NeuroImage 101 (2014) 704–711

Contents lists available at ScienceDirect

NeuroImage

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

The neural basis of stereotypic impact on multiple social categorization

Eric Hehman [a,b,*], Zachary A. Ingbretsen [b,c], Jonathan B. Freeman [a,b,*]

[a] New York University, USA
[b] Dartmouth College, USA
[c] Harvard University, USA

ARTICLE INFO

Article history:
Accepted 25 July 2014
Available online 3 August 2014

Keywords:
fMRI
ACC
mPFC
dlPFC
Stereotypes
Mouse-tracking

ABSTRACT

Perceivers extract multiple social dimensions from another's face (e.g., race, emotion), and these dimensions can become linked due to stereotypes (e.g., Black individuals → angry). The current research examined the neural basis of detecting and resolving conflicts between top-down stereotypes and bottom-up visual information in person perception. Participants viewed faces congruent and incongruent with stereotypes, via variations in race and emotion, while neural activity was measured using fMRI. Hand movements en route to race/emotion responses were recorded using mouse-tracking to behaviorally index individual differences in stereotypical associations during categorization. The medial prefrontal cortex (mPFC) and anterior cingulate cortex (ACC) showed stronger activation to faces that violated stereotypical expectancies at the intersection of multiple social categories (i.e., race and emotion). These regions were highly sensitive to the degree of incongruency, exhibiting linearly increasing responses as race and emotion became stereotypically more incongruent. Further, the ACC exhibited greater functional connectivity with the lateral fusiform cortex, a region implicated in face processing, when viewing stereotypically incongruent (relative to congruent) targets. Finally, participants with stronger behavioral tendencies to link race and emotion stereotypically during categorization showed greater dorsolateral prefrontal cortex activation to stereotypically incongruent targets. Together, the findings provide insight into how conflicting stereotypes at the nexus of multiple social dimensions are resolved at the neural level to accurately perceive other people.

Introduction

Individuals rely on stereotypes as cognitive shortcuts to successfully navigate a complex social world. These shortcuts sort and categorize information so that it is processed more efficiently, but in turn influence thought and decision-making about other people in a manner prone to error (Cuddy et al., 2007). Thus, by economizing on mental resources stereotypes generally afford benefits to individuals, but often unfairly disadvantage the targets of stereotypes and are overgeneralized to targets for which they do not apply. For example, individuals more likely to rely upon the association between Black males and hostility in North America (Devine, 1989) are more likely to perceive an emotionally ambiguous Black face as angry, and an angry, racially ambiguous face as Black (Hugenberg and Bodenhausen, 2003, 2004). Thus, individuals may be anxious to interact with (Dovidio et al., 2002) or hire (Dovidio and Gaertner, 2000) a non-hostile, friendly Black male. Naturally, perceivers readily encounter targets who counter stereotypes, such as a friendly, Black male, and must integrate incongruent emotional (e.g., smiling) and racial (e.g., Black) information in order to perceive these individuals accurately. The goal of the current research was to examine the neural mechanisms involved in integrating multiple facial cues, considered congruent or incongruent due to stereotypes, when perceiving others at the nexus of multiple social dimensions.

A great deal of research has revealed the mechanisms by which stereotypes may influence behavior (Cuddy et al., 2007; Devine, 1989; Fiske et al., 2002). It has long been known, for example, that facial cues (e.g., wide nose) trigger social categories (e.g., Black), which thereafter activate associated stereotypes (e.g., hostile; Macrae and Bodenhausen, 2000). However, recent work suggests that activation between social categories and stereotypes may be reciprocal. Specifically, current models of person perception posit that social categorizations are the end-result of lower-level face processing and higher-order social cognition mutually constraining one another over time. During this process, categories and stereotypes dynamically form a "compromise," as they pass activation back-and-forth (Freeman and Ambady, 2011). Critically, this suggests that activated stereotypes may sometimes considerably impact the basic perception of social categories.

For example, the stereotype of "violent" is equally associated with social categories Black and male, just as "family-oriented" is equally shared between the Asian and female categories. Accordingly, participants were faster to categorize Blacks as male and Asians as female, but slower to categorize Blacks as female and Asians as male, presumably due to race-triggered stereotypes (e.g., Black → aggressive) facilitating

* Corresponding authors at: Department of Psychology, New York University, USA.
E-mail addresses: ericlhehman@gmail.com (E. Hehman), jon.freeman@nyu.edu (J.B. Freeman).

http://dx.doi.org/10.1016/j.neuroimage.2014.07.056
1053-8119/

the activation of sex categories (e.g., aggressive → male) (Johnson et al., 2012). Similar findings of stereotypes altering basic perceptions have been observed at the crossroads of race and emotion. For example, both Blacks and angry individuals are stereotypically associated with hostility, and White and happy individuals with non-hostility (Devine, 1989), and previous studies have found that participants with greater racial prejudice are more likely to perceive an emotionally ambiguous Black face as angry, and a racially ambiguous angry face as Black (Hugenberg and Bodenhausen, 2003, 2004).

Recent research harnessing methodological advances in real-time behavioral techniques such as mouse-tracking provides a useful example of how stereotype associations can bias the basic categorization of faces (Johnson et al., 2012). Asian and Black, male and female faces appeared at the bottom center of a computer screen, and the trajectory of the mouse was recorded while participants categorized each target's gender by clicking a "male" or "female" response at the top left and right corners of the screen. This technique has previously been shown to capture the extent to which multiple social categories are simultaneously activated during the process of categorization, via the mouse trajectory's curvature toward the opposite category (e.g., Dale et al., 2007; Freeman et al., 2011). In support of the hypothesized overlap in stereotype content between Black male and Asian female, participants deviated more toward the male category when categorizing Black targets, and more toward the female category when categorizing Asian targets. Most importantly, individuals harboring stronger Black male and Asian female stereotypical associations exhibited greater mouse-trajectory deviations toward the category response congruent with these stereotypes (Johnson et al., 2012; see also Galinsky et al., 2013). Thus, an incidental overlap in stereotypes between multiple category dimensions (e.g., Black, male, and anger associations with hostility) can alter the perception of each dimension. Further, because individuals vary in the strength of their associations between social categories and associated stereotype content, they also vary in the degree to which stereotypes influence perception.

This prior work highlights an important conundrum. Because social targets are always categorizable along multiple dimensions (e.g., race, emotion), the stereotypes tied to a target's multiple categories will sometimes agree and sometimes conflict. Thus, in many encounters, individuals must resolve conflicting stereotypes to allow for veridical perceptions of others. If, for instance, a happy Black face triggers stereotypes of hostility (due to associations with Black) that in turn facilitate an interpretation of anger, the conflict must be resolved between the stereotype-based, top-down-driven interpretation of "anger" and the cue-based, bottom-up-driven interpretation of "happy." Thus, faces may often trigger conflicting interpretations (one via stereotypes and one via facial cues) that must be rapidly resolved to accurately perceive others (Freeman and Ambady, 2011; Johnson et al., 2012).

Although numerous studies have demonstrated clear behavioral evidence of stereotype-driven conflict due to multiple social categories, how this conflict is detected and resolved at the neural level remains unclear. Several regions repeatedly shown to be important for social cognition and stereotyping would be likely to be involved, such as the medial prefrontal cortex (mPFC) and anterior cingulate cortex (ACC). Greater activation is observed in the mPFC during tasks that would access stereotypic content, and when stereotypes govern judgments (Knutson et al., 2007; Mitchell et al., 2009, 2005; Quadflieg et al., 2009). Research utilizing diverse techniques (Beer et al., 2008; Gehring et al., 1993; Niki and Watanabe, 1979; Pardo et al., 1990) has long converged on the ACC's role in conflict monitoring, as it is readily activated by situations during which multiple responses compete (Botvinick et al., 2001). However, it is unclear whether the mPFC and ACC are sensitive to conflicts between facial cues and stereotypic expectations regarding those cues, which spontaneously arise from perceiving another's multiple category memberships. If true, it would suggest that these regions may have an important role in resolving the natural inconsistencies triggered by faces due to their multiple social categories. This could potentially aid in the ability to perceive others' faces in a veridical fashion (e.g., to perceive a happy Black male face as happy, rather than angry).

The current research

In the present study, we examined neural responses to faces whose multiple categories varied in stereotypic congruency by specifically focusing on two category dimensions, race and emotion. Drawing on the association between Black males and anger (Devine, 1989), and consistent with prior work (Hugenberg and Bodenhausen, 2003, 2004), stereotypic congruency would increase as a Black face becomes angrier or a White face becomes happier and would decrease as a Black face becomes happier or a White face becomes angrier. Participants viewed faces independently varying on race and emotion while blood-oxygenation-level-dependent (BOLD) responses were measured using functional magnetic resonance imaging (fMRI). Specifically, participants were presented with Black, racially-ambiguous, and White faces displaying angry, emotionally-ambiguous, and happy expressions so that sensitivity to incongruency could be examined in a linear, graded manner. We were interested in BOLD responses that might be sensitive to the stereotypic congruency of a face's race and emotion, and how such responses might be moderated by the strength of an individual's stereotypic associations between these category dimensions.

To address the latter question, following scanning participants completed a behavioral mouse-tracking task, during which they categorized a subset of the faces presented in the scanner by race and emotion. This methodology provided a highly sensitive measure of the associative strength between two social category dimensions such as race and emotion. Its demonstrated millisecond-level detection of social category competition during categorization tasks is ideal for examining conflicts between bottom-up facial cues and top-down stereotype-driven responses (Freeman et al., 2011; Johnson et al., 2012).

Methods

Participants

Twenty-three healthy participants participated for partial course credit. Participants were excluded from analysis for head movement exceeding 3 mm during scanning ($n = 1$), structural abnormalities detected during scanning ($n = 1$), and for falling asleep ($n = 1$), leaving 20 for final analysis (4 male; 12 White, 6 Asian, 1 American Indian, 1 multiracial). All participants were right-handed, aged between 18 and 20 ($M = 18.82$, SD = .80), native English-speakers, with normal or corrected-to-normal vision, and no reported history of neurological disorders or use of psychoactive medications. All participants gave informed consistent in a manner approved by Dartmouth College's Committee for the Protection of Human Subjects and were scanned at the Dartmouth Brain Imaging Center in Hanover, NH, USA.

Stimuli

Face stimuli were created using FaceGen (Singular Inversions, 2012). FaceGen creates 3D digital face models based off laser scans of hundreds of individuals' faces (Blanz and Vetter, 1999). One hundred shape and 100 texture principle components were derived from this entire data set, and digital faces can thus be morphed along these continua. These faces can be morphed along emotion continua, and the structure of the faces are adjusted accordingly.

Computer-generated stimuli allowed us to precisely and independently manipulate race- and emotion-related facial cues while controlling for other facial information. A total of 864 unique faces were created. All targets were male, aged approximately between 20 and 30 years of age, directly oriented, and presented on gray backgrounds. Targets varied along 3 levels of race (White, race-ambiguous, Black)

and 3 levels of emotion (happy, emotion-ambiguous, angry) and were vignetted to display only the face (Fig. 1). Ambiguous targets were included to examine BOLD responses sensitive to the stereotypic congruency of race and emotion cues in a potentially graded, linear manner. Ambiguous race or emotion was generated using a race level of approximately 50% White/50% Black or emotion level of approximately 50% happy/50% angry. Non-ambiguous race or emotion was generated at a level of approximately 100% of the appropriate category. There were 96 stimuli in each of the 3 × 3 within-subject conditions.

Procedure

fMRI paradigm

Participants viewed four functional runs of blocks of faces in pseudorandom order, presented in E-Prime (Psychology Software Tools, Inc., Sharpsburg, PA, USA). Two blocks of each type of stimulus were presented within a run, resulting in a total of 18 blocks per run. Blocks consisted of 12 stimuli from the same category and lasted for 18 s. Each stimulus was presented for 500 ms and preceded by a 1 s fixation cross. Between each block a fixation cross appeared for 6 s, and at the end of each run a fixation cross was presented for 12 s. Fixation epochs served as baseline. Order of block presentation within each run was pseudorandomized, with the exception that in the first run, a block involving ambiguous targets was not presented until after non-ambiguous targets so that participants would not misinterpret the condition. To ensure participants would continually attend to the stimuli, they evaluated every face on an innocuous dimension unrelated to our hypotheses, facial symmetry (i.e., symmetrical or asymmetrical?), via response button. This task ensured participants' continued attention on the stimuli and that any effects of race, emotion, or their stereotypical compatibility would be due to the implicit, rather than explicit, encoding of those dimensions.

Mouse-tracking paradigm

Following scanning, participants categorized faces in a mouse-tracking paradigm. This paradigm has been used in a number of studies to measure the simultaneous activation of multiple competing categories (e.g., Dale et al., 2007; see Freeman et al., 2011, for review), and specifically to measure the associative strength between two social category dimensions (Johnson et al., 2012). Here we make use of its real-time sensitivity to category competition to measure the associative strength between race and emotion. A subset of 30 faces from each condition presented during scanning was again presented. On every trial, participants clicked a "Start" button located at the bottom-center of the screen, which was then replaced by a target face. Participants were tasked with categorizing each face as quickly as possible using responses located in the top left and right corners of the screen. Following 5 practice trials, participants categorized faces as "Black" or "White" in one block of trials, and as "Angry" or "Happy" in another block, for a total of 270 trials (Fig. 2). During this process, we recorded the streaming *x* and *y* coordinates of the mouse (sampling rate = 70 Hz). The MouseTracker software package was used to record and analyze trajectory data (Freeman and Ambady, 2010). The 30 faces of each condition were evenly distributed across blocks. Stimulus order was randomized by participant. Block order and which responses appeared in the top left vs. right of the screen were counterbalanced across participants. Mouse trajectories en route to category responses were recorded for analysis. Finally, participants completed a long-established self-report measure of racial bias, the Attitudes Toward Blacks scale ($\alpha = .84$) (Brigham, 1993), and demographic information.

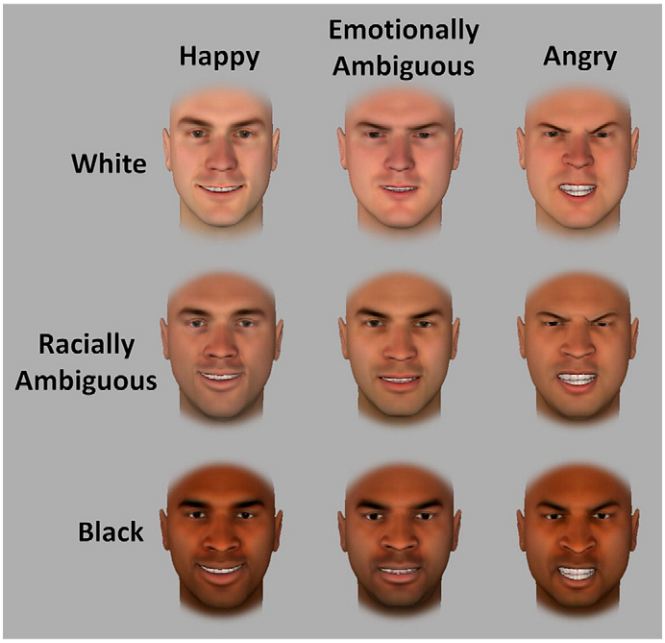

Fig. 1. Example face stimuli. Race and emotion independently varied on 3 levels.

fMRI acquisition

Images were acquired on a Philips Achieva 3.0-T scanner (Philips Medical Systems, Bothell, WA, USA), equipped with a SENSE birdcage head coil. All stimuli were back-projected onto a screen visible via a mirror mounted on the MRI head coil (visual angle ~13.5 × 13.5°). Anatomical images were acquired using a T1-weighted protocol (256 × 256 matrix, 128 1.33 mm transverse slices). Functional images were acquired using a single-shot gradient echo EPI sequence (TR = 2 s, TE = 35 ms). Thirty-five interleaved oblique-axial slices (3 × 3 × 4 mm voxels; slice gap = 0 mm) parallel to the AC–PC line were obtained for each volume.

Data preprocessing

fMRI

Imaging data was processed using BrainVoyagerQX (Brain Innovation, Maastricht, Netherlands). Functional preprocessing included 3D motion correction (trilinear interpolation), slice scan time correction (sinc interpolation), spatial smoothing using a 3D Gaussian filter (8 mm FWHM), and voxelwise linear detrending and high-pass filtering of frequencies (above 3 cycles per time course). Structural and functional data of each participant were transformed to standard Talairach stereotaxic space (Talairach and Tournoux, 1988).

Mouse trajectories

For each trial of the mouse-tracking categorization task, we computed area under the curve (AUC): the geometric area between the observed trajectory and an idealized straight-line trajectory drawn from the start and end points (see Freeman and Ambady, 2010 for further details on mouse-trajectory measures and analytic techniques). For comparison, all trajectories were remapped rightward. Outlier trajectories ±2.5 SD than the mean of all trajectories were removed from analysis (4% of total trajectories). A larger AUC is indicative of greater attraction to the opposite category during categorization.

To assess inter-individual variability in the degree to which participants' stereotypes exerted an influence on the categorization process, AUC difference scores [stereotypically incongruent − stereotypically congruent] were computed. Thus, higher scores indicate that, during emotion categorization, participants deviated more toward the "Angry" response when categorizing a Happy Black face, and toward the "Happy" response when categorizing an Angry White face; and that, during race categorization, participants deviated more toward the "Black" response when categorizing an Angry White face, and toward the "White" response when categorizing a Happy Black face.

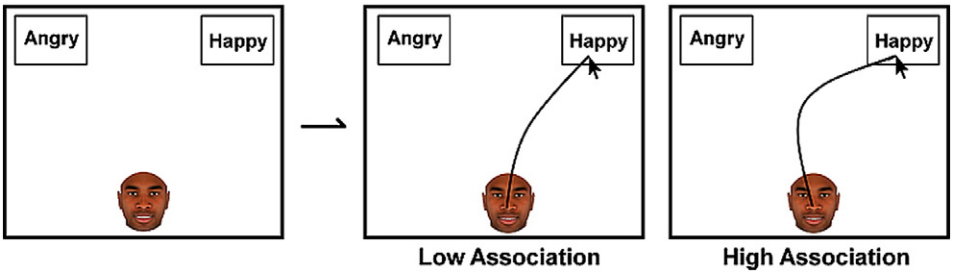

Fig. 2. Schematic illustration of a mouse-tracking trial, displaying mouse-trajectories indicative of low and high associations between race and emotion stereotypes.

These scores therefore index individual participants' stereotypical Black–angry/White–happy associations (also see Johnson et al., 2012).

Data analytic approach

A GLM design matrix was constructed using 9 predictors corresponding to the 9 block conditions (3 race × 3 emotion levels). Conditions were modeled as boxcar functions across block duration and convolved with a two-gamma HRF. First-level GLM analyses conducted on individual participants' fMRI signal were submitted to a second-level random effects analysis, treating participants as a random factor. In all whole-brain analyses, to control for multiple statistical testing we maintained a false positive detection rate of $p < .05$ by using a voxelwise threshold of $p < .005$ and a minimum cluster size of $k \geq 803$ mm^3. The minimum cluster size was empirically determined by a Monte Carlo simulation accounting for spatial correlations between neighboring voxels (Forman et al., 1995).

A whole-brain analysis of variance (ANOVA) testing a 3 (Race: Black, Ambiguous, White) × 3 (Emotion: Angry, Ambiguous, Happy) interaction effect identified regions sensitive to the stereotypic congruency of race and emotion cues. To interpret this interaction, we further probed BOLD responses in these regions, extracting β values for all 9 conditions and analyzing them with follow-up regression analyses. Specifically, β values were regressed on contrast-coded variables representing Race (Black = −1, Race ambiguous = 0, White = 1), Emotion (Angry = −1, Emotion ambiguous = 0, Happy = 1), and their interaction using regression. Simple slopes were decomposed and plotted utilizing techniques specified by Preacher et al. (2006). Given our a priori interest in regions sensitive to stereotypic congruency, we followed this analysis with a more specific whole-brain Incongruent (Black Happy, White Angry) > Congruent (Black Angry, White Happy) contrast, again examining β values in these regions for all 9 conditions using regression analysis. To use similar size regions of the ACC and mPFC for the follow-up regression analyses, we used β values extracted from 20-mm spheres located at the center of the regions elicited by whole-brain analyses. Regression analyses were for descriptive purposes only and for better characterizing the pattern of results in regions elicited by the whole-brain analyses.

In addition, we examined whether stereotypic congruency influenced the functional connectivity between regions by generating connectivity maps for the Incongruent and Congruent conditions. As outlined in Roebroeck et al. (2005), these maps identify voxels whose TR measurements are reliably correlated with simultaneous TR measurements in a given reference region. Specifically, functional connectivity or instantaneous correlation between voxels A and B exists when simultaneous TR measurements A(t) and B(t) improve predictions of the other, while accounting for the past independent time-courses of A and B. To then identify voxels exhibiting greater functional connectivity with relevant regions of interest (ROIs; defined by the whole-brain analyses) when viewing stereotypically incongruent vs. congruent faces, we tested a second-level contrast of the Incongruent > Congruent connectivity maps using a conservative approach examining within, rather than between, subject variation in connectivity, as recommended (Roebroeck et al., 2005).

Finally, we examined the relationship between BOLD responses and mouse trajectories of the behavioral task. A whole-brain analysis of covariance (ANCOVA) was used to identify BOLD responses correlating with the strength of participants' stereotypical associations (trajectory AUC difference scores [Incongruent − Congruent]) during the Incongruent (relative to Congruent) trials.

Results

Neuroimaging

To identify regions sensitive to the stereotypic congruency of race and emotion cues, we conducted a whole-brain ANOVA testing a 3 Race (White, Ambiguous, Black) × Emotion (Happy, Ambiguous, Angry) interaction effect. This revealed activation in the ACC ($p < .05$, corrected; Fig. 3; Table 1). To specify the nature of this interaction, β values for all 9 stimulus conditions were extracted from this region and submitted to regression analysis (see section Data analytic approach for details). As demonstrated in Fig. 3, the ACC exhibited linearly increasing responses as race and emotion became correspondingly more incongruent ($B = -.062$, SE = 0.015, $\beta = -.146$, $t = -4.08$, $p = .001$). Breaking down this interaction, simple slopes revealed that greater ACC activation was displayed as White targets changed from Happy to Angry ($B = -0.079$, SE = 0.024, $t = -3.36$, $p = .003$) and as Black targets changed from Angry to Happy ($B = 0.044$, SE = 0.022, $t = 1.97$, $p = .064$), although this latter effect was marginal. Activation in the ACC did not differ between racially-ambiguous targets of different emotional expressions ($B = -0.018$, SE = 0.017, $t = -1.01$, $p = .323$). Taken together, results indicated that the ACC was sensitive to the stereotypic compatibility of race and emotion cues in a linear, graded manner. No main effects were evident in the ACC.

To more specifically examine neural responses sensitive to stereotypic congruency, we conducted a direct Incongruent (Angry White and Happy Black faces) > Congruent (Angry Black and Happy White faces) whole-brain contrast. Replicating the whole-brain ANOVA, this contrast revealed activation again in the ACC but additionally activation in regions of the mPFC ($p < .05$, corrected; Fig. 4; Table 1). To better specify the pattern of activation in these regions, β values for all 9 stimulus conditions were extracted and submitted to regression analyses.[1] Results revealed a pattern similar to those of the previous analysis, with both the ACC and mPFC exhibiting linearly increasing responses as

[1] We thank an anonymous reviewer for suggesting a complementary parametric analytic approach. A design matrix was created which included parametric predictors for the main effect of race (Black = −1, Race ambiguous = 0, White = 1), the main effect of emotion (Angry = −1, Emotion ambiguous = 0, Happy = 1), and of primary interest, the race × emotion interaction effect (i.e., stereotypic congruency). Positive values of this latter effect therefore indicate congruency and negative values indicate incongruency. We specifically examined the regions of activation elicited by the Incongruent > Congruent whole-brain contrast. Converging with the results reported in the main text, this parametric analysis revealed significant modulation of stereotypic congruency in both the ACC ($M = -.045$, SE = .020, $t = -2.23$, $p = .038$) and mPFC ($M = -.043$, SE = .018, $t = 2.409$, $p = .026$), with these regions showing linearly increasing responses as race and emotion became more stereotypically incongruent.

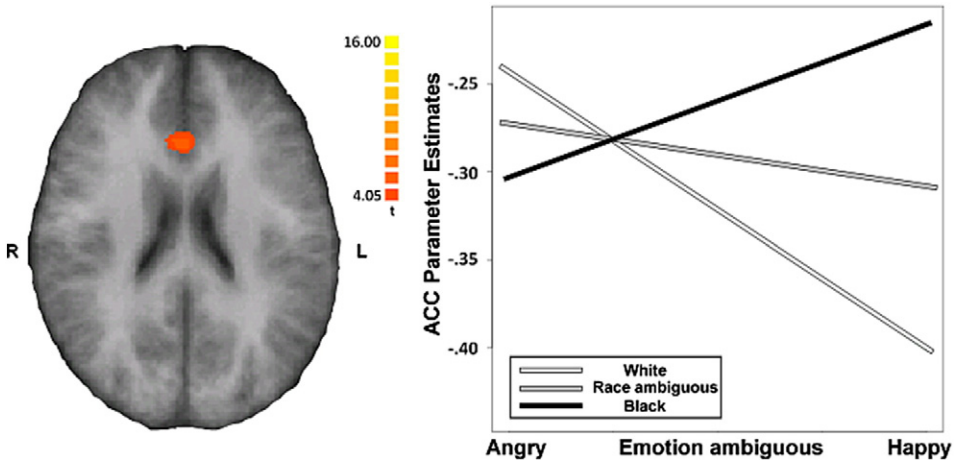

Fig. 3. Whole-brain analysis testing the 3 (Race: Black, Ambiguous, White) × 3 (Emotion: Angry, Ambiguous, Happy) interaction revealing BOLD responses in the ACC ($p < .05$, corrected).

race and emotion become correspondingly more incongruent (Fig. 4). Thus, the ACC and mPFC are highly sensitive to stereotypical violations of racial and emotional facial cues, exhibiting increasing activation in a graded, linear manner as the incongruency between race and emotion increased.[2]

Connectivity analyses

To investigate whether incongruent targets (vs. congruent targets) led to greater instantaneous connectivity between regions, we used the ACC ROI elicited by our whole-brain analysis as a reference in a contrast of Incongruent > Congruent connectivity maps ($p < .05$, corrected; $k \geq 803$ mm^3). Analyses indicated that Incongruent (relative to Congruent) targets elicited stronger simultaneously correlated activity between the ACC and a region implicated in early face processing (Haxby et al., 2000), the lateral fusiform gyrus (FG; BA 37; $x = -40$, $y = -53$, $z = -23$; t = 3.62, k = 1802 mm^3; Fig. 5). These results suggest that viewing stereotypically incongruent (relative to congruent) targets led the ACC to be more functionally coupled with regions associated with the early processing of faces.

Post-scan behavioral data

To examine the possible impact of stereotypical Black–angry/White–happy associations on participants' post-scan categorizations, mouse-trajectory difference scores [Inconguent − Congruent] were analyzed. These scores index the degree to which a participant's trajectories were more attracted to the stereotypically associated race or emotion response while categorizing a stereotypically incongruent (versus congruent) target and, therefore, the extent to which a participant stereotypically associated Black with angry and White with happy. A one-sample t-test comparing these difference scores to 0 indicated that, as a group, participants' categorizations were not overall influenced by Black–angry/White–happy associations, $t(19) = -1.31$, $p = .21$, consistent with previous work (Hugenberg and Bodenhausen, 2003). Differences emerged at the individual level, however. Participants who were more racially biased were more attracted to categorizing targets in a manner consistent with stereotypes, as demonstrated by the significant relationship between [Incongruent −Congruent] AUC difference scores and self-reported racial bias, $\beta = .571$, SE = 0.71, $p = .04$. Thus, these results are consistent with previous work finding that such stereotypic associations particularly impact categorization for individuals with greater racial bias (Hehman et al., 2014; Hugenberg and Bodenhausen, 2003, 2004). They are also consistent with studies finding that mouse-trajectories during social categorization tasks are sensitive to inter-individual variability in the strength of stereotypical associations (Johnson et al., 2012).

Brain–behavior correlations

To examine the relationship between neural responses to stereotypic incongruency and the mouse-tracking measure of stereotypical association strength, a whole-brain ANCOVA identified regions exhibiting a correlation between [Incongruent − Congruent] BOLD contrast values and [Incongruent − Congruent] AUC difference scores ($p < .05$, corrected). Interestingly, this revealed a region of the right dorsolateral prefrontal cortex (dlPFC) often implicated in the inhibition of prepotent responses (Casey et al., 1997; Chee et al., 2000; MacDonald et al., 2000). Responses in the dlPFC to incongruent targets were positively correlated with the strength of stereotypical associations ($r = .64$, $p = .003$; Fig. 6; Table 1). Thus, participants who showed stronger stereotypical associations (Fig. 6d) in the mouse-tracking task also showed stronger dlPFC activation in response to stereotypically incongruent targets than participants with weaker stereotypical associations (Fig. 6c). For example, the degree to which participants' mouse trajectories were attracted toward the "Angry" response while categorizing happy Black

Table 1
Regions of activation elicited by whole-brain analyses ($p < .05$, corrected).

Region	Brodmann	Talairach coordinates				
		x	*y*	*z*	*t*	mm^3
Race × Emotion interaction						
Anterior cingulate cortex	24	2	24	22	4.51	1056
Middle cingulate cortex	24	−15	−18	29	5.28	818
Incongruent > Congruent contrast						
Anterior cingulate cortex	32	2	25	32	3.68	6318
L. Medial prefrontal cortex	6	−20	0	52	3.56	883
L. Medial prefrontal cortex	8	−17	33	43	3.70	1897
L. Anterior insula	13	−35	7	3	3.45	1193
Cerebellum		1	−76	−24	3.75	3256
R. Cerebellum		35	−53	−29	3.49	1703
Correlation with stereotypic association strength	Brodmann	*x*	*y*	*z*	*r*	mm^3
R. Dorsolateral prefrontal cortex	8	38	24	41	.64	1295

[2] We further tested several specific comparisons of BOLD response in the ACC across conditions using repeated measures ANOVA: Angry White vs. Happy Black, ($M = .010$, SD = .239 vs. $M = -.096$, SD = .300), $F(1,19) = .792$, $p = .385$, Angry White vs. Angry Black, ($M = .010$, SD = .239 vs. $M = -.131$, SD = .327), $F(1,19) = 2.044$, $p = .169$, Happy White vs. Angry Black, ($M = -.221$, SD = .239 vs. $M = -.131$, SD = .327), $F(1,19) = 1.244$, $p = .279$, Happy White vs. Happy Black, ($M = -.221$, SD = .239 vs. $M = -.096$, SD = .300), $F(1,19) = 5.852$, $p = .026$. See Supplementary Table 1 for these specific whole-brain contrasts.

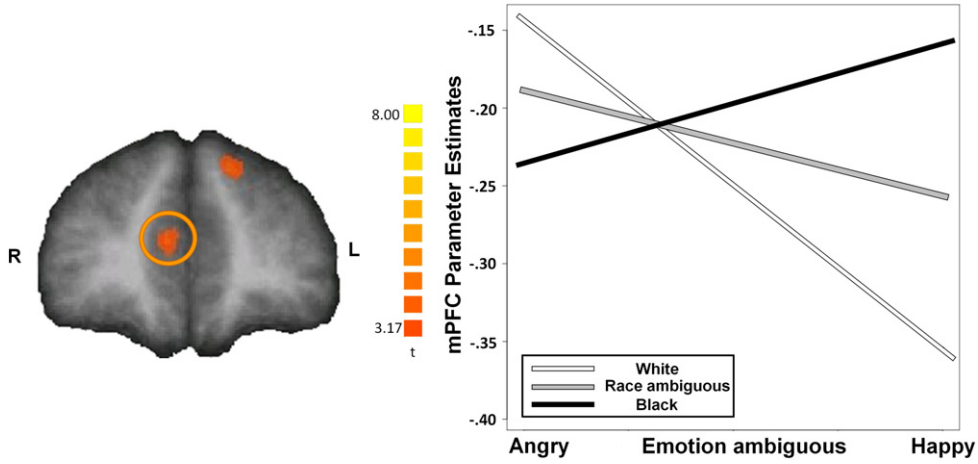

Fig. 4. Whole-brain analysis testing a stereotypically Incongruent > Congruent contrast revealing stronger BOLD responses to incongruent targets in the mPFC ($p < .05$, corrected).

faces (which was ultimately corrected into a "Happy" response) predicted how strongly the dlPFC activated in response to those faces. This correlation was unique to the dlPFC, as it was not present in the ACC ($r = -.03$, $p = .90$) or mPFC ($r = .11$, $p = .64$).

Discussion

Social targets are categorizable along multiple dimensions (e.g., race, emotion), which leads stereotypes linked to these dimensions to sometimes agree and sometimes conflict. Behavioral work has revealed how membership in a social category (e.g., Black) can skew perceptions along other dimensions (e.g., emotion) in a stereotype-consistent manner (Hugenberg and Bodenhausen, 2003, 2004; Johnson et al., 2012), yet the neural mechanisms involved in the detection and resolution of conflict between top-down stereotypes and bottom-up visual cues remain relatively little known. The current research examined neural responses sensitive to routine conflicts between facial cues and stereotypes due to targets' multiple social categories, to elucidate how these conflicts might be resolved and thus facilitating the accurate perception of others.

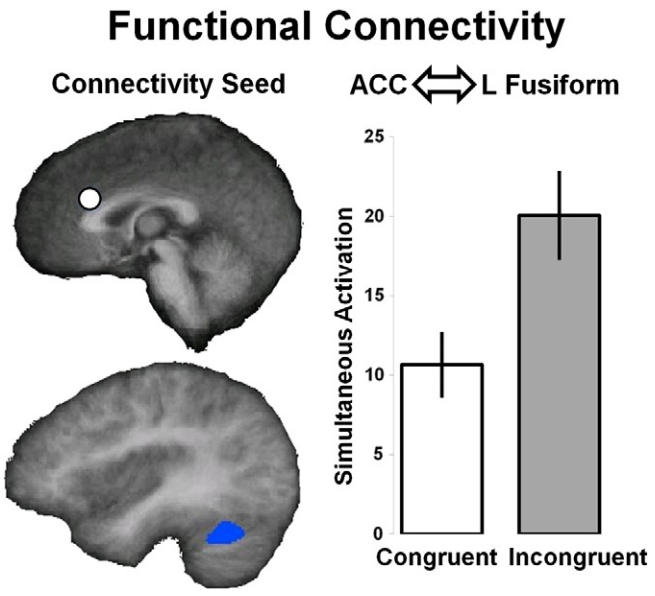

Fig. 5. Functional connectivity results using the ACC region of interest ($p < .05$ corrected, $k \geq 803$ mm^3). There was greater functional connectivity between the ACC and the lateral FG during Incongruent than Congruent trials.

We found that faces with cues along multiple social dimensions (race and emotion) differentially recruited the mPFC and ACC depending on the congruency or incongruency of their stereotypical associations. Faces whose multiple cues were stereotypically incongruent (Happy Black and Angry White targets) more strongly recruited these regions than those whose multiple cues were stereotypically congruent (Angry Black and Happy White targets). Previous studies have found similar mPFC regions to be more strongly engaged when forced to rely upon stereotypes to make judgments (Mitchell et al., 2005, 2006), when completing group-based associations in a stereotype-congruent manner (Knutson et al., 2007), and when evaluating tasks as congruent with gender stereotypes (Quadflieg et al., 2009). Extending such research, the present results suggest that this region is sensitive to the stereotypic compatibility of another person's multiple social category dimensions. Given the mPFC's role in accessing stereotype knowledge (Amodio and Frith, 2006; Mitchell et al., 2009, 2005), its increasing activation to targets' stereotypic incongruency could reflect a greater integration of conflicting stereotype knowledge spontaneously accessed from a face's multiple social category memberships.

Both the mPFC and ACC exhibited increasingly stronger responses as Black faces became happier and White faces angrier (i.e., incongruent with stereotypes). Thus, the present work additionally provides new evidence that these regions are highly sensitive to the stereotypic compatibility of a target's multiple category dimensions (e.g., race and emotion) in a graded, linear fashion. Research has long converged on the ACC's role in conflict monitoring, as it is readily activated by situations during which multiple responses compete (Botvinick et al., 2001). Such work suggests that the ACC activation found here may potentially reflect the conflict between a perceived social category and associated stereotypes (e.g., Black and Angry) when viewing stereotypically incompatible targets (e.g., Black and Happy), and we demonstrate that this response can be quite fine-grained in nature. While the current research treated racial and emotional dimensions as continuous, with targets progressing from angry to happy and from Black to White, we note that these continuous dimensions included only three intervals. Thus, including stimuli that vary on a wider continuum with smaller increments would be an important consideration for future research interested in examining the graded sensitivity of these regions.

In addition, the ACC further demonstrated greater instantaneous connectivity with the left lateral FG, an area associated with holistic face processing (Gauthier et al., 2000; Haxby et al., 2000; Kanwisher et al., 1997). Because this connectivity reflects instantaneous correlations in these regions' time-courses, we can infer that these regions are communicating to a greater extent during incongruent trials but cannot determine the directionality of this relationship. Thus, three possibilities are evident. The first is that the relationship between the ACC

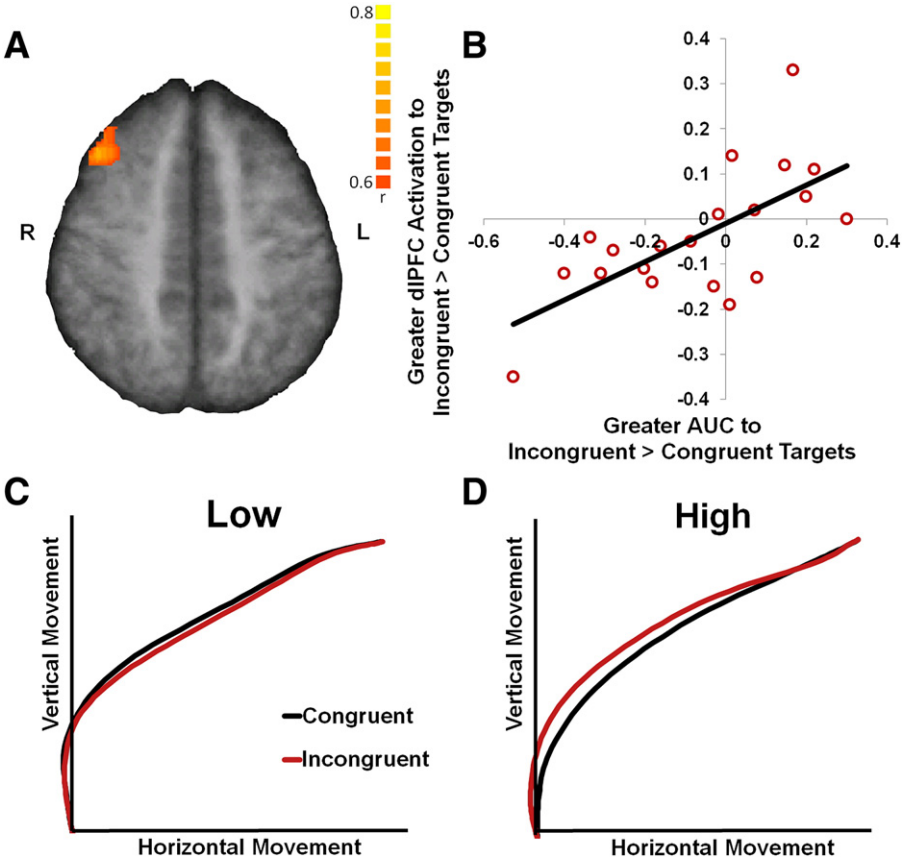

Fig. 6. A) A whole-brain analysis revealing a correlation between dlPFC responses to Incongruent > Congruent targets and the stereotypic incongruency effect in mouse-trajectories ($p < .05$, corrected). B) Correlation between mouse-trajectory area under the curve (AUC) incongruent effect and dlPFC activation to Incongruent > Congruent targets. Average mouse-trajectories for Incongruent and Congruent targets for participants low (C; bottom half) and high (D; top half) in dlPFC BOLD response to Incongruent > Congruent targets. Participants exhibiting greater dlPFC activation to incongruent targets curved more toward the stereotypically associated response when categorizing those targets (e.g., toward "Angry" for happy Black targets). Note. Median split performed on mouse-trajectory incongruency effect for visualization purposes only.

and this FG region involved in face processing reflects the bottom-up perceptual contribution of the fusiform cortex to higher-order regions involved in determining congruency with stereotypes. Perhaps more intriguingly, an alternative possibility is that this relationship represents a top-down modulation on relatively early perceptual representations, such that activated stereotypes available in higher-order regions modulate face representations in lower-level regions. This possibility would be consistent with recent theoretical models of person perception positing that lower-level face processing and higher-order social cognition mutually constrain one another over time (Freeman and Ambady, 2011) and behavioral data demonstrating how stereotypes influence the basic perception of social dimensions (Hugenberg and Bodenhausen, 2003, 2004; Johnson et al., 2012). Finally, these are not mutually exclusive, and a third possibility is that this connectivity is bidirectional, possibly reflecting both feed-forward perceptual input to higher-order regions in addition to feedback of stereotype knowledge on lower-level regions. Indeed, this region has recently been argued to play an important role in the processing of social categories (Freeman et al., 2010; Van Bavel et al., 2008). Resolving these prospects has important theoretical implications for understanding the interplay between stereotypes and face perception, and future research could disentangle these possibilities.

Using a mouse-tracking technique, we found that the extent to which participants stereotypically linked race and emotion was associated with dlPFC activation. Individuals with a stronger behavioral tendency to link angry expressions with Blacks, and happy expressions with Whites, showed greater activation in the dlPFC when viewing targets incongruent with those stereotypes. This region shares numerous functional connections with the ACC (Amodio and Frith, 2006), and previous research has demonstrated its role in suppressing prepotent responses, such as during the incongruent condition of an implicit association task (Chee et al., 2000; Luo et al., 2006) or during response inhibition in go/no-go tasks (Casey et al., 1997; MacDonald et al., 2000). One possibility is that in the present study the dlPFC had a role in inhibiting stereotypical associations that interfered with the social categorization process (e.g., inhibiting the activation of "angry" for a happy Black face, or activation of "Black" for an angry White face). If true, the dlPFC may be important for integrating bottom-up phenotypic/perceptual and top-down stereotypic/conceptual sources of information during social categorization, particularly by playing an inhibitory role in this process. Future research could test more directly the dlPFC's role in integrating stereotypical associations during the perception of social categories.

Importantly, however, this integration was not consistent across all participants. Rather, greater dlPFC activation to targets violating stereotypical expectations was evident only for individuals with stronger stereotypical Black–angry/Happy–white associations, as assessed by the behavioral results. In contrast, all participants demonstrated greater mPFC and ACC activation to incongruent targets, and this activation was unrelated to the strength of stereotypical associations. This dissociation between the dlPFC vs. mPFC and ACC in our results is noteworthy, and consistent with previous research demonstrating that nearly all individuals are aware of the content of stereotypes. Awareness, however, is independent from personal biases, or allowing this information to influence one's behaviors (Amodio and Devine, 2006; Devine, 1989). Accordingly, though the pattern of activity we find in the mPFC and ACC suggests that all participants were similarly sensitive to stereotypic

incongruities, individual differences in dlPFC activation could potentially reflect that not all participants needed to inhibit their prepotent, stereotype-driven interpretations to these targets equally.

In summary, social targets are always categorizable along numerous category dimensions (e.g., sex, race, emotion). The present work examined the neural mechanisms that underlie perceivers' ability to detect and resolve the natural inconsistencies that arise when perceiving targets at the nexus of multiple social categories. The results suggest that the mPFC and ACC respond in a sensitive, graded fashion to the stereotypic incongruities between a target's multiple category memberships. The dlPFC may potentially help to integrate and resolve these by inhibiting the stereotypical associations (e.g., Black → angry) that interfere with veridical perceptions (e.g., happy). More generally, these findings further the understanding of how bottom-up perceptual and top-down social information is resolved at the neural level, ultimately allowing us to perceive others accurately through the veil of stereotypes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.07.056.

References

Amodio, D.M., Devine, P.G., 2006. Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. J. Pers. Soc. Psychol. 91 (4), 652–661. http://dx.doi.org/10.1037/0022-3514.91.4.652.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7 (4), 268–277. http://dx.doi.org/10.1038/nrn1884.

Beer, J.S., Stallen, M., Lombardo, M.V., Gonsalkorale, K., Cunningham, W.A., Sherman, J.W., 2008. The Quadruple Process model approach to examining the neural underpinnings of prejudice. NeuroImage 43 (4), 775–783. http://dx.doi.org/10.1016/j.neuroimage.2008.08.033.

Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3D faces. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques — SIGGRAPH '99, pp. 187–194.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict Monitoring and Cognitive Control. 108 (3), 624–652. http://dx.doi.org/10.1037//0033-295X.I08.3.624.

Brigham, J., 1993. College students' racial attitudes. J. Appl. Soc. Psychol. 23, 1933–1967.

Casey, B.J., Trainor, R.J., Orendi, J.L., Schubert, A.B., Nystrom, L.E., Giedd, J.N., Rapoport, J.L., 1997. A developmental functional MRI study of prefrontal activation during performance of a go—no—go task. Journal of cognitive neuroscience 9 (6), 835–847.

Chee, M.W., Sriram, N., Soon, C.S., Lee, K.M., 2000. Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. Neuroreport 11 (1), 135–140.

Cuddy, A.J.C., Fiske, S.T., Glick, P., 2007. The BIAS map: behaviors from intergroup affect and stereotypes. J. Pers. Soc. Psychol. 92 (4), 631–648. http://dx.doi.org/10.1037/0022-3514.92.4.631.

Dale, R., Kehoe, C., Spivey, M.J., 2007. Graded motor responses in the time course of categorizing atypical exemplars. Mem. Cognit. 35 (1), 15–28.

Devine, P.G., 1989. Stereotypes and prejudice: their automatic and controlled components. J. Pers. Soc. Psychol. 56 (1), 5–18. http://dx.doi.org/10.1037//0022-3514.56.1.5.

Dovidio, J.F., Gaertner, S., 2000. Aversive racism and selection decisions: 1989 and 1999. Psychol. Sci. 11 (4), 315–319. http://dx.doi.org/10.1111/1467-9280.00262.

Dovidio, J.F., Kawakami, K., Gaertner, S., 2002. Implicit and explicit prejudice and interracial interaction. J. Pers. Soc. Psychol. 82 (1), 62–68. http://dx.doi.org/10.1037//0022-3514.82.1.62.

Fiske, S.T., Cuddy, A.J.C., Glick, P., Xu, J., 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. J. Pers. Soc. Psychol. 82 (6), 878–902. http://dx.doi.org/10.1037//0022-3514.82.6.878.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33 (5), 636–647.

Freeman, J.B., Ambady, N., 2010. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. Behav. Res. Methods 42 (1), 226–241. http://dx.doi.org/10.3758/BRM.42.1.226.

Freeman, J.B., Ambady, N., 2011. A dynamic interactive theory of person construal. Psychol. Rev. 118 (2), 247–279. http://dx.doi.org/10.1037/a0022327.

Freeman, J.B., Dale, R., Farmer, T.A., 2011. Hand in motion reveals mind in motion. Front. Psychol. 2 (April), 59. http://dx.doi.org/10.3389/fpsyg.2011.00059.

Freeman, J.B., Rule, N.O., Adams, R.B., Ambady, N., 2010. The neural basis of categorical face perception: graded representations of face gender in fusiform and orbitofrontal cortices. Cereb. Cortex 20 (6), 1314–1322. http://dx.doi.org/10.1093/cercor/bhp195.

Galinsky, A.D., Hall, E.V., Cuddy, A.J.C., 2013. Gendered races: implications for interracial marriage, leadership selection, and athletic participation. Psychol. Sci. (March) http://dx.doi.org/10.1177/0956797612457783.

Gauthier, I., Tarr, M., Moylan, J., 2000. The fusiform "face area" is part of a network that processes faces at the individual level. J. Cogn. Neurosci. 12, 495–504.

Gehring, W., Goss, B., Coles, M., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4 (6), 385–390.

Haxby, J., Hoffman, E., Gobbini, M., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4 (6), 223–233 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827445).

Hehman, E., Volpert, H.I., Simons, R.F., 2014. The N400 as an index of racial stereotype accessibility. Soc. Cogn. Affect. Neurosci. 9, 544–552. http://dx.doi.org/10.1093/scan/nst018.

Hugenberg, K., Bodenhausen, G.V., 2003. Facing prejudice: implicit prejudice and the perception of facial threat. Psychol. Sci. 14, 640–643. http://dx.doi.org/10.1046/j.0956-7976.2003.psci.

Hugenberg, K., Bodenhausen, G.V., 2004. Ambiguity in social categorization: the role of prejudice and facial affect in race categorization. Psychol. Sci. 15 (5), 342–345. http://dx.doi.org/10.1111/j.0956-7976.2004.00680.x.

Johnson, K.L., Freeman, J.B., Pauker, K., 2012. Race is gendered: how covarying phenotypes and stereotypes bias sex categorization. J. Pers. Soc. Psychol. 102 (1), 116–131. http://dx.doi.org/10.1037/a0025335.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.

Knutson, K.M., Mah, L., Manly, C.F., Grafman, J., 2007. Neural correlates of automatic beliefs about gender and race. Hum. Brain Mapp. 28 (10), 915–930. http://dx.doi.org/10.1002/hbm.20320.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., Blair, R.J.R., 2006. The neural basis of implicit moral attitude—an IAT study using event-related fMRI. NeuroImage 30 (4), 1449–1457. http://dx.doi.org/10.1016/j.neuroimage.2005.11.005.

MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288 (5472), 1835–1838.

Macrae, C.N., Bodenhausen, G.V., 2000. Social cognition: thinking categorically about others. Annu. Rev. Psychol. 51, 93–120. http://dx.doi.org/10.1146/annurev.psych.51.1.93.

Mitchell, J.P., Ames, D.L., Jenkins, A.C., Banaji, M.R., 2009. Neural correlates of stereotype application. J. Cogn. Neurosci. 21 (3), 594–604. http://dx.doi.org/10.1162/jocn.2009.21033.

Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005. The link between social cognition and self-referential thought in the medial prefrontal cortex. J. Cogn. Neurosci. 17 (8), 1306–1315. http://dx.doi.org/10.1162/0898929055002418.

Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50 (4), 655–663. http://dx.doi.org/10.1016/j.neuron.2006.03.040.

Niki, H., Watanabe, M., 1979. Prefrontal and cingulate unit activity during timing behavior in the monkey. Brain Res. 171, 213–224.

Pardo, J.V., Pardo, P.J., Janer, K.W., Raichle, M.E., 1990. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proc. Natl. Acad. Sci. U. S. A. 87 (1), 256–259.

Preacher, K.J., Curran, P.J., Bauer, D.J., 2006. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J. Educ. Behav. Stat. 31 (4), 437–448. http://dx.doi.org/10.3102/10769986031004437.

Quadflieg, S., Turk, D.J., Waiter, G.D., Mitchell, J.P., Jenkins, A.C., Macrae, C.N., 2009. Exploring the neural correlates of social stereotyping. J. Cogn. Neurosci. 21 (8), 1560–1570. http://dx.doi.org/10.1162/jocn.2009.21091.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25 (1), 230–242. http://dx.doi.org/10.1016/j.neuroimage.2004.11.017.

Talairach, J., & Tournoux, P. Co—planar stereotaxic atlas of the human brain, 1988. Theime. Germany, Stuttgart.

Van Bavel, J., Packer, D.J., Cunningham, W.A., 2008. The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. Psychol. Sci. 19 (11), 1131–1139.