Social Cognitive and Affective Neuroscience Advance Access published December 20, 2016

Self-distancing Improves Interpersonal Perceptions and Behavior by Decreasing Medial Prefrontal Cortex Activity During the Provision of Criticism

Jordan B. Leitner[1], Ozlem Ayduk[1], Rodolfo Mendoza-Denton[1], Adam Magerman[2], Rachel Amey[2], Ethan Kross[3], Chad E. Forbes[2]

[1]Department of Psychology, University of California, Berkeley

[2]Department of Psychological and Brain Sciences, University of Delaware

[3]Department of Psychology, University of Michigan, Ann Arbor

Author note

This research is supported by the National Science Foundation under Award Numbers 1306709 and 1514510. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Correspondence concerning this article may be addressed to Jordan B. Leitner (jleitner@berkekley.edu); 3210 Tolman Hall, Berkeley, CA, 94720.

Abstract

Previous research suggests that people show increased self-referential processing when they provide criticism to others, and that this self-referential processing can have negative effects on interpersonal perceptions and behavior. The current research hypothesized that adopting a self-distanced perspective (i.e., thinking about a situation from a non-first person point of view), as compared to a typical self-immersed perspective (i.e., thinking about a situation from a first-person point of view), would reduce self-referential processing during the provision of criticism, and in turn improve interpersonal perceptions and behavior. We tested this hypothesis in an interracial context since research suggests that self-referential processing plays a role in damaging interracial relations. White participants prepared for mentorship from a self-immersed or self-distanced perspective. They then conveyed negative and positive evaluations to a Black mentee while EEG was recorded. Source analysis revealed that priming a self-distanced (vs. self-immersed) perspective predicted decreased activity in regions linked to self-referential processing (medial prefrontal cortex; MPFC) when providing negative evaluations. This decreased MPFC activity during negative evaluations, in turn, predicted verbal feedback that was perceived to be more positive, warm and helpful. Results suggest that self-distancing can improve interpersonal perceptions and behavior by decreasing self-referential processing during the provision of criticism.

Keywords: Intergroup dynamics; Prejudice; Racial and ethnic attitudes and relations; Electrophysiology; Mentorship

Self-distancing Improves Interpersonal Perceptions and Behavior by Decreasing Medial Prefrontal Cortex Activity During the Provision of Criticism

Many academic and professional relationships rely on the provision of evaluative feedback. For instance, mentors are often responsible for conveying criticism to their mentees. How might neurocognitive processes during the provision of this criticism relate to interpersonal perceptions and behavior? While research has focused extensively on the neural processes associated with receiving feedback (e.g., Cohen, Elger, & Ranganath, 2007; Hauser et al., 2014; Moser & Simons, 2009; Talmi, Atkinson, & El-Deredy, 2013), little is known about the neural processes that emerge when *conveying* feedback to others. A deeper understanding of these processes is needed, as they may provide insight into factors that disrupt or improve interpersonal relationships.

Self-referential processing during criticism

In relationships that involve the provision of evaluative feedback (e.g., mentorship), one disruptive factor might be that feedback-givers experience increased self-referential processing (i.e., focus on one's own thoughts, feelings, and behaviors) when conveying criticism. Supporting this view, people avoid conveying critical feedback to others, especially when the critic's identity is known (Jeffries & Hornsey, 2012; DePaulo & Bell 1996; Tesser & Rosen, 1972). As such, when people do convey criticism, they may become entangled in thoughts about whether the criticism portrays them as insensitive or overly harsh. Though conveying criticism may increase self-referential processing in a range of contexts, this link may be particularly robust when the feedback-giver is White, and the feedback-receiver is a racial minority. Indeed, research suggests that Whites' criticism in interracial relationships may be interpreted as racially motived (Mendoza-Denton et al., 2002; 2010), and Whites often worry they will appear prejudiced (Butz & Plant, 2006; Plant & Devine, 2003) or offend minorities (Crosby & Monin, 2007; Goff et al., 2008).

Importantly, research suggests that this self-referential processing during the provision of criticism can have negative implications for interpersonal perceptions and behavior. Specifically, increased self-focus is related to increased negative affect (Mor & Winquist, 2002) and lower self-efficacy in social situations (Kashdan & Roberts, 2004; Woody & Rodriguez, 2004). Furthermore, those who ruminate on negative thoughts show more negative perceptions of others (Lam, Schuck, Smith, Farmer, & Checkley, 2003; Pearson, Watkins, Mullan, & Moberly, 2010; Takano, Sakamoto, & Tanno, 2011) and poorer interpersonal problem solving (Lyubomirsky & Nolen-Hoeksema, 1995). Specific to interracial contexts, research suggests that Whites' self-referential processing (e.g., Whites' focus on whether criticism reflects their prejudice) is related to negative perceptions of Blacks (Plant & Devine, 1998), negative inter-group interactions (Vorauer, 2006; Vorauer & Turpie, 2004), and less genuine feedback in mentoring contexts (Crosby & Monin, 2007). Thus, we expected that increased self-referential processing when conveying criticism might contribute to more negative interpersonal perceptions and behaviors.

Self-immersed vs. self-distanced perspectives

Self-referential processing typically occurs from an internal, first-person perspective (referred to as a *self-immersed perspective*). For instance, Kross and Ayduk (2008) showed that when engaging in internal dialogue (i.e., self-talk), people often refer to themselves using the words "I" and "my." However, research has also found that people can adopt a *self-distanced* perspective, wherein they observe and analyze their experience from an external observer's point of view. One way a person can adopt a self-distanced perspective is by changing internal dialogue to refer to the self with the word "you" or one's own name (Kross et al., 2014; also see Grossmann & Kross, 2014). For example, Molly could adopt a self-distanced perspective by saying to herself, "*Molly*, *you* feel nervous," as opposed to saying to herself "*I* feel nervous."

Because a self-distanced perspective psychologically "transforms" the part of the self being observed into an "other," it may attenuate processes typically involved in self-referential thought. Additionally, evidence suggests that self-distancing decreases rumination, mitigates distress, and improves performance (Ayduk & Kross, 2008; 2010; Katzir & Eyal, 2013; Kross & Ayduk, 2008; Kross et al., 2011, 2014; McIsaac & Eich, 2004; Verduyn et al., 2012). Furthermore, adopting a self-distanced perspective has been linked to increased consideration of opposing viewpoints (Kross & Grossman, 2012), constructive behaviors towards dating partners during conflicts (Ayduk & Kross, 2010), and decreased reactivity towards a transgressor (Mischkowski et al., 2012). Thus, we hypothesized that feedback-givers who are led to adopt a self-distanced perspective would show decreased self-referential processing when they provide criticism, and in turn convey feedback that was more positive, warm, and helpful.

Insight from neural activity

One way to gain insight into the magnitude and timing of self-referential processing during the provision of criticism is to measure EEG activity, and conduct source analysis. Source analysis builds upon traditional event-related potential (ERP) analysis in two ways. First, whereas ERP analysis provides little insight into neural generators, source analysis estimates activity in neural generators (see Grech et al., 2008 for a review). Second, researchers who employ ERP analysis in novel paradigms are often faced with ambiguity regarding whether a given voltage deflection in sensor-level data is a well-studied ERP component, or whether they have discovered a new ERP component that is specific to their paradigm (Kappenman & Luck, 2012). This ambiguity is reduced in source analysis since it models activity at the neural source level, as opposed to the sensor level.

Research employing EEG and source analysis has suggested that self-referential processes are supported by the medial prefrontal cortex (MPFC). For instance, EEG activity source-localized to the MPFC is greater when thinking about self-related (vs. other-related) adjectives (Esslen et al., 2008) and pronouns (Shi et al., 2011). EEG results linking self-referential processes to the MPFC are also consistent with fMRI research showing that the MPFC supports self-referential processes (Northoff et al., 2006), and that MPFC activity decreases when people construe past negative emotions as psychologically distant from the self (Kross et al., 2009). Thus, to the extent that adopting a self-distanced perspective reduces self-referential processing during the provision of criticism, adopting a self-distanced perspective should decrease MPFC activity during the provision of criticism.

Furthermore, to the extent that self-distancing mitigates negative emotions, self-distancing may influence neural regions that support the cognitive control of emotion, including the lateral prefrontal cortex (LPFC; see Buhle et al., 2013. for a meta-analysis). However, as some research has found that self-distancing is unrelated to LPFC activity (Moser et al., under review), the possibility remains that self-distancing reduces negative affect relatively effortlessly, without recruiting cognitive control networks.

Current research

The current research tested the hypothesis that priming individuals to self-distance would reduce MPFC activity during the provision of criticism, which in turn would predict more positive interpersonal perceptions and behavior. Participants prepared for a mentoring experience from either a self-distanced vs. self-immersed perspective, after which they conveyed negative and positive evaluations to a mentee while continuous EEG activity was recorded. Finally, we examined whether mentors who showed decreased MPFC activity when conveying negative evaluations would perceive their mentee more positively, and ultimately provide verbal feedback that was perceived to be warmer and more helpful.

We examined these issues in the context of interracial mentorship, consisting of a White participant and Black mentee. We constrained race in this study since the aforementioned research suggests that criticism in an interracial context would evoke self-referential processing. Moreover, a deeper insight into the processes that shape interracial mentorship would be valuable, given that interracial vs. same-race mentoring relationships show poorer outcomes (McDonald & Westphal, 2012; Woolf, Cave, Greenhalgh, & Dacre, 2008). Nevertheless, we posited that the current design might reveal processes that emerge in contexts that generalize beyond interracial mentorship.

While we expected that self-referential processing would emerge when people provide criticism to others, we were unsure about whether self-referential processing would emerge when people convey praise. On the one hand, providing praise might absolve a feedback-giver's concerns about appearing insensitive. As such, people might show higher self-referential processing when they provide criticism vs. praise. On the other hand, praise can portray the feedback-giver as patronizing (Major, Sawyer, & Kuntsman, 2013; Vescio, Gervais, Snyder, & Hoover, 2005). Thus, people may show equivalent self-referential processing when they provide criticism and praise. Due to this ambiguity regarding self-referential processing during praise, our hypotheses focused primarily on MPFC activity during the provision of criticism. Nevertheless, we also explored neural activity during the provision of praise.

Method

Participants and design. 59 White right-handed introductory psychology students (29 male; M_{age} = 18.92, SD_{age} = 1.09) participated in a 2 (perspective: self-immersed vs. self-distanced) x 2 (evaluation: negative vs. positive) mixed-model design with repeated measures on the second factor. Participants were recruited if they reported on a pre-screening survey that they had not experienced traumatic brain injury, brain surgery, or seizures. 4 participants had fewer than 10 valid EEG trials per condition (see below for details), and were thus excluded from analyses. A subsample (*n* = 41) completed the video feedback portion of the study. Thus, 55 participants were included in the main EEG analyses, and 41 participants were included in analyses that examined video feedback. We determined this sample size from a power analysis, which indicated that achieving .80 power at an alpha level of .05 for a medium-size within-between subjects interaction effect required a sample of 34 participants. We exceeded this minimum sample size, as we anticipated that some participants might be excluded due to noisy EEG data.

Procedure. Participants were fitted with a 64-channel EEG cap, and given a description of the task. The cover story indicated that participants were to adopt the role of a mentor, and evaluate a speech given by another student ("the mentee"). Ostensibly, "the mentee" had recorded a speech on the topic of why he was qualified for his dream job, was seated in an adjacent room, and would receive the evaluations that the participant provided. In reality, all participants viewed the same mentee speech given by a Black male confederate.

Perspective manipulation. Participants viewed the entire speech, after which they were assigned to adopt a self-distanced or self-immersed perspective, based on procedures in Kross et al. (2014). Participants in the self-distancing condition were instructed to use their own name and the pronoun "you" to refer to themselves as they reflected on their emotions about evaluating their mentee. Participants in the self-immersed condition were instructed to use the pronouns “I” and “my” as they reflected on their emotions about evaluating their mentee. Participants were given three minutes to silently reflect using this language technique.

Manipulation check. We assessed the effectiveness of the manipulation with an item used in previous work (Kross et al., 2014). Specifically, participants in the self-immersed condition indicated from 1 (*not at all*) to 5 (*exclusively*) the extent to which they used the first-person pronouns *I* and *me* to refer to themselves as they reflected on their emotions. Participants the self-distanced condition indicated from 1(*not at all*) to 5 (*exclusively*) the extent to which they used *you* and *your own name* to refer to themselves as they reflected on their emotions. Thus, for all participants, this manipulation check assessed the degree to which they followed instructions (M = 3.84, SD = 1.07).

EEG evaluation task. Next, participants viewed their mentee’s speech for a second time, and provided evaluations that were ostensibly sent to the mentee. Specifically, after every 6 seconds of the mentee’s speech (70 trials throughout the speech), participants were prompted to indicate whether the preceding speech segment was weak or strong by pressing buttons that sent a negative evaluation (i.e., an image of a “thumbs-down”) or positive evaluation (i.e. an image of a “thumbs-up”) to the mentee. The experimenter informed the participant that this task would help the mentee identify the weakest and strongest moments of his speech. This task was ideal for capturing neural activity associated with the provision of feedback, as it allowed us to time-lock EEG activity to participants’ responses.

To obtain sufficient trials for EEG analysis, it was important that participants gave equal amounts of negative and positive evaluations. Accordingly, participants were informed that their task was to identify the 35 weakest segments, and 35 strongest segments of their mentee’s speech. To ensure that participants followed these instructions, the experimenter tracked the participants' evaluations. Specifically, after every 10 trials, the participant would view a screen that indicated, "please wait for the experimenter." At this point, the experimenter would inform the participants how many negative and positive evaluations they could still convey. For example, if the participant had provided 5 negative and 15 positive evaluations, the experimenter would indicate that the participant had 30 negative and 20 positive evaluations remaining. As a result, all participants gave the same number of positive and negative evaluations.

Verbal feedback. Next, participants were instructed to provide verbal feedback that would ostensibly be sent to the mentee. Participants were instructed to speak freely, and give honest and constructive feedback. To assess whether participants perceived their mentee positively, three blind coders rated these videos for feedback valence (i.e., degree to which mentors' evaluation was negative vs. positive) on a 1 (*very negative*) to 7 (*very positive*) scale (M = 3.85, SD = 1.02). Additionally, three coders rated these videos for warmth (M = 3.82, SD = 1.16) and helpfulness (M = 4.25, SD = 1.26) on a 1 (*not at all*) to 7 (*very much*) scale. We examined feedback valence, warmth, and helpfulness individually (as opposed to creating a composite variable) since these constructs are theoretically independent (e.g., more positive feedback can be perceived as either helpful or condescending). Moreover, examining these constructs individually provided an opportunity to examine whether more positive mentee perceptions might underlie warm and helpful behavior. Inter-coder reliability was acceptable for feedback valence (ICC = .74), warmth (ICC = .69), and helpfulness (ICC = .78).

EEG Recording. Continuous EEG activity was recorded using an ActiveTwo head cap and the ActiveTwo BioSemi system (BioSemi, Netherlands). Recordings were collected from 64 Ag-AgCl scalp electrodes, and 2 electrodes placed around the left eye. A ground electrode was formed by BioSemi's Common Mode Sense active electrode and the Driven Right Leg passive electrode. EEG activity was digitized with ActiView software (BioSemi) and sampled at 2048 Hz. Data were analyzed with Brain Electromagnetic Source Analysis (BESA) 5.3 software (MEGIS Software GmbH, Grafelfing, Germany). Data were down-sampled at 512 Hz, referenced to original average activity, band-pass filtered between .1 and 30 hz, and response-locked to the button-press of participants' evaluations. Epochs extended from 500 ms pre-response to 1000 ms post-response. Ocular artifacts were corrected via the adaptive algorithm implemented in BESA, after which we identified and rejected epochs containing non-ocular artifacts (amplitude > 120 μV, gradients > 75 μV, low signal < 0.01). Epochs were baseline-corrected by subtracting the average value of activity -400 to -200 ms from the entire epoch. Epochs that were free of non-ocular artifacts were considered "valid trials." Participants included in analysis had, on average, 28 negative evaluation trials (SD = 6.53, minimum = 16), and 27 positive evaluation trials (SD = 6.47, minimum = 13). The number of valid trials required to be included in analyses is consistent with previous work (Forbes & Leitner, 2014; Leitner et al., 2014).

EEG analytic approach: Source analysis. We estimated the amplitude of neural regions involved in participants' conveyed evaluations by performing dipole source modeling in BESA, consistent with previous research (Forbes & Leitner, 2014; Leitner et al., 2014). BESA transforms electrode space into source space to fit the location and orientation of dipole sources that explain the greatest variance of scalp EEG. This approach uses data from all electrodes, and acts as a spatial filter to represent ongoing EEG activity as source waveforms. To avoid biasing the results toward any one condition, source analysis was conducted on a grand average waveform comprised of both negative and positive evaluations. In order to apply the same spatial filtering method to eye movement and brain activity, and best model eye-related activity in the EEG signal, source analyses were performed on data that were not corrected for ocular artifacts, consistent with previous work (Leitner et al., 2014).

Given our hypotheses regarding the MPFC and LPFC, we constructed an a priori dipole model, wherein we planted one source in the MPFC, and bilateral regional sources in the left and right lateral PFC. Additionally, to control for activity that did not originate from these frontal sources, we planted bilateral regional sources in the eyes (to account for eye movements), motor cortex (to account for hand movements), and occipital cortex (to account for visual processing; Fig. 1).

To determine the time window for applying this source model, we inspected grand-average ERP waveforms (Fig. 2). Visual inspection of these waveforms showed an early medial-frontal negativity between approximately 200 ms pre-response and 100 ms post-response, and a later frontal positivity (and parietal negativity) between approximately 200-400 ms post-response. As such, we applied our source model to the -200 ms to 400 ms window surrounding participants' responses. This model accounted for 99.4% of total variance in the grand-average EEG waveform between -200 pre-response and 400 ms post-response. To account for individual variability in participants' brain anatomy, regional sources were uniquely oriented for each participant's average waveform.

Next, we examined source waveform amplitudes during two time windows. To capture "early neural activity" involved in the anticipation and administration of mentor evaluations, we computed mean amplitude between 200 ms pre-response to 100 ms post-response. This time window is consistent with the fronto-central ERP voltage deflection (Fig. 2), along with research that has examined response-locked ERP components that emerge between -200 pre-response and

100 ms post-response, including the event-related negativity (Risel et al., 2013), and the N-40 (Vidal et al., 2011). To capture "later neural activity" that occurred after the participant's response, and during the second frontal ERP voltage deflection, we computed mean source amplitude between 200-400 ms post-response. All analyses were conducted on the primary orientation of each regional source.

Results

To explore potential gender effects, we first modeled gender as a between-subjects factor in all ANOVAs predicting early and late amplitude (described below). No effects including gender were significant, *p*s > .07. Thus, we report analyses that did not include gender.

Manipulation check. We examined responses to the manipulation check to determine whether participants in the self-immersed condition used first-person language, and participants in the self-distanced condition used non-first-person language when reflecting on their emotions. A one-sample *t*-test revealed that mean responses to the manipulation check were significantly above the scale midpoint, $t(48)$[1] = 5.49, $p < .001$, indicating that participants followed instructions. Furthermore, a one-way ANOVA indicated that participants in the self-distancing ($M = 3.86$, $SD = 1.18$) and self-immersed ($M = 3.81$, $SD = .93$) conditions reported that they followed instructions to an equivalent degree, $F(1, 47) = .02$, $p = .88$. `

Early neural activity. To determine whether perspective or evaluation valence influenced neural activity during the early response period (-200 to 100 ms), we conducted a 2(perspective: self-immersed vs. self-distanced) x 3(source: MPFC vs. left LPFC vs. right LPFC) x 2(evaluation: positive vs. negative) mixed-model ANOVA with repeated measures on the 2nd and 3rd factors. We applied the Greenhouse-Geisser correction to all analyses. A main effect emerged for evaluation valence, such that amplitudes were greater when conveying negative (*M*

= .18, *SD* = 7.74) compared to positive (*M* = -1.72, *SD* = 7.95) evaluations, $F(1, 53) = 5.36$, p = .024, $\eta^2 = .09$. Additionally, a significant 3-way interaction emerged, $F(1.88, 99.73) = 5.26$, p = .008, $\eta^2 = .09$ (Fig. 3). No other effects were significant, p*s* > .069.

We conducted several follow-up analyses to decompose this 3-way interaction. First, to test whether self-distanced (vs. self-immersed) perspective reduced amplitude at any source when conveying negative feedback, we conducted a 2(perspective) x 3(source) ANOVA predicting neural reactivity during negative evaluations. Including source as a factor in this analysis was valuable, since the perspective x source interaction term tested whether any perspective effects varied across sources. This analysis revealed a main effect for source, *F*(1.6., 84.01) = 3.84, p = .035, $\eta^2 = .07$, which was qualified by a significant 2-way interaction $F(1.6, 84.01) = 6.10$, p = .006, $\eta^2 = .10$. Simple effects analyses indicated that, as hypothesized, MPFC activity when conveying negative evaluations was lower for participants in the self-distanced than self-immersed condition, $F(1, 53) = 4.17$, p = .046, $\eta^2 = .07$. In contrast, left LPFC activity during negative evaluations was greater in the self-distanced than self-immersed condition, $F(1, 53) = 7.77$, p = .007, $\eta^2 = .13$. Perspective was unrelated to right LPFC reactivity during negative evaluations p = .261. Thus, self-distancing decreased MPFC activity and increased left LPFC activity during the anticipation and delivery of criticism.

Second, we tested whether perspective had an effect on amplitude during positive evaluations, and whether such an effect varied across sources. A 2(perspective) x 3(source) ANOVA predicting amplitude during positive evaluations revealed no main effects or interactions, *p*s > .12.

Third, we tested the effects of feedback valence and perspective condition within each source. When predicting early MPFC amplitude, a 2(feedback valence) x 2(perspective)

ANOVA revealed a main effect for feedback valence, $F(1, 53) = 4.79$, $p = .033$, $\eta^2 = .08$, such that MPFC amplitude was greater when conveying negative compared to positive evaluations. Neither the main effect of perspective, $p = .096$, nor the interaction, $p = .204$, were significant.

When predicting early left LPFC amplitude, a 2(feedback valence) x 2(perspective) ANOVA revealed a significant interaction, $F(1, 53) = 11.62$, $p = .001$, $\eta^2 = .18$. In the self-immersed condition, left LPFC amplitude was significantly greater when conveying positive than negative evaluations, $F(1, 22) = 6.01$, $p = .023$, $\eta^2 = .22$. In contrast, in the self-distanced condition, left LPFC amplitude was significantly greater when conveying negative than positive evaluations, $F(1, 31) = 9.05$, $p = .005$, $\eta^2 = .23$. No other effects were significant, *ps* > .183.

Finally, when predicting early right LPFC amplitude, a 2(feedback valence) x 2(perspective) ANOVA revealed no significant main or interaction effects, *ps* > .211.

Later neural activity. To determine whether perspective and evaluation valence predicted later activity, we conducted the same 2(perspective) x 3(source) x 2(evaluation) mixed-model ANOVA described above on the 200-400 ms epoch. Consistent with the early response period, a main effect emerged for evaluation valence, $F(1, 53) = 9.34$, $p = .004$, $\eta^2 = .15$, such that amplitudes were greater when conveying negative compared to positive evaluations. Additionally, a main effect emerged for source, $F(1.47, 77.87) = 34.90$, $p < .001$, $\eta^2 = .40$, such that amplitudes were greater at the MPFC source, compared to the left LPFC and right LPFC sources, helmert contrast: $F(1, 53) = 38.76$, $p < .001$, $\eta^2 = .422$. No other effects were significant, *p*s > .21, indicating that perspective condition did not influence frontal source amplitude during this later time period.

Verbal feedback. The aforementioned analyses indicated that mentors who adopted a self-distanced perspective showed decreased early MPFC and increased early left LPFC activity when conveying criticism. We sought to examine whether these patterns of neural activity predicted the verbal feedback participants conveyed to their mentee. Accordingly, we employed regression-based path analysis using the PROCESS module (Hayes, 2013; model 4). We began by modeling perspective condition as a predictor, video feedback valence as an outcome variable, and the following as simultaneous parallel mediators: early MPFC and left LPFC activity when conveying negative evaluations.

MPFC activity during negative evaluations significantly mediated the relationship between perspective and video feedback valence, indirect effect a_1b_1 = .20, 95% CI [.02, .59] (Fig. 4 A). That is, self-distancing predicted lower MPFC activity when conveying negative (i.e. thumbs down) evaluations, which in turn predicted more positive verbal feedback. Notably, the indirect effect was not significant for left LPFC amplitude during negative evaluations, 95% CI [-.05, .41]. Thus, decreased MPFC amplitude during negative evaluations was the only significant mediator between self-distancing and verbal feedback valence.

Finally, we tested whether mentors who gave more positive verbal feedback appeared more warm and helpful (rather than cold and condescending). Accordingly, we tested for serial mediation (Hayes, 2013; model 6) by modeling perspective as a predictor, early MPFC amplitude during negative evaluations and feedback valence as serial mediators, and verbal feedback warmth and helpfulness as separate outcome variables. The indirect effects were significant for both warmth, $a_1d_{21}b_2$ = .14, 95% CI [.01, .440] (Fig. 4B) and helpfulness, $a_1d_{21}b_2$ = .12, 95% CI [.01, .43] (Fig. 4C). Self-distancing predicted decreased MPFC activity when conveying negative evaluations, which in turn predicted more positive verbal feedback. More positive verbal feedback was, in turn, interpreted as more warm and helpful. See Supplementary Materials for additional analyses.

Discussion

Individuals who adopted a self-distanced, relative to a self-immersed, perspective showed decreased neural activity in regions associated with self-referential processing (i.e., MPFC) during the provision of negative evaluations. Critically, this decreased MPFC activity predicted more positive perceptions, and in turn, verbal feedback that was perceived as more warm and helpful. While previous research has examined neural activity associated with *receiving* feedback (e.g., Cohen et al., 2007; Hauser et al., 2014; Moser & Simons, 2009; Talmi et al., 2013), this is the first study, to our knowledge, to examine the neural processes associated with *conveying* evaluative feedback.

Notably, these findings are consistent with fMRI research that has suggested that self-distancing decreases MPFC activity (Kross et al., 2009), and behavioral research showing that self-distancing contributes to positive interpersonal outcomes (Ayduk & Kross, 2010; Kross & Grossman, 2012; Mischowski et al., 2012). Furthermore, that decreased MPFC activity predicted positive and helpful feedback is consistent with research that has suggested that interracial interactions might be strained by excessive self-referential processing (Butz & Plant, 2006; Goff et al., 2008; Plant & Devine, 2003; Sasaki & Vorauer, 2010). However, the current findings extend this previous work by showing that a self-distanced perspective corresponded with more positive perceptions and behavior by reducing MPFC activity during the provision of negative evaluations.

The current results revealed that self-distancing corresponded with decreased activity in the left LPFC when conveying praise, but increased activity in the left LPFC when conveying criticism. The effect of self-distancing when conveying criticism is consistent with research showing that self-distancing reduces negative affect (Kross et al., 2014), and LPFC activity contributes to the reappraisal of negative affect (Buhle et al., 2013). Nevertheless, the lack of association between LPFC activity and mentoring behavior suggests that LPFC, as compared to MPFC, activity was a less critical factor in determining how participants would behave. Future research might continue to explore the role of LPFC activity when people convey feedback to others.

Our analytic approach modeled neural activity evoked by button press. What psychological processes might this evoked activity reflect? One possibility is that this activity reflects a set of self-referential processes that emerge in the moment before conveying feedback to another person. For instance, people may "self-check" that they are about to convey feedback that is socially appropriate. Supporting this possibility, the grand-average waveform (Fig. 2) showed a medial frontal negativity immediately before the button press, and medial frontal negativities around the time of a response have been linked to increased monitoring of one's own behavior (Luu, Flaisch, & Tucker, 2000). Another possibility is that the evoked activity reflected conflict between a strategy (e.g., "don't offend the mentee") and the impending behavior (e.g., conveying criticism). Supporting this possibility, medial frontal negativities can reflect conflict between a strategy and an action (Barthalow et al., 2005). Moreover, people avoid giving others negative feedback (Jeffries & Hornsey, 2012), and MPFC amplitude was greater following negative vs. positive feedback. However, given the novelty of this paradigm, future research will be needed to further explore these possibilities.

As noted above, we interpret decreased MPFC amplitude to reflect decreased self-referential processing. This notion is supported by forward inference, as self-distancing decreases self-referential processing at a psychological level (i.e., self-distancing "transforms" the part of the self that is being observed into an "other"), and self-distancing decreased MPFC amplitude. Additionally, this interpretation is supported by reverse inference, as the MPFC has been linked to self-referential processes in research employing EEG (Esslen et al., 2008; Knyazev, 2013; Shi et al., 2011) and fMRI (Northoff et al., 2006). However, the MPFC also supports an array of social cognitive processes, including the mapping of relations between others and oneself (Han et al., 2005; Iacoboni et al., 2004; Mitchell et al., 2005; Ochsner et al., 2005; Seger et al., 2004), and taking the perspective of another person (Buckner et al., 2008). Thus, it is possible that MPFC activity in the current study reflected multiple processes. Future research might continue to examine which psychological processes were most active in the current paradigm.

In the current study, we recruited only White participants, and all participants provided feedback to a Black mentee. As such, a remaining question is whether the current effects are specific to interracial mentoring contexts, or generalize to any context in which people convey feedback to others. Future research might examine situational features that moderate (a) the relationship between self-distancing (vs. self-immersion) and MPFC activity when conveying criticism to others, and (b) the relationship between MPFC activity when conveying criticism to others and interpersonal perceptions and behavior.

While the current findings elucidate one pathway through which self-distancing predicted mentoring behavior, it may be important to examine other additional mediating pathways. For instance, reduced distress regarding appearing prejudiced may have been a pathway through which self-distancing predicted more positive mentoring outcomes. Supporting this possibility, self-distancing has been shown to reduce distress (Ayduk & Kross, 2008; 2010; Katzir & Eyal, 2013; Kross & Ayduk, 2008; Verduyn et al., 2012), and reduced distress regarding intergroup interactions has been linked to more effective intergroup communications (Ulrey & Amason, 2001). Thus, it will be important for future research to examine how distress and anxiety are related to the neural activity and behavioral outcomes assessed in the current work.

Finally, while the current research focused on the effects of self-distancing when conveying feedback, future research should also focus on the potential effects of self-distancing when receiving feedback. Specifically, it would be important to examine whether people show more positive interpersonal perceptions and behavior when they adopt a self-distanced perspective prior to receiving criticism. Exploring these research questions will be important for understanding ways of fostering positive relationships.

Footnotes

[1]Due to computer error, manipulation check data were missing for 6 participants.

References

Ayduk, O., & Kross, E. (2008). Enhancing the pace of recovery self-distanced analysis of negative experiences reduces blood pressure reactivity. *Psychological Science*, *19*, 229-231.

Ayduk, O., & Kross, E. (2010). From a distance: Implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology*, *98*, 809-829.

Boykin, M., Mendoza-Denton, R. & Patt, C. (2015). Stigma: Implications for student achievement and mentoring. In C. Rubie-Davies (Ed.), *The Routledge International Handbook of Social Psychology of the Classroom* (pp. 93-101). New York: Routledge.

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., ... & Ochsner, K. N. (2013). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cerebral Cortex*, bht154.

Butz, D. A., & Plant, E. A. (2006). Perceiving outgroup members as unresponsive: implications for approach-related emotions, intentions, and behavior. *Journal of Personality and Social Psychology*, *91*, 1066-1079.

Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage*, *35*(2), 968-978.

Colby, S. L., & Ortman, J. M. (2014). Projections of the Size and Composition of the U.S. Population: 2014 to 2060 Current Population Reports. *Current Population Reports*, 25–1143. Retrieved from http://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf

Costello, K., & Hodson, G. (2011). Social dominance-based threat reactions to immigrants in need of assistance. *European Journal of Social Psychology, 41*, 220-231.

Crosby, J., & Monin, B. (2007). Failure to warn: How student race affects warnings of potential academic difficulty. *Journal of Experimental Social Psychology*, *43*, 663-670.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of personality and social psychology, 71*(4), 703.

Esslen, M., Metzler, S., Pascual-Marqui, R., & Jancke, L. (2008). Pre-reflective and reflective self-reference: A spatiotemporal EEG analysis. *NeuroImage*, *42*, 437–449.

Forbes, C. E., & Leitner, J. B. (2014). Stereotype threat engenders neural attentional bias toward negative feedback to undermine performance. *Biological Psychology, 102,* 98-107.

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology*, *94*, 91-107.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., ... & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation, 5*.

Han, S., Jiang, Y., Humphreys, G. W., Zhou, T., and Cai, P. (2005). Distinct neural substrates for the perception of real and virtual visual worlds. *Neuroimage, 24*, 928–935.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage, 84*, 159-168.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., et al. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage, 21,* 1167–1173.

Jeffries, C. H., & Hornsey, M. J. (2012). Withholding negative feedback: Is it about protecting the self or protecting others?. *British Journal of Social Psychology*, *51*, 772-780.

Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S.J. Luck and E.S. Kappenmann (Eds.) *The Oxford Handbook of Event-Related Potential Components* (pp. 3-30). New York: Oxford University Press.

Kashdan, T. B., & Roberts, J. E. (2004). Social anxiety's impact on affect, curiosity, and social self-efficacy during a high self-focus social threat situation. *Cognitive Therapy and Research*, *28*(1), 119-141.

Katzir, M., & Eyal, T. (2013). When stepping outside the self is not enough: A self-distanced perspective reduces the experience of basic but not of self-conscious emotions. *Journal of Experimental Social Psychology*, *49*, 1089-1092.

Knyazev, G. G. (2013). EEG correlates of self-referential processing. *Frontiers in Human Neuroscience*, *7*, 1–14.

Kross, E., & Ayduk, O. (2008). Facilitating adaptive emotional analysis: Distinguishing distanced-analysis of depressive experiences from immersed-analysis and distraction. *Personality and Social Psychology Bulletin, 34,* 924-938

Kross, E., & Ayduk, O. (2011). Making meaning out of negative experiences by self-distancing. *Current Directions in Psychological Science, 20,* 187-191.

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., ... &

Ayduk, O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology*, *106,* 304-324.

Kross, E., Davidson, M., Weber, J., & Ochsner, K. (2009). Coping with emotions past: the neural bases of regulating affect associated with negative autobiographical memories. *Biological Psychiatry*, *65*, 361-366.

Kross, E., Duckworth, A. L., Ayduk, Ö., Tsukayama, E., & Mischel, W. (2011). The effect of self-distancing on adaptive versus maladaptive self-reflection in children. *Emotion, 11,* 1032–1039.

Kross, E., Gard, D., Deldin, P., Clifton, J., & Ayduk, O. (2012). "Asking why" from a distance: Its cognitive and emotional consequences for people with major depressive disorder. *Journal of Abnormal Psychology*, *121*, 559-569.

Kross, E., & Grossmann, I. (2012). Boosting wisdom: distance from the self enhances wise reasoning, attitudes, and behavior. *Journal of Experimental Psychology: General*, *141*, 43-48.

Landivar, L. C. (2013). Disparities in STEM employment by sex, race, and Hispanic origin. *Education Review*, *29*, 911-922.

Leitner, J. B, Hehman, E., Jones, J., & Forbes. C. (2014). Self-enhancement influences medial frontal cortex alpha power to social rejection feedback. *Journal of Cognitive Neuroscience, 26,* 2330-2341.

McDonald, M. & Westphal, J. 2012. Access denied: Low mentoring of women and minority first-time directors and its negative effects on appointments to additional boards. *Academy of Management Journal, 56,* 1169-1198.

McIsaac, H. K., & Eich, E. (2004). Vantage point in traumatic memory. *Psychological*

Science, *15*, 248-253.

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: implications for African American students' college experience. *Journal of Personality and Social Psychology*, *83*, 896-918.

Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G., & Aceves, M. J. (2010). Group-value ambiguity understanding the effects of academic feedback on minority students' self-Esteem. *Social Psychological and Personality Science*, *1*, 127-135.

Mischkowski, D., Kross, E., & Bushman, B. J. (2012). Flies on the wall are less aggressive: Self-distancing "in the heat of the moment" reduces aggressive thoughts, angry feelings and aggressive behavior. *Journal of Experimental Social Psychology*, *48*, 1187-1191.

Mitchell, J. P., Banaji, M. R., and Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neurosci*ence, *17,* 1306–1315.

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: a meta-analysis. *Psychological Bulletin*, *128*, 638-662.

Moser, J. S., Dougherty, A., Moran, T. P., Guevarra, D., Shablack, H., Ayduk, O…& Kross, E. (under review). Third person self-talk facilitates emotional control without engaging cognitive control.

Moser, J. S., & Simons, R. F. (2009). The neural consequences of flip-flopping: The feedback-related negativity and salience of reward prediction. *Psychophysiology*, *46*, 313-320.

National Center for Education Statistics (NCES; 2014). Total fall enrollment in degree-granting postsecondary institutions, by level of enrollment, sex, attendance status, and race/ethnicity of student: Selected years, 1976 through 2012. Retrieved February 6, 2015: http://nces.ed.gov/programs/digest/d13/tables/dt13_306.10.asp.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*, 440-457.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihsltrom, J. F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage*, *28*, 797-814.

Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin*, *29*(6), 790-801.

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*, *93*, 377-385.

Rosen, S., & Tesser, A. (1970). On reluctance to communicate undesirable information: The MUM effect. *Sociometry,* 253-263.

Sasaki, S. J., & Vorauer, J. D. (2010). Contagious resource depletion and anxiety? Spreading effects of evaluative concern and impression formation in dyadic social interaction. *Journal of Experimental Social Psychology*, *46*, 1011-1016.

Seger, C. A., Stone, M., and Keenan, J. P. (2004). Cortical activations during judgments about the self and an other person. *Neuropsychologia*, *42,* 1168–1177.

Shi, Z., Zhou, A., Liu, P., Zhang, P., & Han, W. (2011). An EEG study on the effect of selfrelevant possessive pronoun: self-referential content and first-person perspective. *Neuroscience letters, 494,* 174-179.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *The Journal of Neuroscience, 33*(19), 8264-8269.

Tesser, A., & Rosen, S. (1972). Similarity of objective fate as a determinant of the reluctance to transmit unpleasant information: The MUM effect. *Journal of Personality and Social Psychology, 23*, 46.

Toosi, N. R., Babbitt, L. G., Ambady, N., & Sommers, S. R. (2012). Dyadic interracial interactions: a meta-analysis. *Psychological Bulletin*, *138*, 1-27.

Ulrey, K. L., & Amason, P. (2001). Intercultural communication between patients and health care providers: An exploration of intercultural communication effectiveness, cultural sensitivity, stress, and anxiety. *Health Communication, 13*, 449-463.

Verduyn, P., Van Mechelen, I., Kross, E., Chezzi, C., & Van Bever, F. (2012). The relationship between self-distancing and the duration of negative and positive emotional experiences in daily life. *Emotion*, *12*(6), 1248-1263.

Vescio, T. K., Gervais, S. J., Snyder, M., & Hoover, A. (2005). Power and the creation of patronizing environments: the stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *Journal of Personality and Social Psychology*, *88*(4), 658-672.

Vidal, F., Burle, B., Grapperon, J., & Hasbroucq, T. (2011). An ERP study of cognitive architecture and the insertion of mental processes: Donders revisited. *Psychophysiology*, *48*, 1242-1251.

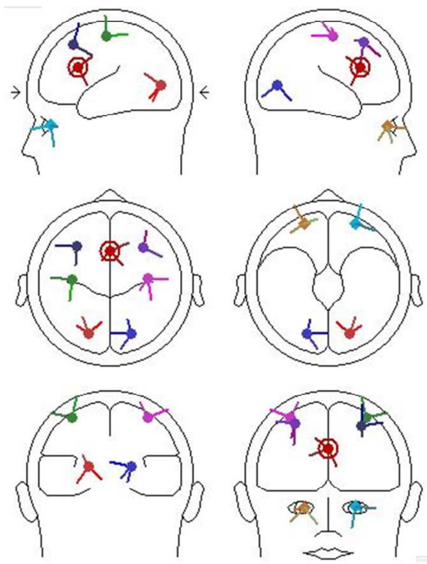
Vorauer, J. D. (2006). An information search model of evaluative concerns in intergroup interaction. *Psychological Review*, *113*, 862.

Vorauer, J. D., & Turpie, C. A. (2004). Disruptive effects of vigilance on dominant group members' treatment of outgroup members: choking versus shining under pressure. *Journal of Personality and Social Psychology*, *87*, 384-389.

Woody, S. R., & Rodriguez, B. F. (2000). Self-focused attention and social anxiety in social phobics and normal controls. *Cognitive Therapy and Research, 24*(4), 473-488.

Woolf, K., Cave, J., Greenhalgh, T., & Dacre, J. (2008). Ethnic stereotypes and the underachievement of UK medical students from ethnic minorities: Qualitative study. *British Medical Journal*, 337, a1220.

A



B

Sources	x	y	z
Left Eye	-26	60	-27
Right Eye	23	61	-27
Left MC	-36	-19	58
Right MC	36	19	58
Left OC	-21	-79	1
Right OC	21	79	1
MPFC	0	19	30
Left LPFC	-32	22	55
Right LPFC	32	22	55

Figure 1. A: Dipole source model. B: Talairach coordinates
Fig. 1

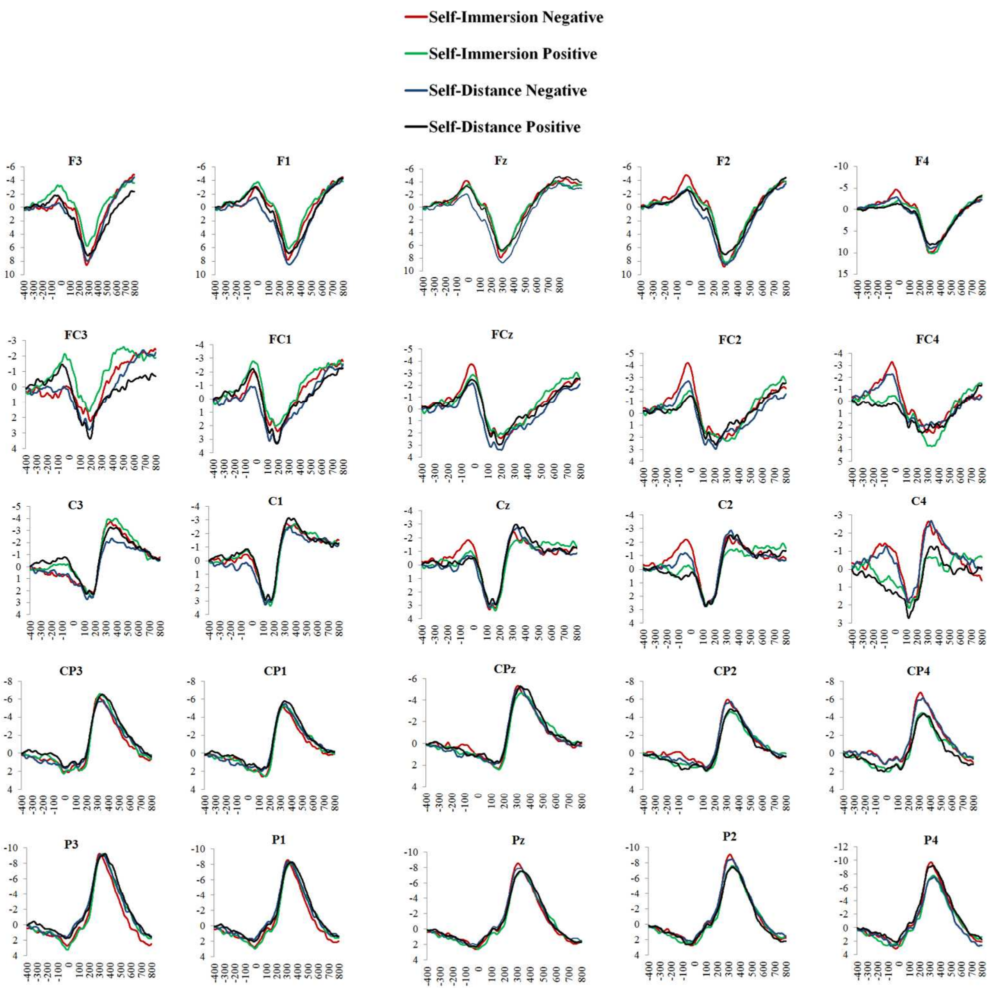


Figure 2. ERP waveforms.
Fig. 2

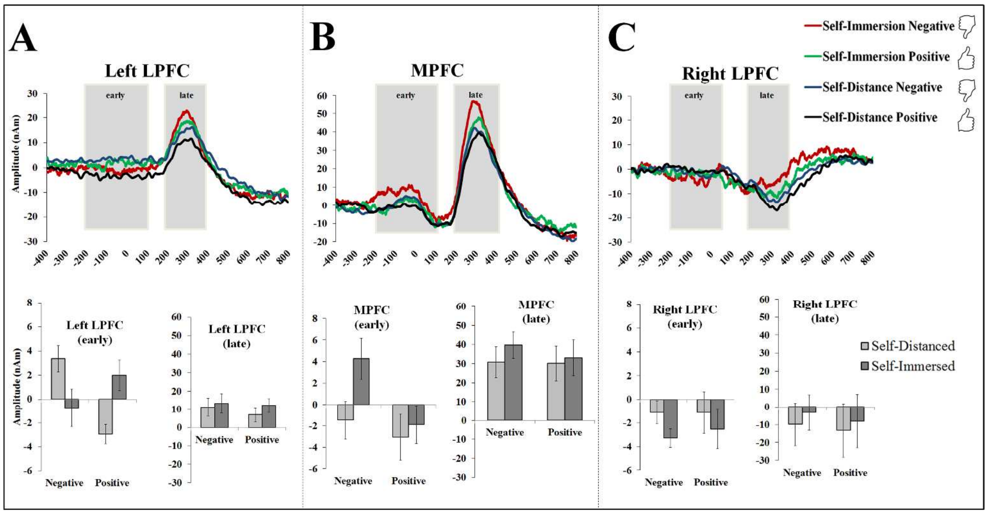


Figure 3. Current source density waveforms of frontal sources: left LPFC (A), MPFC (B), and right LPFC (C). Zero point on top panel represents participants' button-press for conveying feedback. Bottom panel depicts the mean amplitude during early (-200 to 100 ms) and late (200 to 400 ms) time windows. LPFC=lateral prefrontal cortex. MPFC=medial prefrontal cortex. Error bars represent standard error of the mean.
Fig. 3

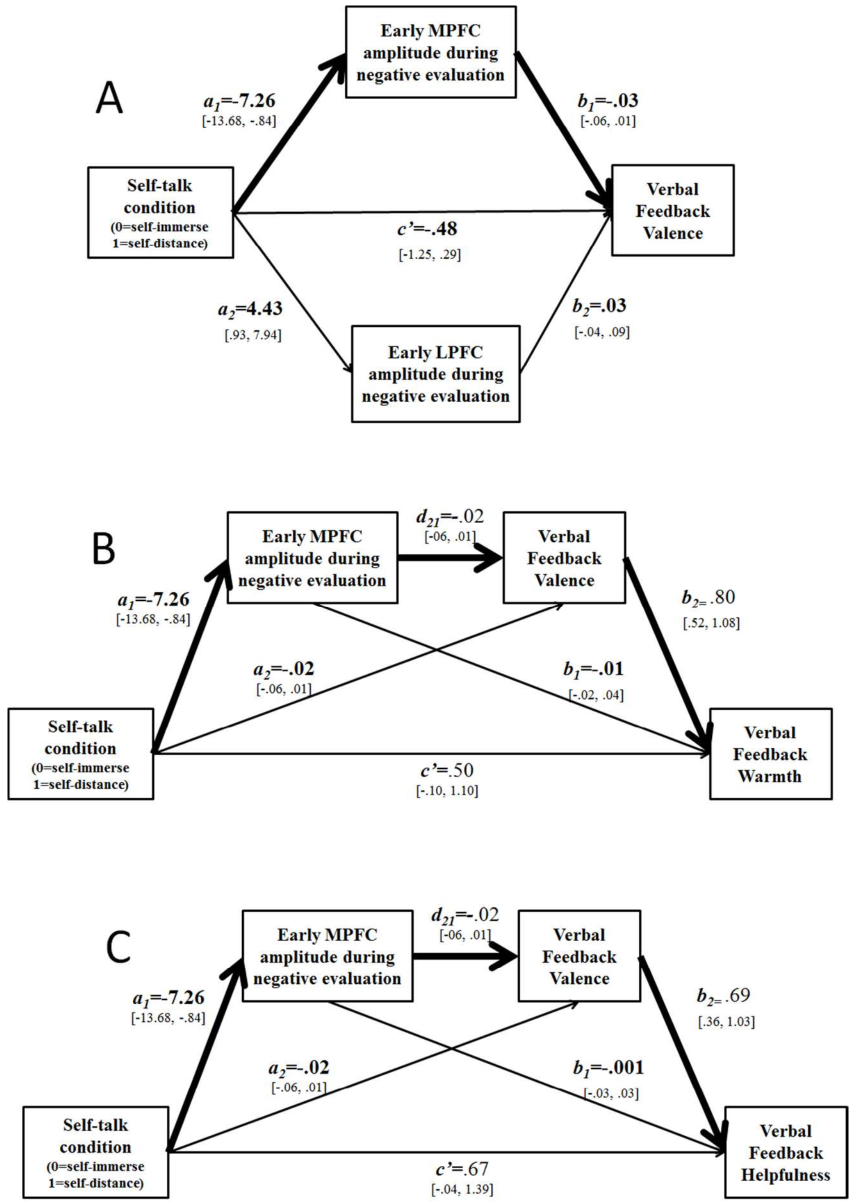


Figure 4. Indirect effects. Values represent unstandardized coefficients, as is convention in path modeling. Bold lines indicate paths of significant indirect effects. Thin lines indicate paths in non-significant indirect effects. The significance of indirect effects is determined by the product of the simple paths that comprise it (Hayes, 2013). 95% CIs of each simple path are indicated in brackets. (A) Parallel mediation predicting verbal feedback valence, (B) serial mediation predicting verbal feedback warmth, (C) serial mediation predicting verbal feedback helpfulness.

Fig. 4