathetic, irrelevant, or hostile attitude towards minorities. Trainees then gave anonymous feedback, they believed, to either a Black or a White student. If self-image motives drive the positive bias, then feedback to a Black student should be biased after egalitarianism was self-impugned, but not after egalitarianism was self-affirmed.

The study also used a design that vigorously reduced social desirability concerns and left self-judgments as the principle motivator for the positive bias. It did so by causing subjects to believe that the pro-, neutral-, or anti-minority attitudes they were induced to express were anonymous, that their feedback was anonymous, and that only they, the participants, could connect their minority-related attitudes to their subsequent feedback.
Do teacher trainees demonstrate the positive feedback bias?

This study recruited students from the Rutgers University School of Education, a leading teacher-training programme. These trainees often get jobs at schools serving minority students. If the teacher trainees show the bias, and do so in order to protect their private self-images, then this would have important ramifications for minority students and for teacher training.

Method

Participants

One hundred and eight White teacher-training students (83.2% female, age \(M = 21.32,\ SD = 4.68\)) participated in this study. Nearly half of the trainees (44.4%) had previous teaching experience, averaging about 2.2 years. Teacher training students (“Trainees”), completed the study individually in 1-hour sessions and received $15.00.

Procedure

Trainees were told that undergraduates in a writing workshop had volunteered writing samples to get trainees’ feedback. Trainees received a poorly written two-page essay, supposedly drafted by their assigned student but actually prepared by experimental staff. Trainees were told that their job was to make comments on the essay draft, which would be returned to the writer.

Egalitarianism manipulation

After delivering the cover story, the experimenter rifled through his/her clipboard, looked chagrined, and asked the trainee if he/she had completed the Social Issues Survey. After the trainee responded negatively, the experimenter produced a one-page Social Issues Survey. The experimenter explained that the survey, which was being developed by an unrelated research team, should have been completed upon arrival. The trainee was asked if he/she would be willing to complete it now. The trainee was told not to write his/her name on the survey, and to deposit the completed survey through a slot in a locked collection box that the other researchers would retrieve later. Thus, trainees thought they were completing a measure that was totally anonymous, unrelated to the feedback study, and inaccessible to the feedback study staff.

The Social Issues Survey manipulated trainees’ concerns about their egalitarianism, and partially followed a design employed by Monin and Miller (2001). There were three versions of this survey, all containing nine opinion-rating questions. The egalitarian boost version, designed to reinforce pro-minority views, included items such as ‘Government offices should be closed on Martin Luther King Day’ and ‘The Confederate flag should not fly over government buildings’. The egalitarian threat survey, designed to express opinions unfavourable to minorities, included items such as ‘It should be legal for businesses to be open on Martin Luther King Day’, and ‘People should be allowed to fly the Confederate flags on their own front lawns’. The civil libertarian framing of these questions (i.e. ‘should be legal’, ‘should be allowed’) pushed the egalitarian threat trainees to endorse items that might be palatable individually but collectively had an anti-minority thrust. The neutral condition survey focused on shopping. It was irrelevant to minorities and neither reinforced nor threatened egalitarian self-image. Sample items include, ‘My supermarket sells a wide variety of canned goods’ and ‘I do most of my clothes shopping at the local mall’.
There were only four response options for the survey questions: ‘totally disagree’, ‘mainly disagree’, ‘mainly agree’, and ‘totally agree’. Because there was no neutral option, trainees in the threat condition could not easily avoid endorsing the pro-libertarian but anti-minority items comprising their survey.

An additional question asked trainees to list, from categories that varied by condition, five exemplars. In the egalitarian boost condition, this item read ‘Please list five famous African Americans in music, poetry, literature, government, or sports’. Trainees were expected to easily identify five African Americans in these categories and infer from their ease of recall (per Winkielman & Schwarz, 2001) their appreciation of minority achievement. Threat condition trainees were asked to list five famous African Americans in the visual arts, sociology, psychology, life sciences, physical sciences, or math. Most people would be pressed to name exemplars of any ethnic group within these fields. However, difficulty naming five African Americans was expected to challenge egalitarianism. Trainees in the neutral condition listed five places where they shop.

Essay review
After completing the Social Issues Survey, trainees engaged in the essay feedback task. Two essays (television violence, environmentalism) were developed for this research to control for essay topic artefacts. The essays, filled with mechanical and content errors, were of equal length and style. Trainees received one of these two essays, which were distributed in a counterbalanced manner. Trainees were told to copy-edit and write comments directly on the essay.

Race cue
Before reviewing their assigned essays, trainees examined a brief writers’ description sheet, supposedly completed by the essay writer. An item towards the bottom of the sheet asked ‘What campus organizations have you joined’. In the Black writer condition, this was answered ‘African Heritage Month Organizing Committee’, while in the White writer condition this item was left blank. The writer’s name (‘Natisha K’/’Heidi K’) served as another race cue.

Writer rating form
After copy-editing their assigned essays, trainees were given the writer rating form, which was described as a direct, confidential, and anonymous communication between themselves and the writer. The writer rating form contained two items for the trainee to indicate how much additional work the essay required on mechanics and on content. An additional item asked trainees to recommend how many hours per week the writer should devote to developing writing skills. Trainees sealed their unsigned writer rating form in an envelope addressed to ‘Writer 26’.

Follow-up questionnaire
After completing the writer rating form, trainees responded to questions regarding the writer’s race and other general background questions. Trainees were then debriefed, thanked, and paid.
Selective bias for essay content
In accord with previous positive feedback bias studies (Harber, 1998, 2004), we predicted that the positive bias would be restricted to ratings of essay content (ideas, beliefs, reasoning) but not ratings of mechanics (grammar, spelling, punctuation). Content criticisms are subjective and less easily justified compared to mechanics (where there are objective standards), and are more interpersonally consequential (people typically value their ideas more than their spelling). These features make content criticism more subjectively risky, and therefore more susceptible to self-image concerns that are predicted to drive the bias.

Results
Preliminary analyses
Essay comments coding
Two raters, blind to condition, coded trainees’ copy-editing comments. The raters tallied positive and negative content comments (arguments, ideas, etc.), positive and negative mechanics comments (spelling, grammar, etc.), general positive comments, general negative comments, buffers, and ‘other’ comments. Buffers were comments that served to moderate or qualify criticism. The reliability of coding was generally satisfactory; content negative $r(106) = .88$, content positive $r(106) = .85$, mechanics negative $r(106) = .86$, and buffer $r(106) = .71$. Mechanics positive $r(106) = .49$, general positive $r(106) = .58$, general negative $r(106) = .31$, and ‘other’ $r(106) = .40$, had low reliability, likely due to their rarity (mean frequency ranged from 0.15 to 0.50).

Because there were many more mechanics related comments ($M = 47.89$, $SD = 18.02$) than content comments ($M = 11.15$, $SD = 7.45$), all comments were transformed into standardized scores. An overall index of feedback favourableness was computed by subtracting all the standardized negative comments from all the standardized positive comments.

Data reduction of Social Issues Survey
The nine items comprising each version of the Social Issues Survey were summed and averaged into single scale scores. Social Issues Survey scores indicated how well the egalitarian boost, threat, and control versions of the survey performed.

Essay topic
A series of three-way analyses of variance indicated that essay topic had no main effect on ratings of mechanics or content, copy-editing favourableness, or the proportion of buffering comments (all $ps > .05$). The one exception was the number of hours recommended for writing skills development, $F(1, 103) = 5.90, p = .017$. Essay topic did not interact with writer race, self-image threat condition, or these two IVs concurrently (all $ps > .05$). Because essay topic was otherwise unrelated to feedback, remaining analyses collapsed across this variable.

Experimenter
d
Experimenter demographics (Asian/Indian female, White female, White male) did not affect the writer rating form outcomes, all $ps > .05$. There was an effect for the
favourableness of copy-edited comments, \( F(2, 102) = 3.48, p = .035 \), due to greater favourability with the White female \((M = 1.94, SD = 3.09)\) than the Asian/Indian female \((M = -0.26, SD = 2.70)\), \( p = .029 \), or (marginally) the White male \((M = -0.18, SD = 2.12)\), \( p = .053 \). Further analyses collapsed across experimenter.

**Manipulation checks**

**Social Issues Survey**

Trainees differed in how enthusiastically they endorsed their respective surveys, \( F(1, 103) = 21.22, p < .001 \), and in their ability to recall exemplars \( F(2, 103) = 146.05, p < .001 \). Tukey post hoc tests showed that trainees in the egalitarian boost condition endorsed survey items more enthusiastically \((M = 3.43, SD = 0.37)\) than did trainees in the neutral condition \((M = 2.98, SD = 0.45)\), \( p < .001 \), or in the threatened condition \((M = 2.84, SD = 0.38)\), \( p < .001 \). Also, trainees in the boost condition \((M = 4.79, SD = 0.61)\) and control condition \((M = 4.95, SD = 0.31)\) recalled more exemplars than trainees in the threat condition \((M = 1.78, SD = 1.36)\), \( ps < .001 \). It is important to note that the threat condition survey (wherein participants were asked to endorse items that collectively had an anti-minority tone) received a mean score of nearly 3 (where 4 = maximum agreement), indicating that these noxious items were endorsed, but unenthusiastically so. Thus, the manipulation apparently operated as intended – it caused threat condition participants to reluctantly endorse items that cumulatively suggested an anti-minority sentiment.

**Identification of writer race**

Due to experimenter error, only 50 of the 108 trainees reported on their recollection of writer race. This was not considered a significant problem because prior feedback studies using this manipulation (Harber, 1998, 2004) obtained over 95% correct race identification. Three trainees who incorrectly identified writer race were excluded from further analyses.

**Suspicion**

No trainee identified the actual purpose of the study.

**Main analyses**

This research predicted that feedback favourableness would be jointly influenced by recipient race (Black or White writer) and trainees’ egalitarian self-image (boosted, unchanged, or threatened). This predicted pattern was tested using a planned contrast analysis (Rosenthal & Rosnow, 1985). The Black/boost, White/boost, White/neutral, and White/threat conditions, which were expected to supply similarly critical feedback, were assigned weights of \(-1\). The Black/neutral condition, which was expected to show some evidence of the positive bias, was assigned a weight of \(+1\), and the Black/threat condition - where the positive bias was expected to be distinctly strong - was assigned a weight of \(+3\). This formula was applied to all remaining outcomes. Table 1 presents the means and SDs for the main outcomes.
Trainees indicated on the Writer Rating Form how much additional work the essay required on content and on mechanics. Recall that content ratings, where the risk of seeming prejudiced is highest, were predicted to show the bias whereas mechanics ratings, where that risk is minimal, were not (per Harber, 1998). Results showed, as expected, that the joint effect of writer race and egalitarian self-image on content ratings was significant, and in the predicted manner, $t(97) = -3.67, p < .001$ (see Figure 1). Tukey post hoc tests show that trainees in the Black/threat condition rated essay content as requiring less added work than did trainees in the Black/boost condition, the White/neutral condition, and (marginally) the White/threat condition. The joint effect of writer race and egalitarian self-image condition on mechanics ratings was, as predicted, negligible, $t(100) = -0.28, p = .783$.

Trainees in the Black writer/self-image threat condition also advised the writer to spend fewer hours developing writing skills than did other trainees, $t(97) = -2.70,$
Furthermore, the amount of time recommended by the Black/threat trainees ($M = 2.82, SD = 1.57$) was roughly a third less than that recommended by trainees in the three White writer conditions, together ($M = 4.22, SD = 1.97$), $p = .032$.

**Copy-edited comments**

The index of copy-edit favourableness showed that copy-edit comments from trainees in the Black/threat condition were more favourable than those supplied by other trainees, $t(98) = 3.13$, $p = .002$ (see Figure 3). Copy-editing comments supplied by trainees in the Black/threat condition ($M = 1.34, SD = 2.47$) were more favourable than those supplied by trainees in the three White writer conditions, analysed as a single group ($M = -0.50, SD = 2.15$), $p = .049$.

The proportion of buffering comments (which mitigate criticism through apology or qualification) was also selectively high in the Black/threat condition, $t(100) = 3.19$,
The most marked disparity in buffers was between Black/threat trainees and Black/boost trainees, \( p = .077 \). Black/threat trainees also produced twice the rate of buffers (\( M = 0.077, SD = 0.081 \)) than did trainees in the three White writer conditions, analysed as a single group (\( M = 0.039, SD = 0.034 \), \( p = .024 \)).

**Adequacy of planned contrasts**

Analyses of residuals on the \(-1 + 1 + 3 - 1 - 1 - 1\) contrast model showed that there was no significant unexplained between-group variance in any of the four significant interactions (all \( ps > .25 \)). Our contrast model therefore fit the data well.

**Discussion**

Ideally, feedback suppliers concentrate their attention on learners’ needs; they monitor themselves, if at all, in order to convey feedback more effectively (Lepper, 1988). The present research tested whether this necessary focus on learners is subverted when feedback suppliers are White and feedback recipients are Black. It predicted that rather than attending solely to the needs of minority learners, White feedback suppliers may be distracted by their own self-image needs, causing them to place a positive bias on their feedback to minorities.

This prediction was confirmed. White teacher trainees whose egalitarian self-images were threatened, and who believed that their feedback recipient was Black, gave selectively more favourable essay-content ratings, recommended less time for developing writing skills, supplied more positive copy-editing comments, and laced these comments with more equivocating buffers. In contrast, trainees in the Black writer condition who had an opportunity to first affirm their egalitarianism did not display the positive bias on any of these measures; their feedback was virtually indistinguishable from feedback supplied by trainees in White writer conditions, who themselves were unaffected by the self-image manipulation. In effect, the trainees’ positive feedback bias to Blacks was ‘turned on’ when their self-image concerns were aroused, and ‘turned off’ when self-image concerns were allayed.
Connecting the positive bias to internal self-image motives is difficult because the conditions that arouse these motives can also introduce external social desirability concerns (MacDonald & Nail, 2005). To address this problem, this study led trainees to believe that (1) the Social Issues Survey (which manipulated self-images) was anonymous, was unrelated to the feedback study, and was not viewed by feedback study staff and (2) that their feedback was an anonymous, exclusive communication between just themselves and their student. Yet the private self-knowledge of having either compromised or confirmed their egalitarian scruples was sufficient to bias anonymous feedback to minorities, or to prevent such bias, respectively.

The self-image effects on feedback have implications for intergroup communication in general. Self-image concerns might contribute to Whites’ intergroup anxiety (Littleford et al., 2005), deterring Whites from interacting with minorities. When interactions do occur, self-image concerns can inhibit the candid disclosures through which lasting social bonds develop (Derlega & Berg, 1987). Also, the out-group deference arising from self-image concerns can place minorities in a perplexing social dilemma. How do they know whether Whites’ congeniality is in response to them, as individuals, or to the self-image threat that their minority-status arouses?

**Implications for minority schooling**

Chronic expression of the positive bias may dilute the academic challenge that is of especial importance to minority advancement (Steele, 1995). Indeed, field research indicates that inflated feedback can mislead minority students about where they need to exert effort (Massey, Scott, & Dornbusch, 1975). Tellingly, trainees in the Black/threat condition recommended one-third less time for writing remediation than did other trainees, reinforcing this point. A corollary liability of the positive bias is that it demoralizes minorities, suggesting that praise is a token to their race rather than a credit to their personal achievement (Crocker et al., 1991).

By identifying self-image concerns as a cause of the positive bias, this research provides an important first step in remediating it. Self-image concerns are fungible (Steele, 1988); if challenged in one domain, they can be supplemented in others. This suggests that shoring up teachers’ feelings of worth in ways unrelated to their egalitarianism may inure them to bias-inducing self-image concerns. Indeed, classroom teachers with more social support are less likely to show the positive bias than are teachers lacking support (Harber et al., 2009).

This study recruited teacher trainees as participants. Most of the trainees are now teachers, many in schools with large minority populations. As a representative sample of new teachers, their susceptibility to self-image concerns suggests that many minority students are exposed to the positive feedback bias. This exposure represents a serious liability if the bias deprives minority students of needed challenge or causes minorities to distrust praise, or be demoralized by it.

**Caveats**

**Manipulation checks**

Manipulation checks directly testing self-image threat, which would likely have produced reactivity, were not included in this study. Thus, although survey responses and feedback patterns were consistent with self-image threat, and
related research (Khan & Dhar, 2006) validated this kind of manipulation, threat itself was not otherwise measured.

Mood
Were trainees’ moods rather than their self-images altered, causing mood-induced stereotyping (e.g. Park & Banaji, 2000)? We think not; the boost condition led to equivalent rather than excessive criticism for Blacks relative to Whites; the bias was restricted to high-threat content and not low-threat mechanics; positive moods induce generosity rather than hostility to strangers (Isen, 2000); and self-image operates independently of mood (Steele, 1988). However, because mood was not measured, its influence cannot be fully determined.

Neutral condition
Trainees in the Black/neutral condition did not show the positive bias, perhaps because of the feedback skills (e.g. Cohen, Steele, & Ross, 1999) and self-efficacy (Bandura, 1977) teacher training provides.

Conclusion
This study does not discount the importance of egalitarianism. Rather, it suggests that for Whites, as for minorities, intergroup contact can present self-image risks that deter the authentic, non self-conscious communications through which true egalitarianism is displayed.

Acknowledgements
This research was sponsored by the National Science Foundation (Grant #BCS 0416889) and the Spencer Foundation (Grant #200400104). We thank Amanda Browning and Thomas O’Connell for assisting this research, and Dr Jeffery Smith and Ms Carrie Ambrecht of the Graduate School of Education, Rutgers New Brunswick, and Donna Henderson-King and Eaaron Henderson-King for their comments.

References


Received 18 February 2009; revised version received 21 August 2009