Inhibition or Ideology? The Neural Mechanisms of Evaluating Race-Targeted Government Assistance[1]

Frank J. Gonzalez
School of Government and Public Policy
University of Arizona
fgonzo@email.arizona.edu

Ingrid J. Haas
Department of Political Science
Center for Brain, Biology, and Behavior
University of Nebraska – Lincoln
ihaas2@unl.edu

[1] Prepared for presentation at the annual meeting of the American Political Science Association, September 1, 2017, San Francisco, CA.

Data collection was funded entirely through a Political Science Doctoral Dissertation Research Improvement Grant (PS DDRIG) provided by the National Science Foundation Division of Social and Economic Sciences (NSF SES; Award #1560432). We would like to thank the NSF for supporting this research. We also owe a debt of gratitude to the Center for Brain, Biology, and Behavior (CB3) at the University of Nebraska – Lincoln, for accommodating this research.

We would like to thank Elizabeth Theiss-Morse for continued support and feedback, as well as John Hibbing, Kevin Smith, Sergio Wals, Sarah Gervais, Johnathan Peterson, Stephen Schneider, and the Political Behavior Research Group at the University of Nebraska – Lincoln.

*****The results presented in this manuscript are preliminary, and so should not be referenced without first consulting the authors.*****

Abstract:

When are people most able to inhibit racial prejudice while evaluating government assistance policies? A common proposal is that by making race salient, individuals can use controlled processing and effortful thought to inhibit automatic racial biases and abide by egalitarian norms. We implemented a 2x2 within-groups design using fMRI to test this hypothesis and identify the neural processes associated with conscious versus nonconscious racial policy evaluations. We also examine variation in these processes across the ideological spectrum. Participants decided whether to support or oppose government assistance to applicants who were either Black or White and whose race was shown either consciously (1000 ms) or nonconsciously (33 ms). We focused on brain regions associated with automatic processing (amygdala), conflict detection (anterior cingulate, insula), and controlled processing (orbitofrontal cortex and prefrontal cortex). We found only minor differences in brain regions associated with automatic processing, but areas associated with conflict detection and controlled processing were more active when race was conscious, as expected. Anti-Black racial biases were also less likely when race was conscious and were associated with lower levels of conflict detection and controlled processing. Critically, the effects of race being conscious on anti-Black biases did not differ by participant ideology, but the effects of conflict detection and controlled processing were moderated by ideology. Some brain regions associated with conflict detection and controlled processing were associated with prejudice inhibition among liberals but exacerbation among conservatives, other regions led to prejudice exacerbation among liberals but not conservatives, and there was no substantial evidence of prejudice inhibition via controlled processing among conservatives. We discuss our findings in terms of interventions for reducing the influence of racial prejudice on policy attitudes, as well as the relationship between prejudice and ideology.

Inhibition or Ideology? The Neural Mechanisms of Evaluating Race-Targeted Government Assistance

For most people, the theory goes, negative racial attitudes affect political thinking automatically or not at all.
- Valentino, Hutchings, and White (2002, p. 77)

A popular view, with deep roots in eighteenth-century enlightenment thinking, is that bigotry and prejudice stem from a deficiency. Whether the deficiencies involve education, reason, awareness, or even cognitive ability, the common assumption is that prejudice is due to a lack of complex, enlightened thought. Accordingly, this view suggests prejudice reflects a default, simplistic way of thinking and can be overcome through effortful thought and learning about the logical fallacies of prejudiced beliefs. Although this viewpoint is primarily evident in lay rhetoric, it is corroborated by work suggesting racial prejudice has its most powerful effects on political attitudes when racial cues are covert and thus go undetected by conscious awareness, and that people are most able to inhibit racial biases when they engage in controlled processing – i.e. when they "think about" race (e.g. Cunningham et al. 2004; Mendelberg 2001; Valentino, Hutchings, and White 2002).

There are two aspects of this line of reasoning that, despite some empirical support, we suggest deserve extra consideration. First, we do not know the extent to which overt (compared to covert) racial cues actually induce controlled processing (i.e. effortful thought) rather than automatic, affective processing when it comes to political evaluations. Some work in psychology has shown that overt racial cues tend to trigger increased controlled processing and reduced automatic processing (Cunningham et al. 2004), no work to date has examined these processes in the realm of politics, where the racial attitudes compete with other factors such as ideology to influence evaluations (e.g. Sniderman, Carmines, Layman, and Carter 1996). Although some work has used experimental manipulations to indirectly stimulate conscious race processing (e.g.

Valentino, Hutchings, and White 2002), we have yet to directly examine the particular neural processes that accompany overt racial cues, and we do not know which precise neural processes lead people to inhibit their racial biases when confronted with overt racial cues.

Second, the consequences of controlled processing for how racial biases impact policy attitudes are not fully understood. Specifically, little work has examined individual-level variation in how controlled processing affects the expression of racial prejudice. Although it may be the case that controlled processing is associated with lower levels of racial bias in the aggregate, this does not necessarily mean that inhibition of prejudice is the universal outcome of controlled processing. It could be the case that for some, effortful racial thought allows for the inhibition of prejudice and adherence to the norms of egalitarianism, but for others, controlled processing actually exacerbates the influence of prejudice by allowing them to justify or rationalize deep-seated biases (e.g. Fazio 1990; Plant and Devine 1998). Indeed, race-targeted policies such as affirmative action inherently make race salient, and yet we know racial prejudice plays a significant role in determining attitudes toward such policies nonetheless (e.g. Henry and Sears 2002; Kinder and Sears 1981; Tesler and Sears 2010).

But among whom should we expect controlled processing to inhibit versus exacerbate biases? Here, we focus on political ideology as a key moderator. Political ideology is among the strongest predictors of attitudes toward race-targeted policies across empirical studies, and a good deal of work has shown that ideological values are often deferred to as socially acceptable justifications for opinions toward government assistance (e.g. Federico and Sidanius 2002; Henry and Sears 2002; Kinder and Mendelberg 2000; Kinder and Sanders 1996; Kinder and Sears 1981; Sidanius, Pratto, and Bob 1996). As such, we look to political ideology as a possible moderator of the effects of controlled processing.

We present findings from a within-subject experiment using functional Magnetic Resonance Imaging (fMRI) to measure brain activity while participants evaluated applicants for government assistance who were either Black or White and whose race was known either consciously (their face shown for 1000 ms) or nonconsciously (their face shown for 30 ms). In doing so, we investigate the neural mechanisms of support for government assistance when race is either covertly or overtly cued as a factor in one's evaluations, as well as the moderating role of ideology.

We set out to accomplish two primary goals with this experiment. First, this investigation provides a significant methodological advancement to existing political science literature by allowing for direct observation of the neural mechanisms assumed to underlay overtly race-related policy evaluations. For the most part, these neural processes have been assumed, ignored, or measured indirectly (e.g. Huber and Lapinski 2006; Hurwitz and Peffley 2005; Mendelberg 2001; Valentino, Hutchings, and White 2002). Second, we aim to adjudicate between two competing hypotheses regarding the consequences of controlled processing for how racial biases affect policy attitudes. On the one hand, much of the existing literature in psychology and political science suggests controlled processing allows individuals to inhibit the influence of racial biases (e.g. Cunningham et al. 2004; Mendelberg 2001; Plant and Devine 1998; Valentino, Hutchings, and White 2002). On the other hand, analytical thinking, effortful thought, and even education are often theorized to be tools individuals can use to rationalize pre-existing opinions (e.g. Kahan et al. 2012; Lodge and Taber 2013) and racial biases (e.g. Federico 2004; Federico and Sidanius 2002), and the effects of controlled processing have been shown to be moderated substantially by motivation (Fazio 1990; Plant and Devine 1998). If this is the case, although controlled processing might inhibit racial biases in support for government assistance for some (e.g. liberals), we should expect controlled processing to exacerbate the role of racial biases in driving support for government assistance for other individuals (e.g. conservatives). These competing hypotheses have starkly different implications for what sorts of interventions are most likely to succeed in reducing the influence of racial prejudice on policy attitudes. Is making race salience in political discourse enough to impede the impact of prejudice?

The Consequences of Controlled Processing about Race

The idea that encouraging effortful thought can reduce prejudice emanates largely from work on when racial cues in political messages are most likely to trigger racial attitudes. This work focuses on political attitudes broadly rather than race-targeted policy attitudes. Mendelberg (2001) developed a framework describing the conflict that most White Americans face when they evaluate a racialized political message. According to this framework, the racial appeals most likely to be influential are those that only nonconsciously prime race because when race is primed consciously, people are cued to reject prejudiced instincts and abide by egalitarian norms encouraged by society. However, when race is primed nonconsciously, the argument is that prejudiced instincts can influence political attitudes uninhibited. This model is supported by an array of empirical evidence showing that a wide range of covert racial cues can trigger race-based evaluations of political messages, and that the mechanism covert racial cues operate through is by making race cognitively accessible (Mendelberg 2001; Valentino, Hutchings, and White 2002). The assumptions of this model are therefore that covert racial cues trigger uncontrollable racial biases, that overt racial cues trigger higher-level cognitive processes, and that higher-level cognitive processes triggered by overt racial cues necessarily yield inhibition of prejudice. The implication is that when race is salient (as it always is with race-targeted policies), prejudice influences race-targeted policies because some individuals simply do not engage in controlled inhibition. As explained by

Valentino, Hutchings, and White in the epigraph at the beginning of this paper, the entirety of the role of racial prejudice is assumed to be through covert, nonconscious means.

Despite evidence supporting this framework, some findings call its assumptions into question. The effects of covert racial cues are diminished when the cue presents a racial group in a counter-stereotypic way (Valentino, Hutchings, and White 2002) – a finding explained as due to greater controlled processing of counter-stereotypic information (Brewer, Dull, and Lui 1981; Hastie 1981). Also, the greater ability of covert racial cues rather than overt racial cues to trigger racial attitudes when evaluating a political message seems concentrated only among the highly educated (Huber and Lapinski 2006). These findings suggest that although the framework constructed by Mendelberg (2001) may be somewhat robust, there are significant limits to its generalizability. It seems that covert cues do not *necessarily* lead to uncontrollable racial biases. Further, the finding that the model holds mainly among the highly educated suggests that the effects of making race salient depend on the person: among highly educated individuals it seems likely that controlled processing will reduce the effects of prejudice, but less educated individuals may hold egalitarian values a bit more loosely, and so controlled processing may often yield an equal or greater role for prejudice. Indeed, work on motivation to respond without prejudice suggests that individuals vary a great deal in the degree to which they are internally or externally motivated to use controlled processes to inhibit prejudice, even when race is salient (e.g. Plant and Devine 1998). Not everyone holds the egalitarian views assumed to yield controlled inhibition when racial appeals are conscious. Nonetheless, this assumption remains largely accepted: that higher-level, controlled processing universally yields prejudice inhibition. The best way to reduce the role of prejudice then is to make people conscious of race and encourage greater controlled processing.

A Neural Framework of the Automatic and Controlled Components of Prejudice

A variety of methodologies have been used to estimate the influence of racial prejudice on political attitudes, but these methods rely heavily on assumptions regarding the mechanisms by which prejudice influences policy attitudes. To understand *how* prejudice influences such attitudes and thus how its role can be diminished, the most promising avenue of research is to directly examine the hypothesized mechanisms at work in the brain. Advances in medical imaging technology have made it possible to obtain quantitative measurements of brain activity by tracking the increases in oxygenated blood that accompany increases in activity in particular parts of the brain. Functional Magnetic Resonance Imaging (fMRI) uses a magnetic field in conjunction with pulses of radio frequencies to track the iron that comes with movement in oxygenated blood (Friston 2009).

There is one simple, overarching finding of research using fMRI methodology – the human brain is extremely complex, and to suggest that any attitude or behavior can be traced solely to a single region of the brain, or conversely that any single region of the brain has just one behaviorally relevant function, would be tremendously misguided. Nonetheless, research has advanced our understanding of neural activity substantially, and we can look at trends in brain activity in response to specific stimuli to infer which parts of the brain are *generally* active in particular contexts, including contexts related to politics (see Haas 2016; Jost, Nam, Amodio, and Van Bavel 2014; Schreiber 2011). Through replication and convergence, we can develop frameworks of specific neural processes so brain activity can be used to predict real-world, ecologically valid outcomes (Berkman and Falk 2013). Neuroscience research on prejudice has provided an excellent starting point for understanding the mechanisms by which prejudice might influence policy attitudes. Findings from this work have essentially left us with a roadmap for how the brain functions when prejudice occurs automatically versus when controlled processes inhibit racial biases.

Group-based categorization seems hardwired into the brain's processing of social information. Some work has shown that even when group membership is meaningless (i.e. created within the confines of an experimental task), the brain reacts to fellow ingroup members by activating certain regions associated with awareness of self (Molenberghs et al. 2013; Shkurko, 2012; Van Bavel, Packer, and Cunningham 2008; Volz et al. 2009;). With regard to reactions to outgroups, the most commonly activated brain regions include the amygdala and insula, suggesting that people immediately categorize outgroup members as emotionally significant (e.g. Cunningham et al. 2004; but also see Wheeler and Fiske 2005) and associate them with uncertainty or danger, which must be avoided (Richeson et al. 2003; Rilling et al. 2008; see Shkurko 2012 for meta-analysis).[2] This reaction – that is, the activation of these brain regions – occurs automatically and even when race is only primed nonconsciously (Cunningham et al. 2004).

Something interesting happens, however, when race is consciously recognized and there is a conflict between people's nonconscious adverse reactions to outgroups and their conscious goals of being or appearing unprejudiced (e.g. "I feel threatened by the Black man walking toward me but know I should not"): the anterior cingulate cortex (ACC) and insula detect the goal conflict, the amygdala response is suppressed, and certain parts of the prefrontal cortex (PFC) kick in that are associated with higher-level cognitive functioning (see Stanley, Phelps, and Banaji 2008).[3] Controlled processes are thus triggered when race makes the jump from covert to overt, and these

[2] It is important to note, here, that the amygdala has been shown to be quite flexible in terms of its functions, and is most appropriately thought of as driving affective or evaluative processing goals rather than threat or danger, per se (Cunningham, Van Bavel, and Johnsen 2008). As such, we want to be careful in this manuscript to emphasize that amygdala activation does not necessarily indicate negative affect, but instead, detection of affective significance.
[3] Note that the role of the insula can be automatic yet this region is still associated with conflict detection. As such, we consider the insula as occupying a sort of "middle ground" between automatic and controlled processes.

processes can allow people to inhibit prejudiced reactions. This basically means the precise mechanisms assumed to be occurring under the framework of existing work in political science are detectable when observing brain data. Overall, we are left with relatively distinct patterns of neural activity when someone reacts adversely to a racial outgroup member, detects a conflict between this adverse reaction and one's conscious goals regarding race, and subsequently attempts to supersede adverse reactions through controlled processing. Note that conflict detection and controlled processing to inhibit adverse reactions should only be evident when race is salient enough so as to be consciously recognized; otherwise, automatic reactions should occur uninhibited (Cunningham et al. 2004).

Critically, the research on the aforementioned neural processes does not suggest that controlled processing in the PFC will *necessarily* yield inhibition of prejudiced reactions – just that it is necessary for such inhibition to be possible. An alternative possibility, which has not been explored thoroughly in the neuroscience literature let alone in political science, is that controlled processes ultimately lead individuals to reinforce their pre-existing biases as a form of rationalization. In terms of the brain, PFC activation may be necessary for inhibition to occur, but in some situations, it is reasonable to expect that it may act as a catalyst for "doubling down" on deep-seated racial animus. The implication for reducing the role of prejudice is glaring: encouraging people to think more about the racial component of race-targeted policies will not diminish the influence of prejudice and may even bolster its influence.

Expectations

In line with the research described thus far, it should be expected that when race is primed covertly, individuals will exhibit greater automatic processing aimed at detecting emotional significance (activation in amygdala), and when race is primed overtly, individuals should exhibit greater conflict detection (activation in ACC and insula) and controlled processing (activation in orbitofrontal cortex[4] and prefrontal cortex). These expectations are in line with the research in social neuroscience on race processing (e.g. Cunningham et al. 2004; Shkurko 2012) and are implied by the political science research on the effects of covert versus overt racial cues.

With regard to the consequences of controlled processing, we adjudicate between two competing hypotheses. According to one hypothesis, the controlled processing associated with conscious racial evaluations will result in diminished anti-Black racial biases in support for government assistance compared to when evaluations are only indirectly racial (i.e. when race is cued covertly). Critically, this should manifest "across the board" ideologically – that is, conscious racial evaluations and controlled processing should be associated with reduced anti-Black biases in support for government assistance for liberals as well as conservatives. If this is the case, then the reason liberals and conservatives disagree on racial policy issues may be because liberals experience relatively high levels of controlled processing, which allows them to inhibit their racial biases when evaluating racial policies, but conservatives simply don't experience enough controlled processing to prevent their biases from driving their policy evaluations. Said differently, if racial biases influence conscious political evaluations, it is because of a *lack of* controlled processing. According to the alternative hypothesis, the reason for differences in the degree to which racial biases influence political evaluations is differences in the *function of* controlled processing. That is, both liberals and conservatives may experience controlled processing during conscious racial policy evaluations, but controlled processing will be associated with reduced anti-Black bias among liberals, but exacerbated anti-Black bias among conservatives.

[4] The orbitofrontal cortex (OFC) has not been shown to be activated "only" at the conscious level and can sometimes be automatic, but is nonetheless implicated in evaluative processes and decision-making rather than affective processes like the amygdala. We use the term "controlled processing" as an umbrella term that encompasses both evaluative processes (OFC) and executive function (PFC).

Data and Methods

Twenty-seven White adults were recruited to participate in the study through flyers placed around the community surrounding a large Midwestern university, as well as a volunteer participant registry created by the facility housing the MRI. Participants were compensated $25. All participants were right-handed and had normal or corrected-to-normal vision. Participants were safety screened to ensure eligibility for MRI and provided informed consent in accord with study approval by the Institutional Review Board. Seventeen participants were female, 19 were Christian, and all were U.S.-born with English as a first language. The median age was 23 years old (range: 19 to 64) and 13 participants had a 4-year college degree or higher. We used some degree of quota sampling to ensure the sample was adequately diverse in terms of political ideology and partisan affiliation. Nonetheless, the sample exhibited a liberal/Democratic skew (16 participants identified as liberal or strongly liberal, 1 identified as moderate, and 10 identified as conservative or strongly conservative; 13 participants identified as Democrats, 2 identified as Independent, 10 identified as Republican, and 2 identified as “something else”).

Experimental Design

The experimental task used a 2 (*race of applicant*: White vs. Black) x 2 (*conscious awareness*: non-conscious vs. conscious) rapid event-related within-subject design consisting of 4 blocks. *Race of applicant* varied at the trial level (i.e. within blocks) and *conscious awareness* varied at the block level (i.e. between blocks). There were 40 trials within each block for a total of 160 trials. Each block lasted approximately 7 minutes. Experimental stimuli were presented in the scanner using PsychoPy2 v1.83.01 (Peirce 2007; 2009). Both trial order and interstimulus interval (ISI) duration were predetermined using Optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq), a software package that maximizes efficiency in modeling the hemodynamic response in rapid event-related designs (see Burock, Buckner, Woldorff, Rosen, and Dale, 1998; Dale, Greve, and Burock, 1999).

Each trial presented participants with a random piece of information about the applicant and asked them to decide whether they strongly oppose, oppose, support, or strongly support monetary aid to that individual using response pads (one response pad for each hand, with buttons under each of participants' pointer and middle fingers). Response options were shown at the top of the screen throughout the task, and their positions on the screen remained constant across trials. Descriptions were constructed in such a way that they would not explicitly indicate any sort of deservingness or merit on the part of the applicant (e.g. "has an outie belly button" or "hums when eating"; see Appendix for complete list of statements). All applicants were males in order to avoid any effects of applicant sex. The phrase "A man that:" was shown on the screen just below the response options at all times.

Prior the description of each individual, a Black or White face was presented for an amount of time that either allowed for conscious recognition of the face (*Conscious* blocks; 1000 ms) or not (*Nonconscious* blocks; 30 ms). The order of *Conscious* and *Nonconscious* blocks was randomized for each participant. All faces were taken from the Chicago Face Database (Ma, Correll, and Wittenbrink, 2015), a database of high-resolution standardized photographs of 158 Black and White individuals between the ages of 18-40. Each image has been coded for 41 physical characteristics (e.g. nose width, face width to height ratio, eye closeness) and rated along 22 perceptual dimensions (e.g. threatening, attractive, masculine) by an independent sample of raters, and was matched as closely as possible across these characteristics and dimensions. Only 80 images were used (40 Black male faces and 40 White male faces), and so each image was shown twice (once during a Conscious block and once during a Nonconscious block). Images were randomly paired with applicant descriptions for each participant.

Each trial began with a fixation cross shown for 500 milliseconds (ms). In Conscious blocks, this was followed by an image of a Black or White face for 1 second, and then an applicant description for 3.5 seconds. In Nonconscious blocks, the fixation cross was followed by a noise mask (a picture of random black and white pixels) for 934 ms, a Black or White face for 30 ms, a second noise mask for 36 ms, and then an applicant description for 3.5 seconds. As such, Conscious and Nonconscious trials were identical except that whereas Conscious trials showed a Black or White face for a full second before each applicant description, Nonconscious trials showed noise masks during that second with a Black or White face shown rapidly in between masks. Thus, participants were expected not to be consciously aware of the influence of the Black or White face in the Nonconscious trials. Prior research has explored similar protocols and shown that a short stimulus presentation time such as the one being used here for the nonconscious face primes is sufficient to yield only nonconscious recognition of a stimulus (see Rohr, Degner, and Wentura 2015). The exact timings used for this experiment were pre-tested on a small sample of undergraduate students as well as during pilot runs in the fMRI scanner.

Survey Measures

Participants were in the scanner for approximately 45 minutes to 1 hour. Before being removed from the scanner, participants were asked a series of questions about whether they were able to see faces during the Nonconscious blocks (8 participants reported being able to see faces, and 11 said they were able to make out the race of faces seen between noise masks). After scanning, they completed a 15-minute long survey, which included, among other things, a 5-point self-report scale measuring political ideology (with higher values indicating more conservative), a three-item survey battery asking the degree to which participants supported various race-targeted policies (affirmative action, race-targeted education spending, and race-based scholarships; $\alpha = 0.92$), and demographic items. For MRI analyses using ideology, the measure was divided into three groups, with anyone marking below a 3 being categorized as "liberal," anyone with a score of 3 being categorized as "moderate," and anyone marking above a 3 being categorized as "conservative."

MRI Data Acquisition

A Siemens Skyra 3.0 Tesla MRI with a 32-channel head coil was used to collect brain activation data. Prior to functional imaging, a high-resolution T1-weighted 3D anatomical image (MPRAGE; *field of view (FoV) read* = 256 millimeters, *slice thickness* = 1.0 x 1.0 x 1.0 mm, r*epetition time (TR)* = 2400 ms, *echo time (TE)* = 3.37 ms, *inversion time (TI)* = 991 ms, *prescan normalize on*, *PAT mode GRAPPA*) was obtained for spatial normalization. Functional MRI data was acquired with acquisition parallel to the AC-PC line to maximize whole-brain coverage (42 slices, *FoV read* = 220 mm, *slice thickness* = 3.0 x 3.0 x 3.0 mm, *TR* = 2500 ms, *TE* = 30 ms, *flip angle* = 80 degrees, *prescan normalize off*). The first 12.5 seconds of each block showed a blank grey screen so that the first five volumes of each block could be discarded to avoid variability due to pre-steady state functional data.

MRI Data Preprocessing and Analysis

MRI data were preprocessed using fMRI Expert Analysis Tool (FEAT) in FMRIB Software Library (FSL; Jenkinson, Beckmann, Behrens, Woolrich, and Smith 2012; Smith et al. 2004) on Mac OS X. The MPRAGE image was skull stripped using FSL's Brain Extraction Function (BET; Smith 2002). Functional data were subjected to normalization, registration to both MPRAGE and standard space (MNI152), spatial smoothing at FWHM of 5mm, slice timing correction (to correct for interleaved data acquisition), and motion correction using MCFLIRT (Jenkinson, Bannister, Brady, and Smith 2002).

General linear models (GLMs) were used as implemented in FSL. Time-series data was modeled at the first level (the trial level) using FMRIB's Improved Linear Model (FILM), and then higher-level analysis (across blocks first, and then across participants) was done using FMRIB's Local Analysis of Mixed Effects (FLAME; see Smith et al. 2004). The blood oxygen level-dependent (BOLD) signal was first modeled at the block-level, as our hypotheses were focused on conscious versus nonconscious racial evaluations rather than evaluations of Black versus White applicants, per se. BOLD signal differences were modeled as a function of *conscious awareness* (Conscious/Nonconscious). The subject-level analyses were then combined into group-level region of interest (ROI) analyses using FSL FLAME1. ROI analyses on left amygdala, right amygdala, bilateral insula, anterior cingulate cortex (ACC), frontal orbital cortex (OFC), and dorsolateral prefrontal cortex (dlPFC) were masked prior to analysis. All masks except the one used for dlPFC were created using anatomical masks from the Harvard-Oxford Cortical/Subcortical Atlases provided with FSL. The dlPFC mask was created by combining the Harvard-Oxford Cortical Atlas masks for the frontal pole and ventromedial prefrontal cortex. Due to the small sample size, these analyses remained uncorrected for multiple comparisons.[5]

Results

Behavioral Task Results

Before turning to the fMRI results, analyses of participants' responses to the task were done to determine whether support for government assistance was influenced by race, conscious awareness, and ideology in the expected directions. Multilevel models were run with trial as the

[5] As such, the results presented in this chapter are preliminary and should not be cited without consulting the authors.

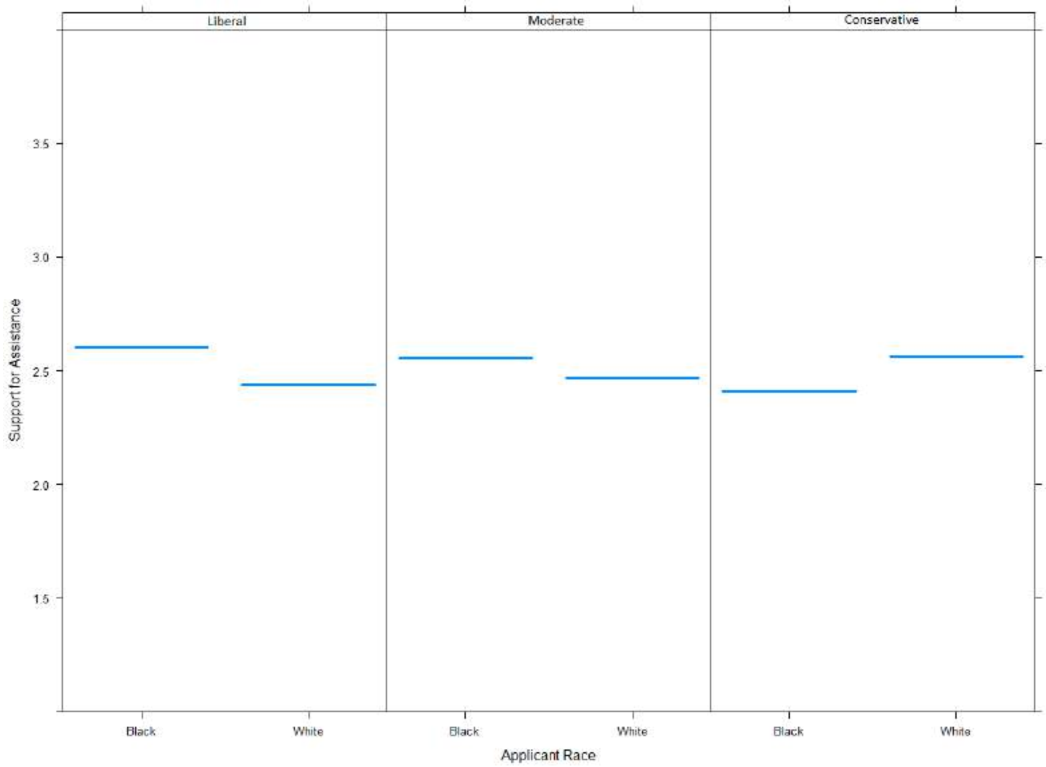
unit of analysis nested within blocks, nested within participant. An empty model was run estimating random effects for participant, block, trial number, and face image (because each image was shown twice). Only the random effects for participant explained a substantial amount of variation in support for government assistance (approximately 10% of the variance), and so all subsequent models include only random effects for participant. All models also controlled for fixed effects of participant's sex, education, religion (Christian or non-Christian), and whether participants reported being able to decipher the race of faces shown in between noise masks in the Nonconscious blocks. Data for two participants are not included in the behavioral analyses due to a technical error.

First, we examined the relationship between racial biases in support for government assistance during the task and support for race-based policies as measured during the post-scan survey in order to determine the degree to which the task reflected evaluations of "real-world" policies. Support for government assistance during the task was regressed on a dummy variable for *conscious awareness* condition (0 = Conscious, 1 = Nonconscious), a dummy variable for race of the applicant (0 = Black, 1 = White), race-based policy support (mean-centered), and the interaction between race of applicant and race-based policy support. Indeed, there was a significant interaction between race of applicant and race-based policy support ($b = -0.047$, $SE = 0.018$, $p < .05$). Once decomposed, it was evident that at low levels of support for race-based policies (-1 SD from the mean), there were no significant racial biases in support for government assistance during the task ($b = 0.025$, $SE = 0.042$, $p = .544$), but at high levels of support for race-based policies (+1 SD from the mean), there was a significant pro-Black bias ($b = -0.125$, $SE = 0.039$, $p < .01$). This suggests that although opposition to race-based policies was not associated with anti-Black bias during the task, support for race-based policies was indeed associated with pro-Black attitudes during the task.

The primary behavioral analyses concern the interaction between race of applicant and conscious awareness. A main effects model indicated no significant main effect of conscious awareness, but there was a significant main effect of applicant race such that on average, participants were less likely to support White applicants than Black applicants (b = -0.055, SE = 0.28, $p < .05$). Next, an interaction was calculated between applicant race and conscious awareness. The interaction was marginally significant (b = 0.103, SE = 0.055, p = .063), and showed that although a pro-White racial bias existed in Conscious blocks (b = -0.093, SE = 0.043, $p < .05$), no significant racial bias existed in Nonconscious blocks (b = 0.010, SE = 0.042, p = 0.813). Thus, although the lack of racial bias in Nonconscious blocks was unexpected, more racially tolerant evaluations were observed in Conscious blocks as hypothesized.

The critical, behavioral analysis relating to the hypotheses about the consequences of conscious racial evaluations and controlled processing involved a three-way interaction between applicant race, conscious awareness, and ideology. If the consequences of conscious racial evaluations depend are different from liberals than for conservatives, then we should expect the pro-Black patterns of support in Conscious trials to be concentrated among liberals, and perhaps anti-Black patterns of support among conservatives in Conscious trials. However, no significant three-way interaction existed (b = 0.019, SE = 0.042, p = 0.655). Yet, a two-way interaction was significant between race and ideology such that among strong liberals ("1"s on the 5-point scale), there was a significant pro-Black bias in support (b = -0.167, SE = 0.046, $p < .001$), and among strong conservatives ("5"s on the 5-point scale), there was a significant anti-Black bias in support (b = 0.154, SE = 0.060, $p < .05$). Thus, although it did not depend on conscious awareness, liberals tended to express pro-Black patterns of support while conservatives tended to express anti-Black patterns of support. This interaction is illustrated in Figure 1.

Figure 1: Interaction between Applicant Race and Ideology



The y-axis indicates support for government assistance, and applicant race is plotted on the x-axis. The left panel indicates liberals, the middle panel indicated moderates, and the right panel indicates conservatives.

fMRI Results

Three sets of analyses were conducted. First, BOLD signal was modeled as a function of conscious awareness to see which brain regions exhibited significant activation for Conscious (compared to Nonconscious) blocks as well as Nonconscious (compared to Conscious) blocks. Second, we constructed a between-subjects *conscious racial bias inhibition* (CRBI) variable to test how controlled processes related to differences in racial biases in support for government assistance between Conscious and Nonconscious trials. The CRBI variable was constructed as follows: a racial bias score was calculated by subtracting the average level of support for Black applicants from the average level of support for White applicants *separately* for Conscious and Nonconscious trials, and then subtracting the racial bias score for Conscious trials from the racial bias score for Nonconscious trials. As such, the CRBI score gauged the degree to which participants were *more* racially biased during Nonconscious trials than Conscious trials (or, the degree to which participants were *less* racially biased during Conscious trials than Nonconscious trials). Said differently, this variable reflected the degree to which the existing literature's hypotheses regarding conscious racial evaluations were true for any given individual. We modeled BOLD signal as a function of the interaction between applicant race, conscious awareness, and CRBI to determine which brain regions exhibited significant activation for Conscious (compared to Nonconscious) blocks among individuals whose anti-Black racial biases were diminished during Conscious (compared to Nonconscious) trials. Third, in order to examine the role of ideology, we conducted the same analyses but separately for liberals and conservatives to see if the consequences of controlled processing during Conscious trials varied across the ideological spectrum. ROI analyses were done for each contrast for amygdala (left and right separately), insula, ACC, OFC, and dlPFC in order to limit analyses to only the brain regions implicated in the critical tests of this paper.

Table 1 contains the results of the analyses of brain activation in Conscious (compared to Nonconscious) trials. As expected, and in line with the assumptions of existing literature, there were significant clusters of activation in brain regions associated with conflict detection controlled processing during Conscious (compared to Nonconscious) trials. Some significant activation in these regions was also associated with Nonconscious (compared to Conscious) trials, but these clusters were very small, especially in comparison to the moderate-large sized clusters associated with Conscious trials. Specifically, substantially sized clusters in the insula, ACC, OFC, and dlPFC were significantly active. The largest and strongest cluster was in dlPFC, which is the ROI most associated with executive function.

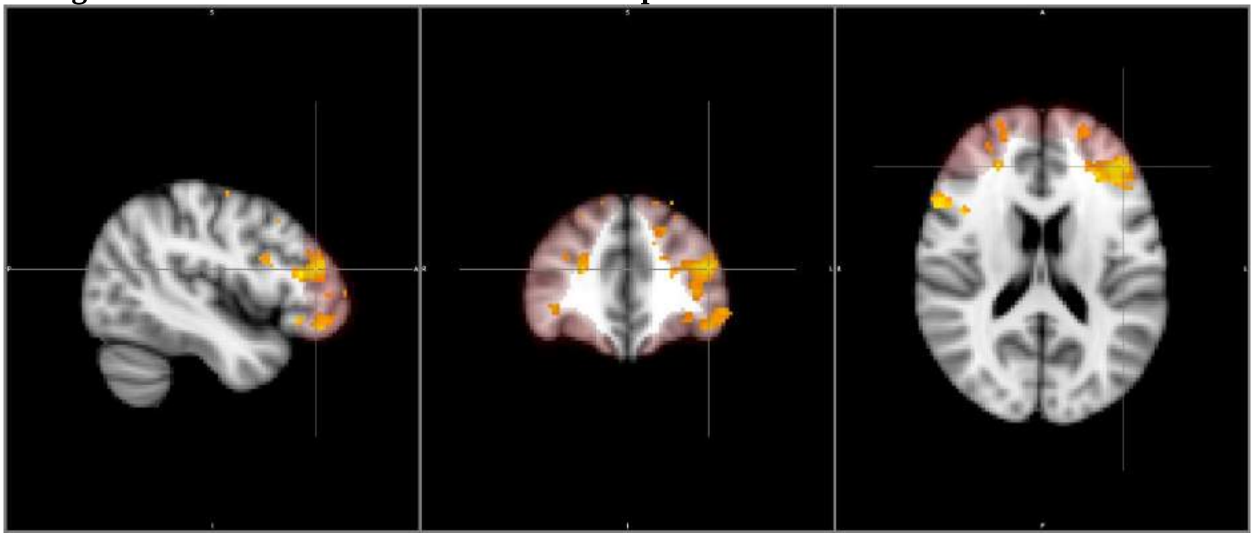
Table 1: Significant Clusters of Bold Activation in ROIs for Conscious vs Nonconscious Trials

Contrast	Anatomical Label(s)	Cluster Size (# Voxels)	Peack Activation (Z-Score)	*p*-value	X	Y	Z
Left Amygdala							
Noncon. > Con.	62% Left Cerebral Cortex, 23% Left Cerebral White Matter, 9% Left Amygdala	0	0	>0.999	57	62	20
Con. > Noncon.	*same as above*	0	0	>0.999	57	62	20
Right Amygdala							
Noncon. > Con.	86% Right Cerebral White Matter, 9% Right Amygdala	2	1.997	0.046	29	63	22
Con. > Noncon.	51% Right Cerebral White Matter, 39% Right Amygdala	10	2.273	0.023	32	58	31
Insula							
Noncon. > Con.	30% Insula, 14% OFC	11	1.945	0.052	29	68	27
Con. > Noncon.	88% Insula	395	2.678	0.007	64	68	32
ACC							
Noncon. > Con.	16% ACC	1	1.671	0.095	44	78	39
Con. > Noncon.	40% ACC	360	2.987	0.003	47	61	52
OFC							
Noncon. > Con.	35% OFC, 12% Temporal Pole	40	2.302	0.021	32	70	23
Con. > Noncon.	28% OFC, 21% Inferior Fonrtal Gyrus, pars triangularis	358	2.698	0.007	20	75	33
dlPFC							
Noncon. > Con.	36% Frontal Pole, 11% Middle Frontal Gyrus	22	1.972	0.049	28	83	58
Con. > Noncon.	42% Frontal Pole, 21% Middle Frontal Gyrus	1438	3.252	0.001	67	82	45

"Noncon. > Con." rows are clusters that were more active in Nonconscious than Conscious trials, and vice versa for "Con. > Noncon."; Anatomical labels are based on the Harvard-Oxford Cortical and Subcortical atlases; cluster size indicates the number of voxels of the cluster, peak activation indicates the z-score for peak activation within the cluster; X, Y, and Z indicate the coordinates of the peak activation in MNI152 space.

Contrary to some expectations from the literature, we did not find evidence of brain activation in amygdala that were more active in Nonconscious than Conscious trials (except for a very small cluster in the right amygdala, which is smaller than the cluster associated with Conscious trials). Figure 2 illustrates the significant cluster of activation in dlPFC for Conscious (compared to Nonconscious) trials.

Figure 2: BOLD Activation in dlPFC in Response to Conscious > Nonconscious Trials



Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for each cluster from the ROI analyses. Areas highlighted in red indicate the ROI for the analysis, and yellow/orange clusters indicate regions of significant activation.

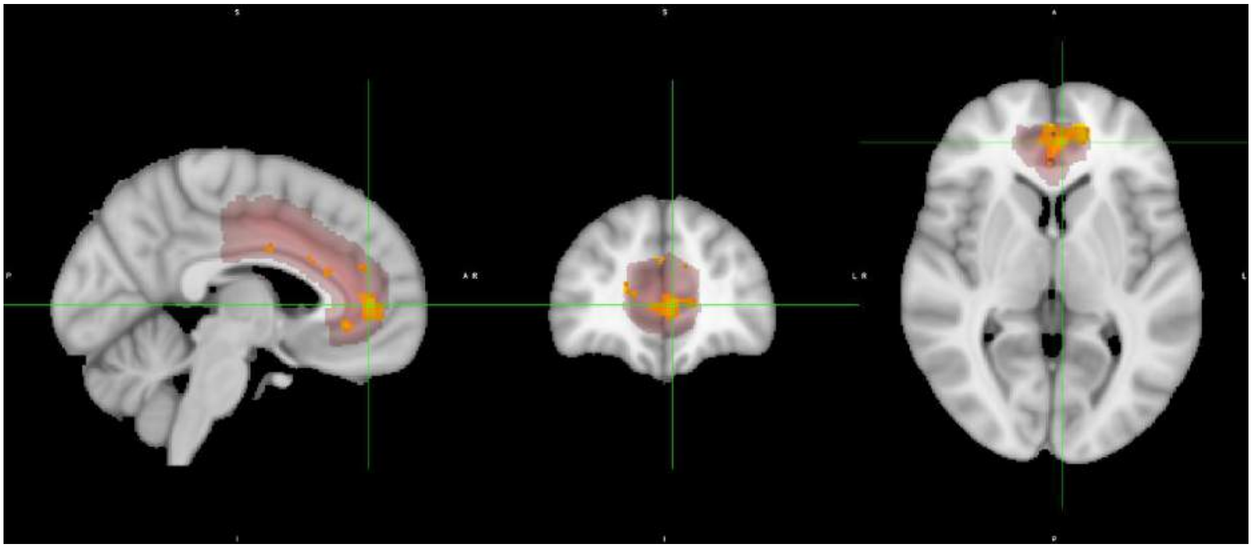
Table 2 contains the results of this same contrast but with CRBI added to the analysis. Only small clusters of activation in amygdala were associated with CRBI. However, substantial clusters of activation in insula, ACC, and OFC were associated with reduced racial biases in support for government assistance in the Conscious (compared to Nonconscious) trials. Thus, in line with the expectations of the existing literature, individuals who exhibited diminished anti-Black racial biases in the Conscious over Nonconscious trials were more likely to have activation in brain regions associated with conflict detection and controlled processing. Notably, though, this was not the case regarding dlPFC. Further, the cluster in OFC was smaller than those in insula and ACC, and so the diminished anti-Black racial biases seem driven to some degree more by regions associated with conflict detection than by regions associated with executive function or controlled processing. Figure 3 shows the significant cluster of activation in ACC associated with reduced anti-Black biases in Conscious over Nonconscious trials.

Table 2: Significant Clusters of Bold Activation in ROIs for Conscious Awareness*Conscious Racial Bias Inhibition Contrast

Contrast	Anatomical Label(s)	Cluster Size (# Voxels)	Peack Activation (Z-Score)	p-value	X	Y	Z
Left Amygdala							
< Bias in Con.	*76% Left Cerebral White Matter, 22% Left Amygdala*	18	2.366	0.018	62	62	26
> Bias in Con.	*63% Left Amygdala, 32% Left Hippocampus*	13	2.019	0.043	53	60	26
Right Amygdala							
< Bias in Con.	*67% Right Cerebral White Matter, 16% Right Cerebral Cortex, 15% Right Amygdala*	2	1.674	0.094	30	64	22
> Bias in Con.	*49% Right Cerebral Cortex, 44% Right Amygdala*	2	1.801	0.072	32	65	25
Insula							
< Bias in Con.	*41% Insula, 30% OFC*	209	2.455	0.014	25	72	31
> Bias in Con.	*20% Planum Polare, 14% Insula*	0	0	>0.999	25	72	31
ACC							
< Bias in Con.	*52% ACC, 42% Paracingulate Gyrus*	298	2.531	0.011	47	85	38
> Bias in Con.	*12% ACC*	3	1.878	0.060	45	58	49
OFC							
< Bias in Con.	*41% Insula, 30% OFC*	75	2.455	0.014	25	72	31
> Bias in Con.	*60% Temporal Pole, 16% OFC*	14	2.488	0.013	64	73	23
dlPFC							
< Bias in Con.	*49% OFC*	1	1.703	0.089	25	77	32
> Bias in Con.	*35% Temporal Pole, 11% OFC*	1	1.948	0.051	64	75	23

"< Bias in Con." rows are clusters that were more active among individuals who exhibited reduced anti-Black racial biases in Conscious over Nonconscious trials; "> Bias in Conscious" rows are the same but for individuals who exhibited increased anti-Black racial biases in Conscious over Nonconscious trials; Anatomical labels are based on the Harvard-Oxford Cortical and Subcortical atlases; cluster size indicates the number of voxels of the cluster, peak activation indicates the z-score for peak activation within the cluster; X, Y, and Z indicate the coordinates of the peak activation in MNI152 space.

Figure 3: BOLD Activation in ACC for Individuals High in Conscious Racial Bias Inhibition for Conscious > NonconsciousTrials



Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for each cluster from the ROI analyses. Areas highlighted in red indicate the ROI for the analysis, and areas highlighted in yellow/orange indicate regions of significant activation.

Thus far, the MRI results have largely supported the assumptions and expectations of the existing literature, but with some caveats. During evaluations regarding applicants for government assistance, brain regions associated with conflict detection (insula, ACC) and controlled processing (OFC, dlPFC) indeed seem to be more active when race is conscious compared to when race is nonconscious. However, the converse – that regions associated with automatic, affective processes would be more active when race is nonconscious – is not supported. When it comes to the relationship between conflict detection/controlled processing and racial biases in support for government assistance, brain regions associated with conflict detection are especially active during Conscious (compared to Nonconscious) trials among individuals who exhibit a reduction in anti-Black racial bias. The same goes for controlled processing in OFC, but the cluster in OFC nearly overlapped with insula, and there was no notable activation in dlPFC.

In order to test whether the outcomes of controlled processing vary across the ideological spectrum, we ran the above analyses on liberals and conservatives separately.[6] Table 3 contains the results for liberals only. We find only small (few voxels), statistically marginal areas of activation in the amygdala that are associated with differences in racial biases between Nonconscious and Conscious trials. As with the aggregate results, we find significant clusters of activation in regions associated with conflict detection (insula and ACC) to be related to a reduction in racial biases during Conscious over Nonconscious trials. The results are mostly similar regarding OFC, except that among liberals, there is also a significant cluster of activation associated with *exacerbation* of racial biases during Conscious (compared to Nonconscious) trials. The same goes for a significant cluster in dlPFC as well. As such, counter to any of our hypotheses, conflict detection among liberals seems associated with inhibition of racial biases in support for government assistance but the consequences of controlled processing are mixed. Figure 4 shows the cluster of activation in ACC associated with diminished racial biases among liberals, and Figure 5 shows the cluster of activation in dlPFC associated with exacerbation of racial biases.

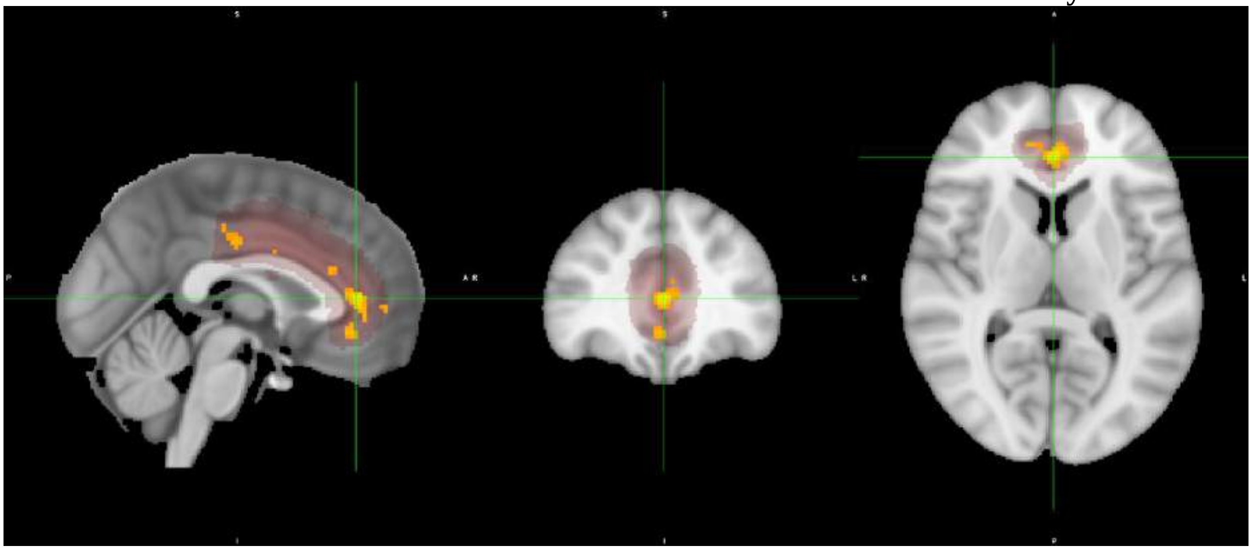
[6] Note that there were only 8 conservatives in these analyses. However, the within-subject design means that these analyses are based on 1,280 trials, and the MRI analyses are done using a multilevel procedure (FLAME) that takes into account variation at the participant level.

Table 3: Significant Clusters of Bold Activation in ROIs for Conscious Awareness*Conscious Racial Bias Inhibition Contrast – Liberals Only

Contrast	Anatomical Label(s)	Cluster Size (# Voxels)	Peack Activation (Z-Score)	*p*-value	X	Y	Z
Left Amygdala							
< Bias in Con.	*41% Left Amygdala, 29% Left Cerebral Cortex*	2	1.730	0.084	55	61	23
> Bias in Con.	*14% Left Amygdala, 10% Left Cerebral White Matter*	1	1.849	0.064	51	58	29
Right Amygdala							
< Bias in Con.	*40% Right Amygdala, 39% Right Cerebral White Matter*	3	1.916	0.055	35	61	31
> Bias in Con.	*53% Right Cerebral Cortex, 28% Right Cerebral White Matter, 9% Right Amygdala*	0	0	>0.999	32	62	20
Insula							
< Bias in Con.	*51% OFC, 19% Insula*	98	2.223	0.026	62	73	31
> Bias in Con.	*35% Planum Polare, 20% Temporal Pole, 12% Insula*	2	1.724	0.085	25	65	26
ACC							
< Bias in Con.	*81% ACC*	198	2.717	0.007	45	82	40
> Bias in Con.	*14% ACC*	1	1.777	0.076	45	57	49
OFC							
< Bias in Con.	*22% Frontal Pole, 14% OFC, 14% Frontal Medial Cortex*	109	2.379	0.017	51	82	26
> Bias in Con.	*36% Frontal Pole, 14% OFC*	50	2.389	0.017	27	81	32
dlPFC							
< Bias in Con.	*22% Frontal Pole, 14% OFC, 14% Frontal Medial Cortex*	4	2.379	0.017	51	82	26
> Bias in Con.	*36% Frontal Pole, 14% OFC*	49	2.389	0.017	27	81	32

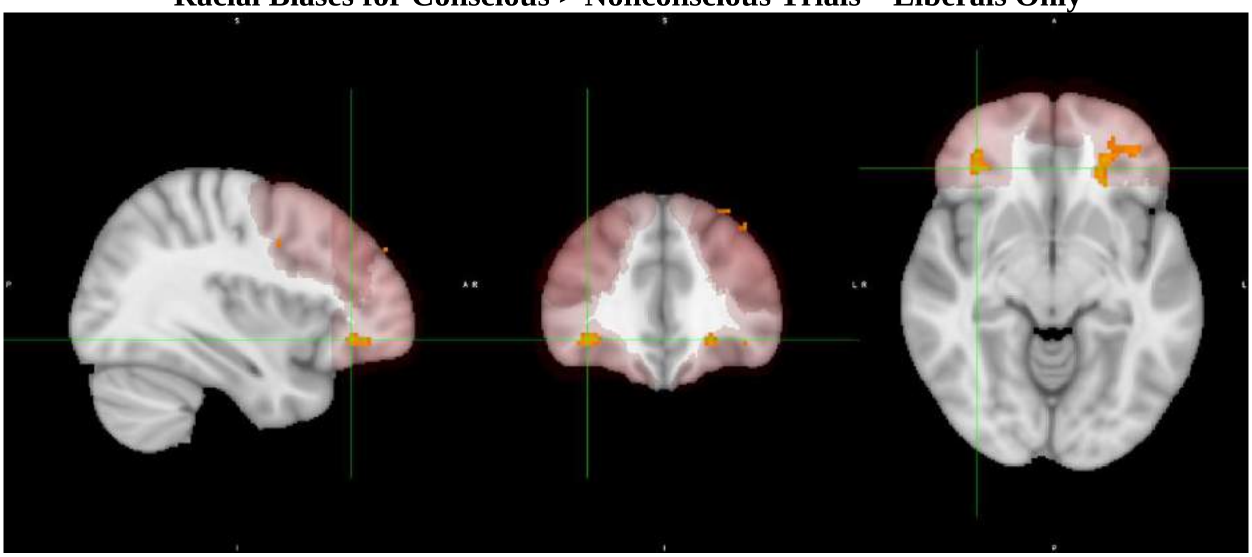
"< Bias in Con." rows are clusters that were more active among individuals who exhibited reduced anti-Black racial biases in Conscious over Nonconscious trials; "> Bias in Con." rows are the same but for individuals who exhibited increased anti-Black racial biases in Conscious over Nonconscious trials; Anatomical labels are based on the Harvard-Oxford Cortical and Subcortical atlases; cluster size indicates the number of voxels of the cluster, peak activation indicates the z-score for peak activation within the cluster; X, Y, and Z indicate the coordinates of the peak activation in MNI152 space.

Figure 4: BOLD Activation in ACC Associated with Decreased Likelihood of Anti-Black Racial Biases for Conscious > Nonconscious Trials – Liberals Only



Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for each cluster from the ROI analyses. Areas highlighted in red indicate the ROI for the analysis, and areas highlighted in yellow/orange indicate regions of significant activation.

Figure 5: BOLD Activation in ACC Associated with Increased Likelihood of Anti-Black Racial Biases for Conscious > Nonconscious Trials – Liberals Only



Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for each cluster from the ROI analyses. Areas highlighted in red indicate the ROI for the analysis, and areas highlighted in yellow/orange indicate regions of significant activation.

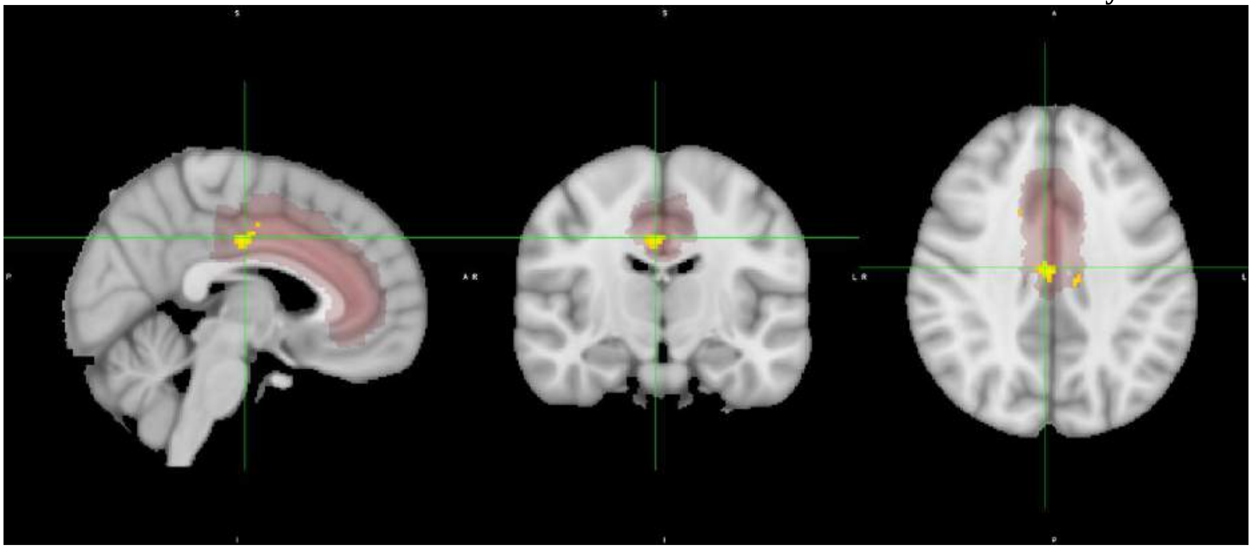
Finally, we turn to the same analyses but with only conservatives instead of liberals. Table 4 contains the results of these analyses. In general, any significant clusters of activation in these analyses were notably smaller and weaker than clusters found among liberal participants. The largest and strongest clusters were in ACC and OFC, and were associated with *increased* racial biases in Conscious (compared to Nonconscious trials). Figure 6 illustrates the significant cluster in ACC associated with exacerbation of racial biases among conservatives for Conscious over Nonconscious trials.

Table 4: Significant Clusters of Bold Activation in ROIs for Conscious Awareness*Conscious Racial Bias Inhibition Contrast – Conservatives Only

Contrast	Anatomical Label(s)	Cluster Size (# Voxels)	Peack Activation (Z-Score)	*p*-value	X	Y	Z
Left Amygdala							
< Bias in Con.	*50% Left Amygdala, 47% Left Cerebral White Matter*	4	1.767	0.077	61	61	26
> Bias in Con.	*72% Left Amygdala, 27% Left Hippocampus*	16	2.050	0.040	54	60	26
Right Amygdala							
< Bias in Con.	*53% Right Cerebral Cortex, 28% Right Cerebral White Matter, 9% Right Amygdala*	0	0	>0.999	32	62	20
> Bias in Con.	*65% Right Hippocampus, 16% Right Amygdala*	5	1.857	0.063	31	58	27
Insula							
< Bias in Con.	*63% Insula*	6	1.773	0.076	66	60	34
> Bias in Con.	*21% Planum Polare, 15% Heschl's Gyrus, 14% Insula*	6	1.785	0.074	64	51	36
ACC							
< Bias in Con.	*27% Paracingulate Gyrus, 14% ACC*	4	1.775	0.076	51	75	50
> Bias in Con.	*39% ACC, 37% posterior Cingulate Gyrus*	40	2.126	0.033	43	56	54
OFC							
< Bias in Con.	*39% OFC, 14% Temporal Pole*	7	1.825	0.068	59	68	24
> Bias in Con.	*62% Temporal Pole, 17% OFC*	39	1.968	0.049	65	73	24
dlPFC							
< Bias in Con.	*49% Temporal Pole, 11% OFC*	0	0	>0.999	60	73	20
> Bias in Con.	*49% Temporal Pole, 11% OFC*	0	0	>0.999	60	73	20

"< Bias in Con." rows are clusters that were more active among individuals who exhibited reduced anti-Black racial biases in Conscious over Nonconscious trials; "> Bias in Con." rows are the same but for individuals who exhibited increased anti-Black racial biases in Conscious over Nonconscious trials; Anatomical labels are based on the Harvard-Oxford Cortical and Subcortical atlases; cluster size indicates the number of voxels of the cluster, peak activation indicates the z-score for peak activation within the cluster; X, Y, and Z indicate the coordinates of the peak activation in MNI152 space.

Figure 5: BOLD Activation in ACC Associated with Increased Likelihood of Anti-Black Racial Biases for Conscious > Nonconscious Trials – Conservatives Only



Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for each cluster from the ROI analyses. Areas highlighted in red indicate the ROI for the analysis, and areas highlighted in yellow/orange indicate regions of significant activation.

Inside the Black Box of Overtly Racial Policy Evaluations

A good deal of research in political science has been focused on when racial biases are most likely to influence political attitudes, how covert racial cues can tinge otherwise nonracial policy evaluations, and when people are most capable of inhibiting their racial biases and abiding by the egalitarian norms of society (e.g. Huber and Lapinski 2006; Hurwitz and Peffley 2005; Mendelberg 2001; Valentino, Hutchings, and White 2002). The overarching consensus (but see Huber and Lapinski 2006) has been that racial biases are most likely to influence political attitudes when racial cues are covert because this allows racial prejudice to be triggered outside of the control of conscious awareness. When race is cued overtly, the story goes, people are able to recognize the attempt at triggering a prejudiced reaction (conflict detection), inhibit their racial biases (controlled processing), and express egalitarian views. We have argued that two assumptions of this model deserve testing.

First, the idea that overtly racial political evaluations trigger conflict detection and controlled processing has yet to be directly tested, largely because of the tremendous cost of directly examining neural activity and because the field of social neuroscience is new enough that it would be difficult to know what activation in particular areas of the brain means. Yet the growth of social neuroscience research on how individuals process race has given us an outline for what we should expect when people have an automatic, affective reaction regarding a racial outgroup, when they recognize the conflict between this automatic reaction and their conscious egalitarian values, and when they utilize executive functioning to “override” racial biases (Cunningham et al. 2004; Rilling et al. 2008; Risheson et al. 2003; Shkurko 2012; Stanley, Phelps, and Banaji 2008; Van Bavel, Packer, and Cunningham 2008). In this paper, we utilized this neuroscientific framework to make predictions about what we should see in the brain when, as predicted by the political science literature, individuals inhibit their racial biases in the face of overtly racial political evaluations.

Our results did not provide evidence for automatic neural processes or racial biases associated with nonconscious racial evaluations, but our findings regarding conscious racial evaluations supported the framework of the existing literature. The behavioral task results suggested that individuals were, as expected, less likely to express bias against a racial outgroup when racial primes were conscious than when they were nonconscious. However, contrary to the hypothesis that anti-Black racial biases would be strongest during nonconscious blocks, anti-Black racial biases in support for government assistance were observed only among conservatives and this did not depend on conscious awareness. This could have been due to multiple reasons. First, although brain activation patterns indicated that participants did respond differently to Black than White faces in nonconscious blocks regardless of ideology, it could be the case that the primes were shown too fast for all of the subsequent behavioral consequences to manifest. This would be somewhat surprising given the pilot testing as well as the fact that quicker primes have been used in prior research to produce racial biases (e.g. Rohr, Degner, and Wentura 2015), and a number of participants reported seeing faces (and even being able to distinguish the races of faces) during Nonconscious trials. Another possibility is simply limited sample size. Although the within-subjects design allowed for a large sample in terms of trials, we could have a problem of restriction of range – the limited between-subject variation may have made it difficult to detect anti-Black biases among certain individuals. Nonetheless, pro-Black biases were evident only in conscious blocks, which suggests that on average, overt racial primes led to more pro-Black evaluations, and conservatives did express anti-Black biases in support while liberals expressed pro-Black biases regardless of whether the cues were overt or covert.

The fMRI results did not show evidence of greater amygdala activity in Nonconscious (compared to Conscious) trials, but conflict detection (insula and ACC) and controlled processing (OFC and dlPFC) were significantly more active in Conscious than Nonconscious trials, as expected. Further, conflict detection and, to a lesser degree, controlled processing, were associated with exhibiting diminished anti-Black racial biases in support for government assistance in Conscious (compared to Nonconscious) trials. In the aggregate, these findings provide the first direct evidence of the psychological mechanisms theorized to be at work by various scholars studying the influence of racial prejudice on political attitudes. However, they also call into question assumptions about what happens psychologically when racial cues are covert.

The second, critical assumption of the existing literature's overarching framework that we have argued deserves scrutiny is the assumption that the consequences of the controlled processing triggered by overt racial cues are largely invariant. We tested this assumption of the existing literature against a competing hypothesis, which stated that although controlled processing will be associated with a reduction in racial biases among liberal-minded individuals, it will be associated with an exacerbation of racial biases among conservative-minded individuals. The results regarding these competing hypotheses were mixed. On the one hand, there were ideological asymmetries in the behavioral correlates of conflict detection and controlled processing, and as expected, conscious racial bias inhibition (CRBI) was associated with activation in regions associated with conflict detection and controlled processing among liberals. Further, among conservatives, there was notably less evidence of CRBI being associated with conflict detection or controlled processing, and the associations that did exist suggested conflict detection and controlled processing are associated with exacerbation of racial biases among conservatives.

On the other hand, there was also evidence that some clusters of activation in brain regions related to controlled processing are associated with an increased likelihood of anti-Black racial biases among liberals. This suggests two things: 1) there are some instances in which encouraging controlled processing among liberals can exacerbate the influence of anti-Black racial biases on policy evaluations, and 2) the neuroanatomy associated with conflict detection and controlled processing is complex and should not be oversimplified. We would like to especially highlight the latter point, as it should be noted that throughout most of the MRI analyses reported in this paper, there was some degree of activation associated with CRBI in every ROI. Despite the fact that ROI analyses allowed us to focus our analyses on theoretically relevant parts of the brain, much work has yet to be done in social neuroscience that can shed light on the functions of particular brain regions as well as the functions of distinct clusters within particular brain regions. For example, one part of the PFC may be associated with the executive function commonly attributed to the frontal cortex, but another part of the PFC may be more involved with functions similar to those of the insula. Future research should be (and in many ways is currently being) done to establish a higher resolution map of the brain's functions in relation to racial thought.

The findings presented here have substantial implications for not only understanding the psychological mechanisms of race-targeted policy attitudes, but also for evaluating the utility of interventions aimed at reducing the role of prejudice in politics. A common argument, to some degree based in academic literature, is that to reduce racial prejudice, individuals should be encouraged to exert effortful thought, deliberation, and controlled processing toward issues of race. In other words, get people to "think about race" more and they will come naturally to realize the logical flaws of their prejudices and come to more enlightened, egalitarian opinions. And how can we encourage people to "think about race"? Make race salient; put it "front and center" when people make political evaluations. The findings of this study support this proposition to some degree. Conscious racial cues were found to trigger conflict detection and controlled processing, and in the aggregate, this was associated with diminished racial biases in support for government assistance during trials where race was conscious compared to when race was nonconscious. However, further analysis suggested these effects of conflict detection and controlled processing were concentrated among liberals, and among conservatives, conflict detection and controlled processing were associated either with no differences in racial biases or exacerbation of racial biases. Further, some forms of controlled processing were associated with exacerbation of racial biases among liberals. The implication is that getting people to "think about race" is not a universal tool for diminishing anti-Black racial biases, and in some cases for some individuals, it can make biases worse.

Works Cited

Berkman, E. T., & Falk, E. B. (2013). Beyond brain mapping using neural measures to predict real-world outcomes. *Current Directions in Psychological Science*, *22*(1), 45-50.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, *41*(4), 656.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport, 9*(16), 3735-3739.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). "Separable neural components in the processing of black and white faces." *Psychological Science, 15*(12), 806-813.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: evaluative processing goals shape amygdala activity. *Psychological Science*, *19*(2), 152-160.

Dale, A. M., Greve, D. N., & Burock, M. A. (1999). Optimal stimulus sequences for event-related fMRI. *NeuroImage, 9*, S33-S33.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, *23*, 75-109.

Federico, C. M. (2004). "When do welfare attitudes become racialized? The paradoxical effects of education." *American Journal of Political Science, 48*(2), 374-391.

Federico, C. M. & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: the antecedents and consequences of "principled objections" to affirmative action. *Journal of Personality and Social Psychology*, *82*, 488-502.

Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science, 326*(5951), 399-403.

Haas, I. (2016). Political Neuroscience. In John R. Absher and Cloutier, Jasmin (Eds.), *Neuroimaging Personality, Social Cognition, and Character* (pp. 355-370). Academic/Elsevier.

Hastie, R. (1981). Schematic Principles in Human Memory. In E. Tory Higgins, C. Peter Herman, and Mark P. Zanna (Eds.), *Social Cognition: The Ontario Symposium*, Vol. 1 (pp. 39-88). Hillsdale, NJ: Erlbaum.

Henry, P. J. & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23, 253-283.

Huber, G. A., & Lapinski, J. S. (2006). The "race card" revisited: Assessing racial priming in policy contests. *American Journal of Political Science*, *50*(2), 421-440.

Hurwitz, J., & Peffley, M. (2005). Playing the race card in the post–Willie Horton Era the impact of racialized code words on support for punitive crime policy. *Public Opinion Quarterly*, *69*(1), 99-112.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*, 825-841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*, 782-790.

Jost, J. T., Nam, H. H., Amodio, D. M., & Van Bavel, J. J. (2014). Political neuroscience: The beginning of a beautiful friendship. *Political Psychology*, *35*(S1), 3-42.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732-735.

Kinder, D. R. & Mendelberg, T. (2000). Individualism reconsidered: Principles and prejudice in contemporary American opinion. In J. S. David O. Sears and L. Bobo (Eds.), *Racialized politics: The debate about racism in America* (pp. 44-74). Chicago, IL: University of Chicago Press.

Kinder, D. R., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals*. University of Chicago Press.

Kinder, D. R. & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40, 414-431.

Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. New York: Cambridge University Press.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135.

Mendelberg, T. (2001). *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press.

Molenberghs, P., Halász, V., Mattingley, J. B., Vanman, E. J., & Cunnington, R. (2013). Seeing is believing: Neural mechanisms of action–perception are biased by team membership. *Human Brain Mapping, 34*(9), 2055-2068.

Peirce, J. W. (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Peirce, J. W. (2009) Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2:10.

Plant, E. A., & Devine, P. G. (1998). "Internal and external motivation to respond without prejudice." *Journal of Personality and Social Psychology, 75*(3), 811.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, *6*(12), 1323-1328.

Rilling, J. K., Dagenais, J. E., Goldsmith, D. R., Glenn, A. L., & Pagnoni, G. (2008). Social cognitive neural networks during in-group and out-group interactions. *Neuroimage*, *41*(4), 1447-1461.

Rohr, M., Degner, J., & Wentura, D. (2015). The "emotion misattribution" procedure: Processing beyond good and bad under masked and unmasked presentation conditions. *Cognition and Emotion*, *29*(2), 196-219.

Schreiber, D. (2011). From SCAN to neuropolitics. In P. Hatemi and McDermott, R. (Eds.), *Man is by Nature a Political Animal: Evolution, Biology and Politics* (pp. 273-299). Chicago and London: The University of Chicago Press.

Shkurko, A. V. (2013). Is social categorization based on relational ingroup/outgroup opposition? A meta-analysis. *Social Cognitive and Affective Neuroscience*, *8*(8), 870-877.

Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology, 70*, 476-490.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*, 143-155.

Smith, S. M., Jenkinson, M., Woolrich, M. w., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I, Flitney, D. E., Niazy, R. K., Saunder, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23,* S208-S219.

Sniderman, P. M., Carmines, E. G., Layman, G. C., & Carter, M. (1996). Beyond race: Social justice as a race neutral ideal. *American Journal of Political Science*, 33-55.

Stanley, D., Phelps, E., & Banaji, M. R. (2008). “The neural basis of implicit attitudes.” *Current Directions in Psychological Science*, *17*(2), 164-170.

Tesler, M., & Sears, D. O. (2010). *Obama's race: The 2008 election and the dream of a post-racial America*. Chicago, IL: University of Chicago Press.

Valentino, N. A., Hutchings, V. L., & White, I. K. (2002). Cues that matter: How political ads prime racial attitudes during campaigns. *American Political Science Review*, *96*(01), 75-90.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience*, *23*(11), 3343-3354.

Volz, K. G., Kessler, T., & von Cramon, D. Y. (2009). In-group as part of the self: in-group favoritism is mediated by medial prefrontal cortex activation. *Social Neuroscience*, *4*(3), 244-260.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56-63.