The Faces of Race-Bias: Awareness of Racial Cues Moderates the Relation Between Bias and In-Group Facial Mimicry

Dana R. Carney, Kristina R. Olson, Mahzarin R. Banaji, & Wendy Berry Mendes

Harvard University

Report: 5000 words or less (not incl abst, refs, tables, figures but INCL footnotes and ackn). No more than 40 refs,

Curr at 2355

Abstract

In a study investigating the influence of external cues such as others' expectations and societal norms (i.e., "correction cues") on electromyography (EMG) measured smiling to Black and White smiling faces, White participants viewed Black and White faces while smiling behavior (i.e., zygomaticus major activity) was measured. When the faces were presented subliminally, for 17 ms, implicit pro-White bias predicted smiling more at Whites. When faces were presented supraliminally, for 4000 ms, explicit pro-White bias predicted a contrast effect in which Whites smiled more at Blacks. These results provide evidence to suggest that "correction cues" trigger corrective processes that work against held attitude, can be miscalibrated and expressed in behavior that is proportionally inversely related to the held attitude. Further, this study is the first evidence to suggest that facial EMG can measure facial mimicry in an event-related paradigm.

The Faces of Race-Bias: Awareness of Racial Cues Moderates the Relation Between Bias and In-Group Facial Mimicry

The human smile conveys great meaning in a single, simple, pulse. Each smile signals affiliation, safety, and warmth. Sometimes smiles are elicited automatically and other times are deliberately crafted to adaptively navigate through everyday social encounters. Internal cues such as happiness or a preference for a person or group can unknowingly elicit a smile. Likewise, external cues such as others' expectations and societal norms (i.e., "correction cues") can force us to consciously correct against true feelings in the service of a smile. The goal of the current research was to test whether the presence of external "correction cues" would modulate the influence of White preference on smiling toward Blacks and Whites.

Preference for Whites over Blacks is often expressed as antipathy—behaving more negatively toward Blacks relative to Whites. However, both Gordon Allport and Kurt Lewin told us long ago that behavior is not only a function of internal characteristics such as race-bias, but is also influenced by the environments which press upon us as we behaviorally express those characteristics. Both Fazio's Motivation and Opportunity as Determinants Model (1990) and Wegener and Petty's flexible correction model (1995) suggest that when "correction cues" are present, behavior and perception are adjusted accordingly. Thus, it ought to come as no great surprise to find that in a world in which biased behavior is socially unacceptable, the expression of bias as negative behavior is often flexibly corrected—sometimes *over*corrected—into positivity. Indeed, currently emerging work suggests that when motivated by a "correction cue," such as an obvious societal norm, Whites' bias is corrected. For example, it has been shown that White preference predicts stronger amygdala activity in response to viewing Black relative to White faces, and such activity is modulated by areas of the prefrontal cortex recruited in affect regulation (i.e., ventrolateral and dorsolateral prefrontal cortex regions; Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004). However, likely lacking an accurate correction system, Whites sometimes appear to *over*correct for their White preference by expressing it as apparent positivity (Carney, 2004; Shelton, Richeson, Salvatore, & Trawalter, 2005). Each of these studies in which evidence of correction or overcorrection emerged allowed Whites to be consciously aware of the Black stimuli—either faces presented on a computer screen or a live person. Conscious awareness of the Blacks was likely to have reminded Whites that prejudiced behavior was unacceptable which led to bias predicting overcorrection. In contrast, White preference predicted unmodulated amygdala activity when Whites were not consciously aware of the Black faces they were seeing (Cunningham et al.). Although Cunningham et al. manipulated White participants' conscious awareness of the Black and White stimuli, the outcome measures of brain activation did not allow for the disambiguation of correction from overcorrection. Manipulating awareness and measuring a behavior such as smiling would allow for the differentiation of corrective versus overcorrective processes. Indeed, research has shown that one of the strongest indicators of overcorrection is smiling (Carney, 2004). Smiling, like other lower-face nonverbal behaviors, is fairly controllable (Cacioppo, Bush, & Tassinary, 1992).

In the current research, we predicted that consciously accessible "correction cues" would direct Whites as to whether or not to flexibly correct behavioral expressions of bias. Specifically, when "correction cues" were consciously unavailable, we expected that Whites' preference for other Whites would predict more smiling at Whites. In contrast, we expected that when "correction cues" were consciously available, and White preference also available to consciousness would be overcorrected resulting in relatively more smiling at Blacks.

To test this hypothesis, we measured smiling behavior with facial EMG in response to the presentation of standardized photographs of smiling Blacks and Whites. Two photograph presentation conditions (within-participants) were used: subliminal (17 ms; no "correction cues" consciously available), and supraliminal (4000 ms; "correction cues" consciously available). Participants also completed measures of implicit and explicit race-bias.

Method

Participants and Design

Twenty-three participants from Harvard University were recruited (2 were Black and were excluded from all analyses leaving 21). The experiment took place in a sound-attenuated laboratory and participants sat in a large, comfortable recliner for the duration of the experiment.

Participants were told that the study was ostensibly called "attention and memory" in which we were interested in measuring electrical activity during picture-viewing. Facial EMG sensors were attached to the right side of the face to three muscle regions. In addition to measuring activity in the zygomaticus major ("smile" muscle that pulls the mouth laterally and slightly upward), additional EMG sensors were applied to the corrugator supercilii and obicularis oculi regions, and the middle of forehead (ground) in order to help with the cover story. Activity from the corrugator and obicularis oculi were not relevant to the current investigation and these data are not discussed.

After the EMG sensors were affixed, participants took all measures used in this study (i.e., the design was fully within-participants). Because the subliminal task required that no "correction cues" were consciously available, it was always given first and the supraliminal task was always given second. Both the subliminal and supraliminal tasks used the same photographs of smiling Black and White females and males. Then, participants completed a subliminal threshold sensitivity task (to gauge whether participants' perceptual threshold was below the ms unit used for the subliminal presentation), and measures of implicit (2 randomly ordered measures) and explicit (2 randomly ordered measures) Black/White race-bias.

Subliminal task. The subliminal task was always presented first. Smiling faces of Blacks (n = 8) and Whites (n = 8; equal numbers of females and males; NimStim: Tottenham, Borscheid, Ellertsen, Marcus, Nelson, 2002) were embedded in 48 additional stimuli including non-smiling faces (n = 32) and neutral non-face objects (n = 16) taken from the IAPS (Lang, Bradley, & Cuthbert, 1998).[1] All trials began with a 1000 ms fixation cross in the middle of the screen. Stimuli were shown for 17 ms and sandwich-masked between semi face-looking abstract pictures used in previous research by Cunningham et al. (2004). Masks were shown for 200 ms each. Each trial ended with a blank dark-colored screen which was shown for 3000 ms to allow facial muscle recovery time after each trial (SOA = 4417 ms). Participants were instructed to pay close attention to the images (i.e., masks were perceptible).

Supraliminal task. The supraliminal task was always presented second. All trials began with a 1000 ms fixation cross. The same stimuli used in the subliminal were used in the supraliminal task except the stimuli were unmasked and shown for 4000 ms and were also followed by a 3000 ms blank recovery screen (SOA = 8000 ms). Participants were asked to pay close attention to the images.

Sensitivity task. Following the subliminal and supraliminal tasks, a sensitivity task was administered to determine whether the subliminally presented stimuli were, indeed, below each individual participant's perceptual threshold. All face stimuli (n = 48) were individually presented using the same procedure as the subliminal task. Following each individual trial, participants made a simple judgment of whether the face was Black or White. By setting the sensitivity criterion so low with an easy discrimination task, we could be sure to exclude any participants for whom the 17 ms presentation was not entirely subliminal. Because there were an equal number of Black and White faces shown, chance level of the Black-White judgment was 50%. Participants' average sensitivity was at chance, 50%, and ranged from 31 – 69% with a standard deviation of 8%. There were only 2 subjects who scored above chance (67 and 69%). Removal of these subjects did not change any effects; therefore all subjects were retained in all analyses.

Implicit measures of race-bias. The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) was used to measure degree of implicit race-bias against Blacks. The IAT provides a measure of the strength of association between mental categories such as “Black” or “White” on the one hand and attributes such as “cold” or “warm” on the other. A race IAT of this type functions as a measure of implicit attitude. It measures strength of association between category and attribute by using the time it takes to make the pairings, and the number of errors in classifying, while respondents are trying to respond rapidly. The strength of association between categories and attributes such as *warm* and *cold* provides an attitudinal measure, and with attributes such as *self* and *other*, provides a measure of implicit affiliation of self to Blacks and Whites.

Two IATs were used in this research.[2] Procedure for IAT construction (e.g., 7-block design) and scoring procedures (e.g., the IAT “D”” score was calculated) followed Greenwald, Nosek, and Banaji (2003; additionally, order of congruent and incongruent blocks was counterbalanced across participants). Both IATs used stock photos of Black and White faces. The two IATs used were: (1) Self / Other – Black / White (“self” items: we, us, our; “other” items: they, them, their), and (2) Warm / Cold – Black / White (“warm” items: warm, compassionate, kind; “cold” items: cold, distant, reserved). Order in which participants received the two IATs was randomized across participants. The average of the two IATs was taken to form a composite IAT measure with a mean D-score of .14 and a standard deviation of .25, which was significantly different from the scale midpoint of zero, $t = 2.48$, $p < .03$ showing participants, on average, preferred Whites to Blacks.

Explicit measures of race-bias. Two explicit (i.e., self-report) race-bias measures, analogous in content to the two IATs were administered in a random order (explicit tasks always followed implicit tasks). Participants indicated on a 7-point scale whether Blacks were perceived as more or less cold as Whites, and more or less affiliated with themselves as Whites. For example, the explicit warm / cold item ranged from “I feel strongly that White people are more interpersonally warm than Black people” to “I feel strongly that Black people are more interpersonally warm than White people.” Likewise, the explicit self / other items began, “I strongly feel that White people are more ‘one of us’ than Black people.” Intermediate attitude anchors were “moderately” and “slightly” and the mid-point was, “I feel that White and Black people are equally ‘one of us’ [interpersonally warm].” The explicit measures were averaged to form a composite measure with a mean score of 4.00 and a standard deviation of .89, which was not significantly different from 4.00, the midpoint of the 7-point scale ($p = 1.00$) showing participants, on average, did not explicitly report a preference for Blacks or Whites.

Facial EMG. Skin was cleaned and abraded and then a female experimenter attached 4mm paired Ag-AgCl electrodes placed on zygomaticus muscle regions to record surface EMG activity. Locations of sensor placement were guided by previous recommendations and guidelines (Fridlund & Cacioppo, 1986). EMG signals were relayed through shielded cable to

Biopac EMG amplifiers (Santa Barbara, CA), where signals were amplified 5,000x. Signals were digitized at 2000 Hz, then recorded and displayed on a computer in a separate control room.

Offline, the data were examined and scored using MindWare EMG 2.1 scoring module, with low and high filter cutoffs set to 3Hz and 30Hz, respectively. We examined EMG activity using the event-related trigger option, which allowed us to determine the temporal windows in which the muscle activity would be assessed. For both subliminal and supraliminal data we used the same time window (5s after stimulus onset) so comparable contrasts could be made across the presentation conditions. The scoring program produces filtered and rectified EMG, and the spectral output. Zygomaticus peak amplitude was used for all subsequent analysis and was *z*-scored within subject. We calculated EMG reactivity as the difference between EMG activity when presented with White faces minus EMG activity when presented with Black faces, so higher scores indicate greater zygomaticus region activity to White faces.

Results

Consistent with our hypothesis, Figure 1 shows that implicit race-bias against Blacks predicted more smiling toward *Whites* in the subliminal (17 ms) condition when “correction cues” were not consciously available. In contrast, when “correction cues” were consciously available in the supraliminal (4000 ms) condition, explicit white preference predicted more smiling toward *Blacks*, $z = 2.41$, $p < .02$ (test for difference between correlated correlations; Meng, Rosenthal, & Rubin, 1992). Specifically, in the subliminal condition in which no “correction cues” were present, implicit (but not explicit) White preference was expressed through more smiling to White and less smiling to Black smiling faces, $r(19) = .41$, $p < .08$. In contrast, when correction cues were present, explicit (but not implicit) White preference predicted more smiling toward Black and less toward White, $r(19) = -.38$, $p < .09$.[3]

Discussion

These data suggest that consciously available social cues can trigger rules of conduct that then modulate behavioral responses. Further, these data suggest that the correction processes engaged are proportionately opposite to held belief. Finally, these data suggest that this correction system, although flexible as Fazio (1990) and Wegener and Petty (1995) suggest, miscalibrates in its correction leading to *over*correction. When no correction cues were present, and both the attitude and stimuli resided outside of conscious awareness, Whites predictably smiled back more at other Whites than they did at Blacks. This finding is not only theoretically consistent, but is empirically consistent with previous work on the relation between an unobtrusive race-bias measure (Black and White faces were subliminally presented) and smiling toward a Black experimenter (Dovidio, Kawakami, & Gaertner, 2002). In contrast, and also predictably, When correction cues were present, and stimulus faces were clearly available to consciousness, self-reported pro-White bias (but not implicit bias, which is less available to consciousness) was overcorrected and expressed as smiling more toward Blacks.

An important methodological side note is that this study is the first to demonstrate what appears to be facial mimicry to subliminally presented faces using an event-related design (whereas past research showing EMG-measured mimicry in response to subliminally presented faces employed a block design; Dimberg, Thunberg, & Elmehed, 2000).

In summary, these findings extend theoretical models of corrective processes (Fazio, 1990; Wegener & Petty, 1995) by suggesting that when the outcome measure is a behavior such as a smile, the corrective system is not likely to be accurately calibrated on average, which can result in ovecorrection.

References (check again before submission)

Cacioppo, J. T., Bush, L. K., & Tassinary, L.G. (1992). Microexpressive facial actions as a function of affective stimuli: Replication and extension. *Personality and Social Psychology Bulletin*, *18*, 515-526.

Carney, D. R. (2004). In the face of prejudice: Implicitly and explicitly measured race-bias and interracial interaction. Unpublished doctoral dissertation, Northeastern Univesity, Boston, MA.

Cunningham, W.A., Johnson, M.K., Raye, C.L., Gatenby, J.C., Gore, J.C., & Banaji, M.R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806-813.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science, 11*, 86-89.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. (Vol. 23, pp. 75-109). New York: Academic Press.

Fridlund, A. J., & Cacioppo, J. T. (1986). Publication guidelines for human electromyographicresearch. *Psychophysiology, 23*, 567-589.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit social cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.

Lang, P. J., Bradley, M. M. & Cuthbert, B. N. (1998). *The International Affective Picture System: Technical Report and affective ratings*. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172-175.

Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science, 16*, 397-402.

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D. J., & Nelson, C. A. (2002). Categorization of facial expressions in children and adults: Establishing a larger stimulus set. *Journal of Cognitive Neuroscience, S74*.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68*, 36-51.

Footnotes

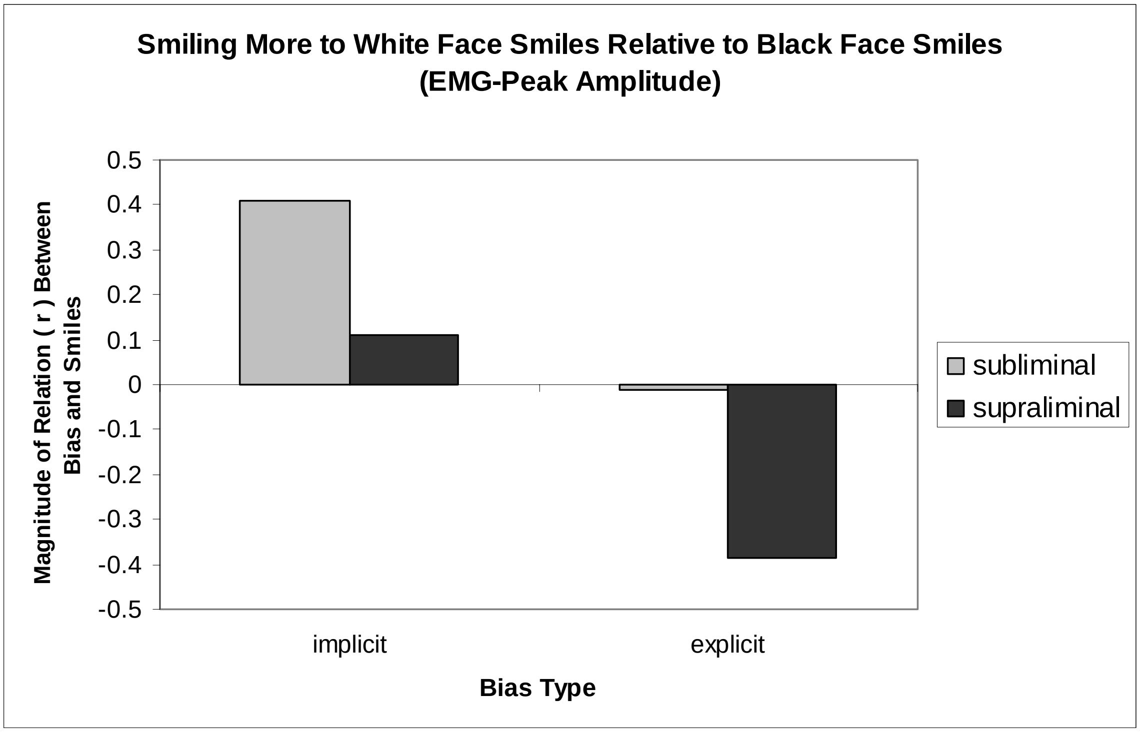
[1] Neutral (n = 16) and angry (n = 16) Black and White female and male stimulus faces were also presented as filler pictures.

[2] An additional IAT, measuring Black-White / Approach-Avoid, was developed for exploratory purposes and was determine *a prori* to contain poor items, thus, no results from this IAT were reported.

[3] Regression analyses showed that the unique contribution of each implicit and explicit race-bias to EMG activity within condition were also marginally significant (i.e., controlling for one method of measuring bias did not attenuate the effects of the other method).

Figure Caption

Figure 1. The magnitude of the relation between race-bias (implicit and explicit) and smiling behavior (peak amplitude of zygomaticus major activity) during two stimulus presentation conditions (subliminal and supraliminal). Values above the zero-point on the y-axis represent more smiling toward White relative to Black (a difference score) and negative values represent the opposite.


Smiling More to White Face Smiles Relative to Black Face Smiles
(EMG-Peak Amplitude)
Magnitude of Relation (r) Between Bias and Smiles
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
implicit
explicit
Bias Type
subliminal
supraliminal