

## The impact of motivation on race-based impression formation



Tianyi Li <sup>a</sup>, Carlos Cardenas-Iniguez <sup>a</sup>, Joshua Correll <sup>b</sup>, Jasmin Cloutier <sup>a,\*</sup>

<sup>a</sup> Department of Psychology, University of Chicago, Chicago, IL 60615, USA

<sup>b</sup> Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO 80309, USA

### ARTICLE INFO

#### Article history:

Received 29 May 2015

Accepted 17 August 2015

Available online 21 August 2015

#### Keywords:

Impression formation

Race perception

Dorsomedial prefrontal cortex

Evaluative incongruence

Motivation

### ABSTRACT

Affective biases toward racial out-group members, characterized by White perceivers' negative evaluations of Black individuals, prevail in U.S. culture. Such affective associations have been found to guide race-based impression formation. Accordingly, individuals may strive to resolve inconsistencies when perceiving targets violating their expectations. The current study focuses on the impact of evaluative incongruence on the activity of the dorsomedial prefrontal cortex (dmPFC) — a brain region previously shown to support impression formation. When asking participants to form impressions of positively and negatively evaluated Black and White individuals, we found preferential dmPFC activity in response to individuals paired with information that violates race-based affective associations. Importantly, individual differences in internal motivation to respond without prejudice (IMS) were found to shape the extent to which dmPFC activity indexes the interactive effects of race and affective associations during impression formation. Specifically, preferential dmPFC activity in response to evaluatively incongruent targets (i.e., Black-positive & White-negative) was present among participants with lower, but not those with higher, levels of IMS. Implications and future directions are discussed in the context of dmPFC involvement in social cognition.

© 2015 Elsevier Inc. All rights reserved.

Negative evaluations of Black individuals by White perceivers frequently characterize the emotional undercurrent of Black–White relations in U.S. culture (Dovidio and Gaertner, 2004; Weisbuch et al., 2009). Despite controversies over the underlying mechanisms, the prevalence of such affective biases and their impact on race-based impression formation has been extensively documented (Allport, 1954; Cottrell and Neuberg, 2005; Hugenberg and Bodenhausen, 2003; Sherman et al., 2008). Indeed, racially biased expectations often guide various facets of person perception. Whereas expectancy-consistent information is preferentially attended to in some instances (Sherman et al., 1998, 2000; Stangor and McMillan, 1992), people will often strive to resolve perceived incongruence when encountering individuals violating their social expectations (Hamilton and Sherman, 1996; Hastie, 1980; Macrae et al., 1999). In such instances, perceivers can be motivated to form a more individuated impression of the targets (Brewer, 1988; Fiske and Neuberg, 1990).

Studies of the neural correlates of impression formation converge on the involvement of the medial prefrontal cortex (mPFC) (Amodio and Frith, 2006; Mitchell, 2008). The dorsal region of the mPFC (dmPFC) is sensitive to the perceived relevance of social information during impression formation (see Baron et al., 2011; Cloutier et al., 2011b; Mitchell et al., 2004) and responds preferentially to targets paired with unexpected person-knowledge (Cloutier et al., 2011a; Ma et al.,

2011). Although recent studies also suggest a role for the mPFC in resolving inconsistencies arising from stereotypical expectations linking race and facial expressions during face perception (Cassidy and Gutchess, 2015; Hehman et al., 2014), the impact of race-based evaluative congruence on the neural substrates of impression formation has, to our knowledge, not been investigated. To the extent that both societal and personal prejudice often cast Blacks in a negative light, it is likely that valenced descriptions differentially shape neural responses during the formation of Black and White individuals. Indeed, previous brain-imaging research suggests that individuals paired with information that violates race-based expectations (i.e., Black-positive & White-negative) may elicit greater dmPFC activity (Cloutier et al., 2011a; Ma et al., 2011).

Despite the pervasiveness of racial prejudice, individuals vary in their evaluations of racial out-group members and in their motivations to avoid being prejudicial toward others. Previous research using event-related potentials has demonstrated that brain activity registering race-based evaluative incongruence may depend on individual differences in motivation and racial attitudes (Amodio et al., 2006, 2008). The dmPFC, which has been found to index the violation of social expectancy, may also be expected to be impacted by the perceiver's motivation but little is known about whether or how dmPFC activity is shaped by individual differences in attitudes and motivation associated with prejudice reduction (see Ames and Fiske, 2013; Kubota et al., 2012). For instance, targets violating affective expectations may evoke preferential dmPFC activity among less prejudiced perceivers as they attempt to resolve incongruence (Cloutier et al., 2011a; Ma et al., 2011). Alternatively,

\* Corresponding author at: 5848 S. University Ave, Chicago, IL 60637, USA. Fax: +1 773 702 0886.

E-mail address: [jcloutier@uchicago.edu](mailto:jcloutier@uchicago.edu) (J. Cloutier).

reduced prejudice, driven by internalized beliefs to be egalitarian or attempts to behave in nonbiased ways, may attenuate their sensitivity to evaluative incongruence, leading to relative diminished differential dmPFC activity to evaluatively incongruent compared to evaluatively congruent targets (Sherman et al., 2008). In the latter case, we would also expect to see preferential dmPFC activity to evaluative incongruence among perceivers with higher levels of prejudice.

The current study aims to examine the neural response to both positive and negative impressions of Black and White individuals. Importantly, the current study is the first to consider how variability in an individual's orientation to race (e.g., attitudes and motivation) moderates dmPFC responses to social targets that violate, as opposed to conform to, evaluative expectations during impression formation (Plant and Devine, 2009).<sup>1</sup> We hypothesize that there may be two competing hypotheses:

- 1) *greater* dmPFC activity to incongruent targets may be found among perceivers motivated to exert greater effort to resolve expectation violations (Cloutier et al., 2011a; Macrae et al., 1999); alternatively,
- 2) *diminished* dmPFC activity to evaluative incongruence may be found among these individuals (highly motivated to avoid prejudice) because they do not view positive Black targets (and negative White targets) as incongruent in the first place (by this logic, perceivers with anti-Black motives should exhibit greater activity for incongruent targets, who violate relatively prejudiced expectations) (Monteith et al., 2009; Moskowitz et al., 1999).

## Materials and methods

### Participants

We recruited 47 participants who self-identified as Caucasian American (one participant indicated both Caucasian and Latino identity) from the greater Chicago area. Each participant received \$50 for participating in the study. Of these 47 participants, two were excluded from data analysis due to excessive movement in the scanner, and one was excluded for failing to follow task instructions. The remaining 44 participants between the ages of 19 and 34 (24 female; mean age = 24.4 years, SD = 4.32 years) were included in the analysis. Participants reported normal or corrected-to-normal vision, were right-handed, and reported no abnormal neurological history. The study was approved by the Institutional Review Board at the University of Chicago and participants provided informed consent.

### Stimuli

A stimulus set consisting of 20 male full-color faces (50% Black) with neutral facial expressions was used. Stimuli were cropped in an oval shape (features such as hair and clothing were removed). An independent group of 18 participants provided ratings for the faces and words in a pilot study. Ratings of attractiveness, threat, masculinity, and prototypicality for their respective racial groups were used to select the facial stimuli (e.g., "to what extent do you think this person looks attractive/threatening/masculine/prototypical to his race?", 1 = Not at all, 7 = Very much). Stimuli from each racial group did not significantly differ on these dimensions (see Table S1 in Supplementary Material for comparison between conditions).

In addition, a word stimulus set was created from a pool of affective words (Anderson, 1968). Twenty positive and twenty negative adjectives were selected and pilot-rated to not differ on perceived valence (i.e., comparison between words ascribed to Black and White targets within each valence), stereotypicality, or arousal on a 7-point Likert scale (Valence: 1 = Very Negative, 7 = Very Positive; Stereotypicality: 1 = Very Stereotypic to African Americans, 7 = Very Stereotypic to Caucasian Americans; Arousal: 1 = Not aroused at all, 7 = Very aroused). Mean ratings of the word stimuli were reported in Table S1

(Supplementary Material). Positive and negative words did not significantly differ on arousal or stereotypicality. In addition, within each valence, words ascribed to Black and White faces did not significantly differ on perceived valence.

### Procedure

In the scanner, participants took part in the impression formation task during which they formed impressions of Black and White faces paired with either positive or negative words. Before this task, participants completed an Evaluative Priming Task outside the scanner to assess their levels of implicit prejudice and an unrelated fMRI task during which participants viewed Black and White faces in a 1-back task. After the scan session, participants were given a set of questionnaires, including explicit measures of prejudice, internal and external motivation to respond without prejudice, and a measure of contact with racial out-group members during childhood. For the purposes of the primary fMRI analysis, this ultimately yielded a 2 (Race: Black vs. White) × 2 (Valence: Positive vs. Negative) within-subject design, with measures of implicit and explicit prejudice, internal and external motivation to respond without prejudice, as well as the amount of interracial contact with Black individuals during childhood (see Cloutier et al., 2014). Details on the impression formation task and individual difference measures (including the Evaluative Priming Task and post-scan questionnaires) are described in greater details below.

### Impression formation task

Participants received instructions and performed practice trials prior to the event-related fMRI session during which they completed the impression formation task (Fig. 1). Participants were instructed to form impressions based on both the faces and the words paired with the faces based on their gut reactions, while giving evaluations from 1 to 8 (1 is very negative; 8 is very positive). The labels of the scale were counterbalanced between participants, and recoded to form an evaluation rating for each condition (i.e., higher value indicates more positive rating). Participants were instructed to rest each hand on a button box. Across two runs, 20 unique male faces (5 White faces paired with positive words, 5 Black faces paired with positive words, 5 White faces paired with negative words, and 5 Black faces paired with negative words) were presented 4 times each pseudorandomly. The stimulus face was superimposed on a black background, with a positive or negative adjective presented in white beneath the face. Each face was consistently paired with 2 positive or 2 negative words. The face-valence pairing was counterbalanced between participants. The trial sequence

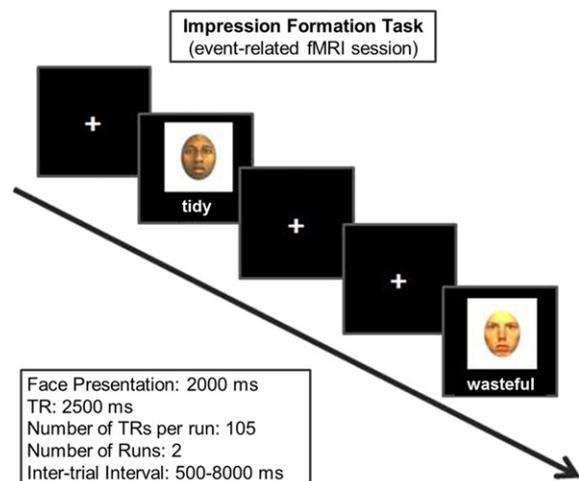


Fig. 1. Schematic representation of the task structure for the impression formation task.

was presented as the following: face-and-word pair (2000 ms) and fixation cross (500 ms). Null events were created by presenting a fixation cross for 2500 ms for a varying number of times and were interspersed in a pseudorandom fashion to introduce jitter into the fMRI time-series to create inter-trial intervals of either 500 ms, 3000 ms, 5500 ms, or 8000 ms. Stimuli were projected onto a screen measuring 240 mm by 180 mm (faces measuring 75 mm by 75 mm). E-Prime 2.0 ([www.pstnet.com/eprime](http://www.pstnet.com/eprime)) was used for stimulus presentation and data collection.

### Individual difference measures

The Evaluative Priming Task was administered to implicitly assess racial bias (Fazio et al., 1995). Participants were instructed to categorize a word as positive or negative following the presentation of a Black or a White face (we used a different face stimulus set as the impression formation study). An implicit index of prejudice (IP) was computed from the differences in reaction times (RTs) in the Evaluative Priming Task using the following formula:  $IP = (RT_{Black\_Positive} - RT_{Black\_Negative}) - (RT_{White\_Positive} - RT_{White\_Negative})$ . Greater scores indicate higher levels of pro-White or anti-Black bias. Further details about the Evaluative Priming Task procedure are described in the Supplementary Material.

Participants completed self-report questionnaires to assess their prejudicial attitudes. Specifically, participants completed the following questionnaires: the six-item Modern Racism Scale (MRS, McConahay, 1986), which was designed to assess the participants' explicit prejudice against Black individuals on a 5-point Likert scale (e.g., "Over the past few years, the government and news media have shown more respect for Blacks than they deserve"), the internal and the external motivation to respond without prejudice scales on a 9-point Likert scale (IMS and EMS, Plant and Devine, 1998), and an interracial contact measure which assesses individual differences in contact with members from different races during three separate stages of childhood (age 0–6, 6–12, and 12–18) for adults and peers in their social network respectively (e.g., "Not including your family, what percentage of the children/adults you knew belonged to each of the following categories? Asian, Blacks, Latino, etc."). The participants' childhood exposure to Black individuals was used as a covariate in the study to isolate the potential impact of interracial exposure on race perception (e.g., see Cloutier et al., 2014). Upon completing the questionnaires, participants were debriefed and compensated.

### fMRI data acquisition and analysis

Anatomical and functional whole-brain imaging was performed on a 3-T Philips Achieva Quasar scanner. High resolution T<sub>1</sub>-weighted anatomical images were acquired in the sagittal plane using a 3D Turbo Field Echo (TFE/MP-RAGE) anatomical scan (TR = 8.5 ms, TE = 4.0 ms, FOV = 240 × 228 mm, 1.0 mm slice thickness, no gap, 240 × 228 mm matrix, 1.0 × 1.0 × 1.0 mm voxel size). Functional images were collected in 2 runs of 105 TRs each, using pulse sequence parameters (TR/TE = 2500/28 ms, flip angle = 81°, contiguous slices with 3.28 mm thickness, gap 0.72 mm, FOV = 210 × 210 mm, approximately 64 × 64 mm matrix). MRI data was collected at the University of Chicago Brain Research Imaging Center.

Functional MRI data was analyzed using the general linear model (GLM) in the SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Data were preprocessed to remove sources of noise and artifacts prior to analysis. Images were realigned within and across runs to correct for head movement. Functional data were transformed into a standard anatomical space (3 mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute) which approximates the Talairach atlas space (Talairach and Tournoux, 1988). Normalized data were then spatially smoothed using an 8-mm Gaussian kernel to increase the signal-to-noise ratio and to reduce the

impact of anatomical variability that was not corrected for by the stereotaxic normalization.

For each participant, a GLM was constructed to examine condition-specific brain activity, e.g., brain regions preferentially responsive to positive Black faces, positive White faces, negative Black faces, and negative White faces. This GLM, incorporating task effects and covariates-of-no-interest (a session mean, a linear trend to account for low-frequency drift, and six movement parameters derived from realignment corrections), was convolved with a canonical hemodynamic response function and was used to compute parameter estimates ( $\beta$ ) and *t*-contrast images (containing weighted parameter estimates) for each condition at each voxel. Contrast images-of-interest for each participant (i.e., positive vs. negative Black faces, positive vs. negative White faces, positive Black vs. positive White faces, and negative Black vs. negative White faces) were submitted to a second-level random-effects analysis to create *t*-images across participants.

Empirically supported Region of Interest (ROI) analyses were conducted to identify the hypothesized impact of prejudicial attitudes on the dmPFC response to negative White, negative Black, positive White, and positive Black faces. To examine the impact of evaluative congruence, we identified the 8-mm spherical ROI for the dmPFC (MNI: 2, 54, 30) from a study that also examined the neural basis underlying violations of social expectancies (Cloutier et al., 2011a). To examine how individual variability in prejudicial attitudes relates to dmPFC response while partitioning out other prejudice-related covariates, parameter estimates for each condition were extracted to examine the effect of individual differences on dmPFC responses in the contrasts-of-interest.

In addition, we performed an exploratory second-level whole-brain regression analysis to identify the impact of the perceivers' amount of IMS on brain responses associated with the contrast images of interest (Incongruent vs. Congruent faces). The focus on IMS aimed to complement results from the correlation analyses examining associations between individual-difference measures of prejudicial attitudes and dmPFC activity (see Table 1). We used the Monte Carlo simulations using AlphaSim (Ward, 2000) to calculate the minimum cluster size at an uncorrected threshold of  $p < .001$  (required for a whole-brain correction of  $p < .05$ ). Minimum cluster size was determined to be 66 contiguous voxels.

## Results

### Behavioral results

The means for MRS, IMS, EMS, and the mean percentage of Blacks who made up the participants' social network during childhood measured by interracial contact were 1.78 (SD = .65,  $\alpha = .84$ ), 7.41 (SD = 1.48,  $\alpha = .87$ ), 4.6 (SD = 1.96,  $\alpha = .86$ ), and 7.81% (SD = 7.91%), respectively.<sup>2</sup> The means for IP (the implicit index of prejudice calculated based on the RT differences to conditions of interest) was

**Table 1**

Partial correlations between each individual-difference variable (controlling for all other four individual difference variables) and effects of Race, Valence and their interaction in the dmPFC.

	Race (White > Black)	Valence (Positive > Negative)	Race × Valence (Incongruent > Congruent)
IMS	.058 (.186)	.121 (.087)	-.419** (-.508***)
EMS	-.026 (-.008)	-.056 (-.002)	-.072 (-.168)
IP	-.087 (-.002)	.062 (-.010)	-.030 (-.102)
MRS	-.185 (-.205)	.076 (-.018)	.002 (.280)
Childhood Interracial Contact	.106 (.131)	-.185 (-.190)	-.239 (-.185)

The partial correlation coefficients are presented above. The values in parenthesis represent bivariate correlation coefficients.

$M_{IP} = 19.10$  ms,  $SD = 47.29$  ms (for the formula of the index of IP, see Procedure: Individual Difference Measures).

#### Mean level effects

The participants' evaluation ratings and RT data during the impression formation task were cleaned by removing trials in which RTs were 3.5 SD away from the individual mean RT. Evaluations and RT data were then submitted to a 2 (Race: Black vs. White)  $\times$  2 (Valence: Positive vs. Negative) repeated-measures ANOVA. Although no main effect of Race on evaluations was found, there was a main effect of Valence on evaluation ratings,  $F(1,43) = 369.973$ ,  $p < .001$ ,  $\eta^2_p = .896$ . Specifically, faces paired with positive words ( $M = 6.131$ ,  $SE = .097$ ) were rated more positively than faces paired with negative words ( $M = 2.636$ ,  $SE = .104$ ),  $t(43) = 19.235$ ,  $p < .001$ , on average across race. In addition, an interaction between Race and Valence was found,  $F(1,43) = 4.701$ ,  $p = .036$ ,  $\eta^2_p = .099$ . Black faces paired with negative words ( $M = 2.739$ ,  $SE = .110$ ) were rated more positively than White faces paired with negative words ( $M = 2.533$ ,  $SE = .114$ ),  $t(43) = 2.559$ ,  $p = .014$ . There was no difference in ratings between White faces and Black faces paired with positive words. When examining RT data, we found no evidence of a main effect of Valence, Race, or an interaction.

#### Individual differences

To examine the impact of motivation on performance during the impression formation task, we estimated the evaluation as a function of Race, Valence, IMS, and all interactions. There was no significant interaction between IMS and Valence, or between IMS and Race. Critically, there was a significant three-way interaction between Race, Valence, and IMS,  $F(1,42) = 7.974$ ,  $p = .007$ ,  $\eta^2_p = .160$ . Similar tests of the moderating influence of MRS and IP revealed no evidence of an interaction with Race, Valence, or Race  $\times$  Valence Interaction,  $F < .472$ .

To clarify the interaction between IMS and Race  $\times$  Valence Interaction, we examined the relationship between IMS and evaluative congruence and incongruence separately. We found a negative correlation between IMS and ratings of evaluatively congruent faces,  $r(42) = -.341$ ,  $p = .024$ . On the other hand, IMS did not correlate with ratings of evaluatively incongruent faces,  $r(42) = .081$ ,  $p = .601$ , suggesting divergent evaluations between incongruent and congruent faces with decreasing levels of IMS (see Fig. 2). Partial correlation between IMS and the Race  $\times$  Valence interaction on evaluations remained when controlling for EMS,  $r(41) = -.411$ ,  $p = .006$ . In addition, the relationship

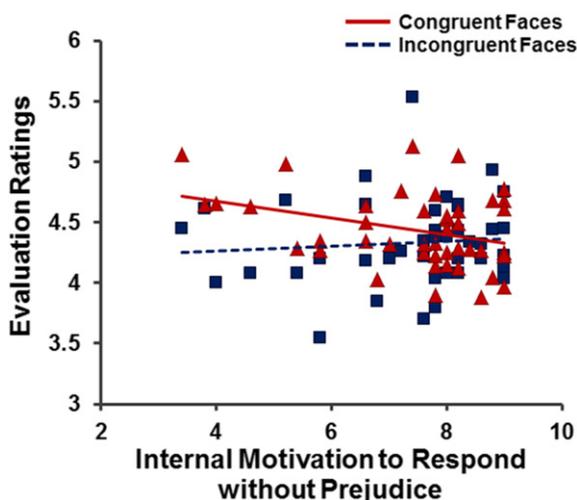


Fig. 2. Divergent evaluations of congruent and incongruent faces were found among perceivers with lower levels of IMS.

between IMS and the Race  $\times$  Valence interaction on evaluations held when including other individual-difference measures as covariates (i.e., EMS, MRS, IP, and childhood interracial contact),  $r(37) = -.358$ ,  $p = .025$ .

Additionally, we examined the impact of individual-difference measures on RTs in response to targets varying in Race and Valence. There was a significant interaction between IMS and Valence on RT,  $F(1, 42) = 4.192$ ,  $p = .047$ ,  $\eta^2_p = .091$ . Specifically, RTs in rating negative versus positive faces tended to converge for perceivers with higher levels of IMS,  $r(42) = -.301$ ,  $p = .047$ . There was no interaction between IMS and Race, or between IMS, Race, and Valence.

#### fMRI results

##### ROI analysis: Mean level effects

A 2 (Race: Black vs. White)  $\times$  2 (Valence: Positive vs. Negative) ANOVA was performed on dmPFC activity. No main effect of Race or Valence was found,  $F < 3.598$ . Critically, a significant two-way interaction between Race and Valence emerged,  $F(1, 43) = 6.654$ ,  $p = .013$ ,  $\eta^2_p = .134$ . We observed an increased dmPFC activity in response to evaluatively incongruent than congruent faces. Specifically, greater dmPFC response was found for Negative White faces ( $M = .602$ ,  $SE = .179$ ) than for Negative Black faces ( $M = .207$ ,  $SE = .178$ ),  $t(43) = 2.344$ ,  $p = .024$ . A trend suggesting increased dmPFC response to Positive Black faces ( $M = .391$ ,  $SE = .180$ ) than to Positive White faces ( $M = -.036$ ,  $SE = .179$ ) was also found,  $t(43) = 1.820$ ,  $p = .076$ . In addition, greater dmPFC response was found for Negative White faces than for Positive White faces,  $t(43) = 3.234$ ,  $p = .002$ , but not between Positive Black and Negative Black faces,  $|t| < .92$ .

##### ROI analysis: Individual difference effects

As mentioned previously, individuals with high levels of IMS, who tend to be more egalitarian, may be less likely to rely on race when forming their impressions about others, and are thus less sensitive to evaluative incongruence. Therefore, we chose to selectively focus on the impact of IMS on dmPFC responses to evaluatively congruent and incongruent faces. To help isolate the effects of IMS from other measures of prejudice-related individual differences, such as MRS, IP, EMS, and childhood interracial contact, we additionally included a set of covariates in our multiple regression models. The relationship between IMS and the dmPFC response to the Race  $\times$  Valence interaction remains significant with or without controlling for the covariates (EMS, IP, MRS and childhood interracial contact; see Table 2). In addition, none of the other individual difference measures, including EMS, IP, MRS, and childhood interracial contact, predicts dmPFC response in the same way as IMS,  $|t| < 1.89$ .

##### Effects of internal motivation to respond without prejudice

In order to identify the impact of IMS on dmPFC response to evaluative incongruence, we ran a simple regression on IMS and dmPFC activity. Although there was on average more pronounced activity in dmPFC to incongruent as compared to congruent faces, higher levels of IMS attenuated this effect,  $b = -.726$ ,  $SE = .190$ ,  $t(42) = -3.821$ ,  $p < .001$ . In other words, the impact of evaluative incongruence on impression formation, as indexed by the dmPFC activity, was more pronounced among perceivers with lower levels of IMS. Conversely, having internalized an egalitarian view by holding higher levels of IMS precluded an otherwise preponderant dmPFC activity in response to evaluative incongruence during race-based impression formation.

In order to further characterize the impact of IMS on the effects of Race, Valence, and their interaction in the dmPFC response, we plotted the Race  $\times$  Valence interaction for high and low IMS respectively. Specifically, we centered IMS scores on low IMS ( $M_{IMS} - SD_{IMS}$ ) and on high

**Table 2**  
Identification of BOLD signal as a function of Valence and Race.

BA	Brain region	k	t	p	MNI coordinates		
					x	y	z
<i>White &gt; Black</i>							
	L VTC/Cerebellum	277	5.82	<.001	-24	-57	-24
BA37	R VTC/Cerebellum	76	4.87	<.001	24	-54	-15
	R Cerebellum	74	4.11	<.001	36	-69	-27
<i>Negative &gt; Positive</i>							
BA17	L Calcarine	291	5.76	<.001	-9	-90	0
BA21	L STS	102	4.66	<.001	-54	-39	0
<i>White Negative &gt; White Positive</i>							
	L STS	113	5.47	<.001	-57	-39	0
BA17	L Middle occipital	87	5.04	<.001	-12	-96	0
	dmPFC	107	4.30	<.001	-3	45	30
	L vIPFC	178	4.23	<.001	-42	18	-18
<i>Black Negative &gt; Black Positive</i>							
BA18	L Cuneus	255	4.71	<.001	-9	-93	12
<i>White Positive &gt; Black Positive</i>							
	R VTC/Cerebellum	352	6.97	<.001	27	-48	-24
	L VTC/Cerebellum	294	5.53	<.001	-30	-45	-27
	R Precuneus	156	4.95	<.001	18	-63	24
	R Middle occipital gyrus	471	<.001	27	-78	27	
BA3	R Postcentral gyrus	176	4.38	<.001	33	-30	72
<i>Congruent &gt; Incongruent</i>							
	R VTC	86	4.81	<.001	39	-63	-6
	Posterior cingulate gyrus	76	4.80	<.001	-3	0	33
	R Precuneus	100	4.57	<.001	18	-66	24

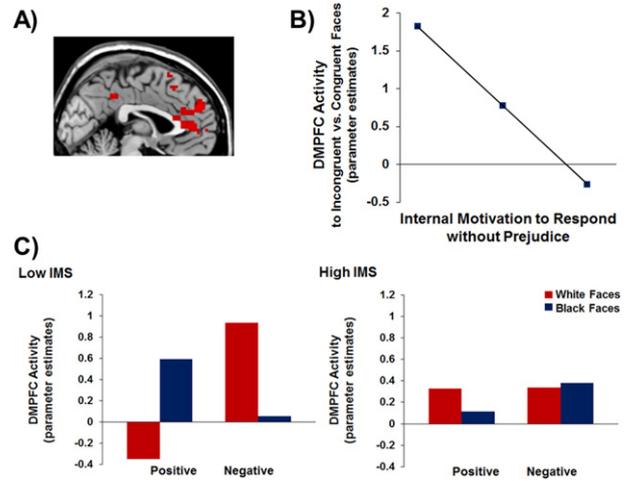
Exploratory whole-brain analysis reporting activations determined to be significant on 44 participants (threshold =  $p < .001$ , uncorrected; clusters  $\geq 66$  voxels determined by AlphaSim; actual values are reported in the table). BA = approximate Brodmann's area location. Coordinates are from the MNI atlas. Locations of the activations are determined based on the functional responses superimposed on averaged anatomical MRI images. Contrasts that do not identify suprathreshold regions are not reported.<sup>4</sup>

IMS ( $M_{IMS} + SD_{IMS}$ ). The dmPFC activities in response to each condition (i.e., White Positive, Black Positive, White Negative, and Black Negative) versus baseline were fitted into regression models where the low-IMS centered and the high-IMS centered scores served as the predictor respectively to avoid dichotomizing the continuous predictor (see Irwin and McClelland, 2003), controlling for covariates related to prejudicial attitudes and interracial contact. The intercepts for each regression model were plotted in Fig. 3C.

Given that both IMS and EMS constitute measures of motivation to respond without prejudice, we included EMS as the covariate in the regression model. IMS still served as a significant predictor for dmPFC response to incongruent vs. congruent faces,  $b = -.710$ ,  $SE = .198$ ,  $t(41) = -3.583$ ,  $p = .001$ . On the other hand, when controlling for IMS, EMS did not predict dmPFC activity. In addition, when controlling for other individual-difference variables in addition to EMS (MRS, IP, and childhood interracial contact), the impact of IMS on dmPFC activity to incongruent vs. congruent faces holds,  $b = -.705$ ,  $SE = .251$ ,  $t(37) = -2.809$ ,  $p = .008$ .<sup>3</sup> This pattern of results suggests that the influence of internal motivation to control prejudice on dmPFC recruitment in forming impressions of evaluative-incongruent and congruent exemplars cannot be easily attributed to other individual-difference factors related to race perception and attitudes (Fig. 3B).

**Whole-brain analysis**

Brain regions preferentially recruited as a function of the Race and Valence conditions are reported in Table 2. In accordance with the incongruence effect obtained from the dmPFC ROI analyses, greater dmPFC activity was found for Negative White than for Positive White faces. Interestingly, greater activity in the ventral temporal cortex (VTC) was found for White compared to Black faces, corroborating previous findings suggesting that the VTC is preferentially recruited when



**Fig. 3.** The dmPFC activity in response to evaluative incongruence varies as a function of perceivers' levels of IMS. (A) The figure presents the sagittal sections illustrating the dmPFC activation obtained from the whole-brain regression analysis. Decreasing dmPFC activity was found with increasing levels of IMS. (B) Greater dmPFC response to evaluatively incongruent than to congruent faces was only found for perceivers with lower levels of IMS, above and beyond other relevant individual-difference variables. (C) The figure represents the Race  $\times$  Valence interaction on dmPFC activity as a function of levels of IMS. The dmPFC ROI used was identified from an independent study (MNI: 2, 54, 30; see Cloutier et al., 2011a). Higher levels of internal motivation to respond without prejudice attenuated differential dmPFC response to evaluative incongruence vs. congruence, controlling for covariates related to prejudicial attitudes and interracial contact.

individuals perceive same-race as compared to other-race faces (Golby et al., 2001). The ventrolateral prefrontal cortex (vlPFC) was also found to show greater activation to Negative White than to Positive White. The vlPFC has previously been shown to be responsive when perceiving targets paired with person knowledge (Cloutier et al., 2011b; Satpute et al., 2014), and updating impressions based on diagnostic behaviors (Mende-Siedlecki et al., 2013), thus possibly integrating various sources of information available during impression formation (see Cloutier et al., 2011b). In addition, this region has also been found to play a role in cognitive and emotion regulation operations (Ochsner et al., 2009; Wager et al., 2008). Therefore, preferential vlPFC activity when perceiving incongruent White faces may suggest a tendency to resolve inconsistency between information and expectation for in-group members. However, the design of the current study does not allow for an understanding of the strategy applied in resolving the

**Table 3**  
Identification of BOLD signal showing differential activity to evaluative incongruence in relation to increases and decreases in internal motivation to respond without prejudice.

BA	Brain region	k	t	p	MNI coordinates		
					x	y	z
<i>Decreasing activity with increasing IMS</i>							
	dmPFC/Anterior cingulate	757	4.72	<.001	-6	45	9
R	Supplementary motor area	402	4.49	<.001	12	12	66
R	Temporal pole	338	4.82	<.001	45	24	-27
L	Insula	85	4.82	<.001	-39	18	3
L	Putamen	369	<.001	-18	6	0	
L	Middle occipital gyrus	98	3.83	<.001	-24	-81	3
L	Cerebellum	322	4.74	<.001	-27	-69	-27
BA8	L Middle frontal gyrus	149	4.47	<.001	-24	21	45
<i>Increasing activity with increasing IMS</i>							
NA							

Exploratory whole-brain regression analysis reporting activations determined to be significant on 44 participants (threshold =  $p < .001$ , uncorrected; clusters  $\geq 66$  voxels determined by AlphaSim); BA = approximate Brodmann's area location. Coordinates are from the MNI atlas. Locations of the activations are determined based on the functional responses superimposed on averaged anatomical MRI images.

inconsistency – whether participants rely on prior knowledge and discount incongruent information or update their impression.

We also performed an exploratory whole-brain regression analysis to identify all brain regions activated to incongruent vs. congruent faces as a function of IMS (Table 3). In accordance with the results of the ROI analysis on dmPFC, we found decreased dmPFC activity to incongruent versus congruent targets in relation to increased levels of IMS (Fig. 3A).

## Discussion

When asking White perceivers to form impressions of White and Black faces paired with valenced descriptive words, we found greater dmPFC activity to evaluatively incongruent targets (i.e. White faces paired with negative traits and Black faces paired with positive traits) compared to congruent targets (i.e. White faces paired with positive traits and Black faces paired with negative traits). These results converge with recent findings showing greater dmPFC responses to stereotypically incongruent faces based on the interaction of race and facial expressions of emotion (Hehman et al., 2014). Importantly, this pattern of result was attenuated by the perceivers' motivation to avoid being prejudiced, such that only perceivers low in IMS showed a difference in dmPFC activity when forming impressions. Indeed, high-IMS perceivers did not exhibit such difference in dmPFC activity. This influence of IMS on differential dmPFC activity cannot be attributed to other individual differences, including EMS, explicit prejudice, implicit prejudice, or childhood interracial contact. Importantly, EMS neither predicted neural and behavioral responses in the same way as IMS, nor did it alleviate the relationship between IMS and dmPFC response (i.e., the association between IMS and dmPFC activity held after controlling for EMS), indicating that tendencies to avoid prejudice driven by external motivation (e.g., social norms to appear unbiased) are not as effective in shaping race-based impression formation as having internalized goals to be nonprejudiced (Amodio et al., 2006). Accordingly, these findings not only provide implications for our understanding of the neural substrates of race-based impression formation but also represent a novel demonstration of how individual differences in motivation can impact dmPFC activity during social cognition.

The dmPFC is engaged when attributing psychological states to conspecifics (Amodio and Frith, 2006; Corradi-Dell'Acqua et al., 2015; Mar, 2011), supports the individuation of in-group members during impression formation (Freeman et al., 2010; Harris and Fiske, 2006), and preferentially responds to the availability of person-knowledge when perceiving others (Cloutier et al., 2011b; Todorov et al., 2007). To our knowledge, the current study is the first fMRI investigation of race-based evaluative congruence during impression formation. Previous research has shown that the dmPFC responds preferentially to targets paired with unexpected person-knowledge (Cloutier et al., 2011a; Ma et al., 2011; Mende-Siedlecki et al., 2012), but none of them have specifically looked at violations of race-based affective expectations. Critically, however, the preferential recruitment of dmPFC underlying impression formation of incongruent targets was absent in high-IMS perceivers. In addition, results from the whole-brain regression analysis demonstrate decreased anterior cingulate cortex activity with increased levels of IMS (Table 3), suggesting that control of race-based incongruence was not higher among high-IMS perceivers during the impression formation task (see Amodio et al., 2008). These results give rise to the possibility that high-IMS perceivers were less sensitive to evaluative incongruence as a result of their internalized goal to be nonprejudiced, rather than by exerting greater efforts to regulate racial bias when forming impressions of targets violating affective expectations. This interpretation is consistent with the proposition that internally motivated, but not externally motivated, individuals may be initially training themselves to reduce biased affective associations distinguishing in-group and out-group members (see Sherman et al., 2008).

Despite evidence that internal motivation attenuates prejudiced behaviors (Amodio et al., 2003; Devine et al., 2002; Plant et al., 2010; Riek et al., 2013), this is to our knowledge the first study demonstrating that highly internally motivated individuals, as a result of their internalized goal and personal beliefs to be nonprejudiced, become less sensitive to what others perceive to be evaluatively incongruent. In apparent contrast with our current findings, a previous study using event-related potentials during a Weapon Identification Task found that high-IMS individuals exert greater effort to avoid potential prejudicial responses as indexed by a larger error-related negativity (ERN) component (Amodio et al., 2008). However, the Weapon Identification Task and the impression formation task performed in the current study differ in important ways. Impression formation may not typically involve conflict monitoring (see Cloutier et al., 2011a; Mende-Siedlecki et al., 2012), particularly among perceivers holding internalized egalitarian beliefs. Whereas the Weapon Identification Task has been used to measure motivation to monitor and regulate biased responses by isolating errors from accurate trials, impression formation processes may instead be susceptible to be shaped by persistent efforts to avoid prejudice.

This interpretation of our fMRI results is further supported by the fact that preference ratings for congruent and incongruent targets in the current study only differed for low-IMS perceivers, whereas evaluative congruence did not impact preference ratings given by high-IMS perceivers. Further examination reveals that the divergence in preference ratings among low-IMS perceivers was driven by more positive ratings for positive White faces than Black faces. On the contrary, high-IMS perceivers viewed White and Black faces paired with positive traits in a similarly favorable light, lending support for previous research suggesting increased preference for positive Black targets among high-IMS individuals (Riek et al., 2013). In line with previous research suggesting that high-IMS individuals are more likely to engage in approach strategies to achieve positive outcomes during interracial interaction (Plant et al., 2010), our findings suggest that individuals with higher levels of IMS may be less inclined to use race-based expectations to form impressions of targets.

The obtained results offer a novel demonstration of how individual differences can impact the recruitment of the dmPFC, a brain region believed to be central to social cognition. Our results demonstrate that internally motivated individuals may be less sensitive to race as a result of the internalization of egalitarian attitudes. These findings offer insights into potential mechanisms by which persistent efforts to avoid prejudicial responses may efficiently alter race-based impression formation.

## Footnote

1. We also examined amygdala activity to explore the potential relationship between valence and race during impression formation (Kubota et al., 2012). Amygdala ROI was identified from a previous study on race perception (Cloutier et al., 2014). Results and discussions of findings on amygdala were reported in Supplementary Material.
2. Our samples could be considered relatively high on IMS, because mean IMS is significantly higher than the midpoint of the 9-point scale,  $t(43) = 10.836, p < .001$ ; our samples scored slightly less than average EMS,  $t(43) = -1.356, p = .182$ .
3. One participant did not complete the childhood interracial contact measure; data from this participant was only excluded when adding childhood interracial contact as a covariate in analysis.
4. In the Incongruent > Congruent contrast, dmPFC emerges using a more liberal threshold ( $p < .005$ , uncorrected; cluster size  $\geq 40$ ).

## Acknowledgments

We would like to thank Dr. Jennifer Kubota for her valuable suggestions on the manuscript, as well as Ivo Gyurovski, Jonathan Feng, Nia Sotto, Cayla Echols, Linda Lin-shiu, and Katy Carlyle for their help in

implementing the project, In addition, we would like to thank Dr. Jean Decety, Dr. Xiaodong Guo, Stanley Peters, and Bruce Jamison from the Brain Research Imaging Center (BRIC) at the University of Chicago.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.08.035>.

## References

- Allport, G., 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Ames, D.L., Fiske, S.T., 2013. Outcome dependency alters the neural substrates of impression formation. *NeuroImage* 83, 599–608.
- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.
- Amodio, D.M., Harmon-Jones, E., Devine, P.G., 2003. Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *J. Pers. Soc. Psychol.* 84, 738.
- Amodio, D.M., Kubota, J.T., Harmon-Jones, E., Devine, P.G., 2006. Alternative mechanisms for regulating racial responses according to internal vs external cues. *Soc. Cogn. Affect. Neurosci.* 1, 26–36.
- Amodio, D.M., Devine, P.G., Harmon-Jones, E., 2008. Individual differences in the regulation of intergroup bias: the role of conflict monitoring and neural signals for control. *J. Pers. Soc. Psychol.* 94, 60.
- Anderson, N.H., 1968. Likableness ratings of 555 personality-trait words. *J. Pers. Soc. Psychol.* 9, 272.
- Baron, S.G., Gobbini, M., Engell, A.D., Todorov, A., 2011. Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Soc. Cogn. Affect. Neurosci.* 6, 572–581.
- Brewer, M.B., 1988. A Dual Process Model of Impression Formation. *Advances in Social Cognition*. Erlbaum, Hillsdale, NJ.
- Cassidy, B.S., Gutchess, A.H., 2015. Neural responses to appearance-behavior congruity. *Soc. Cogn.* 33, 211–226.
- Cloutier, J., Gabrieli, J.D., O'Young, D., Ambady, N., 2011a. An fMRI study of violations of social expectations: when people are not who we expect them to be. *NeuroImage* 57, 583–588.
- Cloutier, J., Kelley, W.M., Heatherton, T.F., 2011b. The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Soc. Neurosci.* 6, 63–75.
- Cloutier, J., Li, T., Correll, J., 2014. The impact of childhood experience on amygdala response to perceptually familiar black and white faces. *J. Cogn. Neurosci.* 26, 1992–2004.
- Corradi-Dell'Acqua, C., Turri, F., Kaufmann, L., Clément, F., Schwartz, S., 2015. How the brain predicts people's behavior in relation to rules and desires. Evidence of a Medio-Prefrontal Dissociation (Cortex).
- Cottrell, C.A., Neuberg, S.L., 2005. Different emotional reactions to different groups: a sociofunctional threat-based approach to "prejudice". *J. Pers. Soc. Psychol.* 88, 770.
- Devine, P.G., Plant, E.A., Amodio, D.M., Harmon-Jones, E., Vance, S.L., 2002. The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *J. Pers. Soc. Psychol.* 82, 835.
- Dovidio, J.F., Gaertner, S.L., 2004. On the nature of contemporary prejudice. *Race, class, and gender in the United States: An integrated study*. 132.
- Fazio, R.H., Jackson, J.R., Dunton, B.C., Williams, C.J., 1995. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* 69, 1013.
- Fiske, S.T., Neuberg, S.L., 1990. A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. *Adv. Exp. Soc. Psychol.* 23, 1–74.
- Freeman, J.B., Schiller, D., Rule, N.O., Ambady, N., 2010. The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Hum. Brain Mapp.* 31, 150–159.
- Golby, A.J., Gabrieli, J.D., Chiao, J.Y., Eberhardt, J.L., 2001. Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845–850.
- Hamilton, D.L., Sherman, S.J., 1996. Perceiving persons and groups. *Psychol. Rev.* 103, 336.
- Harris, L.T., Fiske, S.T., 2006. Dehumanizing the lowest of the low neuroimaging responses to extreme out-groups. *Psychol. Sci.* 17, 847–853.
- Hastie, R., 1980. In: Hastie, R., Ostrom, T.M., Ebbesen, E.B. (Eds.), *Memory for behavioral information that confirms or contradicts a personality impression*, pp. 155–178.
- Hehman, E., Ingbreten, Z.A., Freeman, J.B., 2014. The neural basis of stereotypic impact on multiple social categorization. *NeuroImage* 101, 704–711.
- Hugenberg, K., Bodenhausen, G.V., 2003. Facing prejudice implicit prejudice and the perception of facial threat. *Psychol. Sci.* 14, 640–643.
- Irwin, J.R., McClelland, G.H., 2003. Negative consequences of dichotomizing continuous predictor variables. *J. Mark. Res.* 40, 366–371.
- Kubota, J.T., Banaji, M.R., Phelps, E.A., 2012. The neuroscience of race. *Nat. Neurosci.* 15, 940–948.
- Ma, N., Vandekerckhove, M., Van Overwalle, F., Seurinck, R., Fias, W., 2011. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Soc. Neurosci.* 6, 123–138.
- Macrae, C.N., Bodenhausen, G.V., Schloerscheidt, A.M., Milne, A.B., 1999. Tales of the unexpected: executive function and person perception. *J. Pers. Soc. Psychol.* 76, 200.
- Mar, R.A., 2011. The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134.
- McConahay, J.B., 1986. *Modern racism, ambivalence, and the modern racism scale*.
- Mende-Siedlecki, P., Cai, Y., Todorov, A., 2012. The neural dynamics of updating person impressions. *Soc. Cogn. Affect. Neurosci.* <http://dx.doi.org/10.1093/scan/nss040>.
- Mende-Siedlecki, P., Baron, S.G., Todorov, A., 2013. Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *J. Neurosci.* 33, 19406–19415.
- Mitchell, J.P., 2008. Contributions of functional neuroimaging to the study of social cognition. *Curr. Dir. Psychol. Sci.* 17, 142–146.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *J. Neurosci.* 24, 4912–4917.
- Monteith, M.J., Lybarger, J.E., Woodcock, A., 2009. Schooling the cognitive monster: the role of motivation in the regulation and control of prejudice. *Soc. Personal. Psychol. Compass* 3, 211–226.
- Moskowitz, G.B., Gollwitzer, P.M., Wasel, W., Schaal, B., 1999. Preconscious control of stereotype activation through chronic egalitarian goals. *J. Pers. Soc. Psychol.* 77, 167.
- Ochsner, K.N., Hughes, B., Robertson, E.R., Cooper, J.C., Gabrieli, J.D., 2009. Neural systems supporting the control of affective and cognitive conflicts. *J. Cogn. Neurosci.* 21, 1841–1854.
- Plant, E.A., Devine, P.G., 1998. Internal and external motivation to respond without prejudice. *J. Pers. Soc. Psychol.* 75, 811.
- Plant, E.A., Devine, P.G., 2009. The active control of prejudice: unpacking the intentions guiding control efforts. *J. Pers. Soc. Psychol.* 96, 640.
- Plant, E.A., Devine, P.G., Peruche, M.B., 2010. Routes to positive interracial interactions: approaching egalitarianism or avoiding prejudice. *Personal. Soc. Psychol. Bull.* 36, 1135–1147.
- Riek, B.M., Mania, E.W., Gaertner, S.L., 2013. Reverse subtyping: the effects of prejudice level on the subtyping of counterstereotypic outgroup members. *Basic Appl. Soc. Psychol.* 35, 409–417.
- Satpute, A.B., Badre, D., Ochsner, K.N., 2014. Distinct regions of prefrontal cortex are associated with the controlled retrieval and selection of social information. *Cereb. Cortex* 24, 1269–1277.
- Sherman, J.W., Lee, A.Y., Bessenoff, G.R., Frost, L.A., 1998. Stereotype efficiency reconsidered: encoding flexibility under cognitive load. *J. Pers. Soc. Psychol.* 75, 589.
- Sherman, J.W., Macrae, C.N., Bodenhausen, G.V., 2000. Attention and stereotyping: cognitive constraints on the construction of meaningful social impressions. *Eur. Rev. Soc. Psychol.* 11, 145–175.
- Sherman, J.W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T.J., Groom, C.J., 2008. The self-regulation of automatic associations and behavioral impulses. *Psychol. Rev.* 115, 314.
- Stangor, C., McMillan, D., 1992. Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychol. Bull.* 111, 42.
- Talairach, J., Tournoux, P., 1988. *Co-planar stereotaxic atlas of the human brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging*.
- Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V., 2007. Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45, 163–173.
- Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N., 2008. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* 59, 1037–1050.
- Ward, B.D., 2000. *Simultaneous Inference for fMRI Data*.
- Weisbuch, M., Pauker, K., Ambady, N., 2009. The subtle transmission of race bias via televised nonverbal behavior. *Science* 326, 1711–1714.