Journal of Experimental Political Science (2024), 1–16
doi:10.1017/XPS.2024.12

RESEARCH ARTICLE

The neural mechanisms of race priming in American politics

Frank J. Gonzalez[1] and Ingrid J. Haas[2]

[1]School of Government and Public Policy, University of Arizona, Tucson, AZ, USA and [2]Department of Political Science, Center for Brain, Biology, and Behavior, University of Nebraska – Lincoln, Lincoln, NE, USA
Corresponding author: Frank J. Gonzalez; Email: fgonzo@arizona.edu

Abstract

What are the psychological mechanisms of racial "dog whistles" in American politics? Literature on race priming in American politics argues when race is primed implicitly, racial biases influence political evaluations, but when race is made salient, individuals can use controlled processing to inhibit automatic biases and abide by egalitarian norms. However, the neural mechanisms underlying these processes have yet to be examined directly. In a 2 × 2 within-groups experiment using functional magnetic resonance imaging, we examine these neural mechanisms. We find brain areas associated with conflict detection, evaluative processing, and controlled processing are more active when race is primed explicitly rather than implicitly, as expected, although we do not find substantial brain activation associated with automatic responses to be more active during implicit than explicit primes. Results are discussed in terms of understanding how racial cues influence political evaluations while considering America's ever-changing racial norms.

Keywords: political neuroscience; racial prejudice; race priming; government assistance; magnetic resonance imaging; neuroimaging

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

Scholars and pundits alike often talk about racial "dog whistles," or subtle racial cues meant to stoke racial biases in political evaluations. Indeed, work in psychology and political science – often referred to as the implicit–explicit (IE) model (Mendelberg 2001) – suggests racial biases are most evident in political

This article has earned badges for transparent research practices: Open data and Open materials. For details see the Data Availability Statement.

evaluations 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 and consider race explicitly. Some recent work suggests race has become chronically accessible in the post-Obama era and the effects of racial priming may not apply in the same way as before (Valentino, Neuner, and Vandenbroek 2018), but other work shows these dynamics are often still at play depending on the context (Reny, Valenzuala, and Collingwood 2020). Yet we have never directly examined the proposed mechanisms of the IE model. We do not know the extent to which, in the domain of politics, overt (compared to covert) racial cues actually induce controlled processing (i.e. effortful thought) and covert cues prompt automatic, affective processing. Although some work has used experimental manipulations to indirectly stimulate conscious race processing or inferred controlled processing from measures such as response latencies (e.g. Valentino, Hutchings, and White 2002), we have yet to more directly examine the particular neural processes that accompany overt versus covert racial cues.

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 primed either explicitly or implicitly. 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.

A neural framework for race priming

According to the IE model, the racial appeals most likely to influence political evaluations are those that only implicitly prime race because when race is primed explicitly, people are cued to reject prejudiced instincts and abide by egalitarian norms encouraged by society. However, when race is primed implicitly, the argument is that racial biases can influence political attitudes automatically (e.g., Hurwitz and Peffley 2005; Mendelberg 2001; Valentino, Hutchings, and White 2002). A variety of methodologies have been used to test this model and its psychological mechanisms, such as using response latency measures to gauge the salience of race during political evaluations (Valentino, Hutchings, and White 2002), but these methods rely heavily on assumptions regarding the mechanisms by which race primes impact political evaluations. Here, we seek to directly examine the hypothesized mechanisms at work in the brain. fMRI makes 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 (Friston 2009). Using fMRI, 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). Neuroscience research on race processing has provided an excellent starting point for understanding the presumed mechanisms of the IE model.

Group-based categorization seems hardwired into the brain's processing of social information. With regard to reactions to outgroups, including racial outgroups, the most commonly activated brain regions include the amygdala and insula (Cunningham et al. 2004), which tracks with social psychology work on how people immediately categorize outgroup members as emotionally significant (e.g. Brewer 2007) and often associate them with uncertainty or danger, which must be avoided (Richeson et al. 2003; see Shkurko 2013 for meta-analysis).[1] This reaction – that is, the activation of these brain regions – occurs automatically and even when race is only primed nonconsciously (Cunningham et al. 2004).

A notable literature in social neuroscience has shed light on what happens when race is made salient. There is a conflict between people's nonconscious adverse reactions to stigmatized outgroups and their conscious goals of being or appearing unprejudiced. The anterior cingulate cortex (ACC) detects the goal conflict, the insula integrates one's automatic emotional response with the decision-making process, the amygdala response is suppressed, and certain parts of the prefrontal cortex (PFC) kick in that are associated with higher-level cognitive functioning (see Amodio 2010; Amodio, Devine, and Harmon-Jones, 2007; 2008; Amodio et al. 2004; Stanley, Phelps, and Banaji 2008).[2] Controlled processes are thus triggered when race makes the jump to conscious awareness, and these processes can allow people to inhibit prejudiced reactions.

Expectations

Although social neuroscience has done an excellent job of identifying crucial processes in how people process race, it has not tested this framework in the domain of political attitudes and decisions. Political psychology research tells us political attitudes and decisions likely exist in the context of a broader constellation of factors such as ideology, partisanship, and other political values. The race priming literature is not solely about how individuals respond to being prompted with Black versus White individuals or faces, but rather, how implicitly versus explicitly priming the broad concept of race in the domain of politics (e.g., in campaign messages, the news, or when thinking about racialized issues and policies) impacts political evaluations. As such, instead of just looking at neural responses to Black versus White face primes, we are more interested when race, broadly, is explicitly versus implicitly part of a political evaluation. We base our hypotheses on the neural roadmap provided by the social neuroscience research that examines how people process race in apolitical contexts, but in the domain of politics, it is unknown how much inhibitory and controlled processes kick in (and automatic emotional responses are diminished) when race is made explicit despite the other factors that come into play in political contexts, such as ideology, partisan motivated reasoning, and the fact that political decisions can always be guised as "nonracial" and ideological even when they are based on race.

We had participants engage in a task in which they decided how strongly they would support or oppose government assistance to various individual applicants whose race (White or Black) was either explicitly or implicitly primed. In line with

[1]It is important to note, here, that the amygdala is most appropriately thought of as detecting emotional significance rather than threat or danger, per se (Cunningham, Van Bavel, and Johnsen 2008).

[2]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.

the IE model, we expect behaviorally for policy evaluations to be less biased against Black applicants when race is primed explicitly than when it is primed implicitly. In terms of the brain, we expect that when race is primed implicitly, individuals will exhibit greater automatic processing aimed at detecting emotional significance (amygdala), and when race is primed overtly, individuals should exhibit greater conflict detection and emotional integration (ACC and insula) as well as controlled processing (orbitofrontal cortex and prefrontal cortex). We also expect these brain activation patterns to track with behavior in terms of racial biases in aid support (e.g., activation in ACC when race is primed explicitly is associated with lower anti-Black bias when race is primed explicitly).

Data and methods

Twenty-seven adults were recruited from the community surrounding a large Midwestern university. Given our interest in anti-Black prejudice, and as is common in race priming research (e.g., Valentino, Hutchings, and White 2002), we prescreened individuals to only recruit White participants who had no issues with claustrophobia or permanent metal in their body and were right-handed (to avoid confounds related to brain lateralization related to handedness). We were conscious in our recruitment and prescreening to ensure the sample was adequately diverse in terms of political ideology and partisan affiliation. Nonetheless, the sample exhibited a slight liberal/Democratic skew (16 participants identified as liberal and 8 as conservative; 13 participants identified as Democrats and 10 as Republicans). Due to a technical error, the behavioral response data for two participants was not properly recorded during the task, limiting our sample to just 25 for analyses involving behavioral response data (our behavioral task results and MRI analyses that involve "conscious pro-Black evaluations [CPBE]" – see below). See the Online Appendix for additional sample information and this project's Harvard Dataverse site for survey items included in the postscan survey.

Experimental design

The experimental task used a 2 (*race of applicant*: White vs. Black) $\times$ 2 (*conscious awareness*: nonconscious vs. conscious) rapid event-related within-subject design consisting of four blocks of trials.[3] *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. Each trial presented participants with a random piece of information about the applicant and asked them to decide whether they support (on a 4-point scale from "Strongly Oppose" to "Strongly Support") monetary aid to that individual using response pads while in the scanner. 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 Online Appendix for complete list of

[3]We use the terms "conscious" and "nonconscious" here rather than explicit and implicit, as these terms more accurately describe the type of primes used here.

statements). Participants were told the applicant descriptions would seem random but that this was deliberate for the purposes of the study. All applicants were men to avoid any effects of applicant gender.

Prior to 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 Black and White individuals. 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 totaled 5 seconds and began with a fixation cross shown for 500 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. Further, in the instructions preceding each Conscious block, participants were told the image was of the applicant. 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. 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). In the instructions for Nonconscious blocks, participants were simply told they would see images of black and white dots prior to the descriptions. 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. Further, the instructions for Conscious blocks stated the faces belonged to the applicants, whereas this was not the case for Nonconscious blocks. The exact timings used for this experiment were pretested on a small sample of undergraduate students as well as during pilot runs in the fMRI scanner.[4] The protocol was built using PsychoPy (Peirce 2009) and is illustrated in Figure 1.

MRI data and analysis

A Siemens Skyra 3.0 Tesla MRI with a 32-channel head coil was used to collect functional MRI data. For MRI data analysis, we use a statistical package called FSL that was specifically developed for this purpose and is widely used (e.g., Haas, Baker, and Gonzalez 2017, 2021). The analytical techniques we use are summarized in

[4]Eight participants reported being able to see faces, and 11 said they were able to make out the race of faces seen between noise masks. We still consider the primes nonconscious for these individuals as the instructions for Conscious blocks stated the faces belonged to the applicants but this was not the case for Nonconscious blocks, and so in Nonconscious blocks, faces were still not consciously associated with the proceeding applicant description or evaluation.

https://doi.org/10.1017/XPS.2024.12 Published online by Cambridge University Press

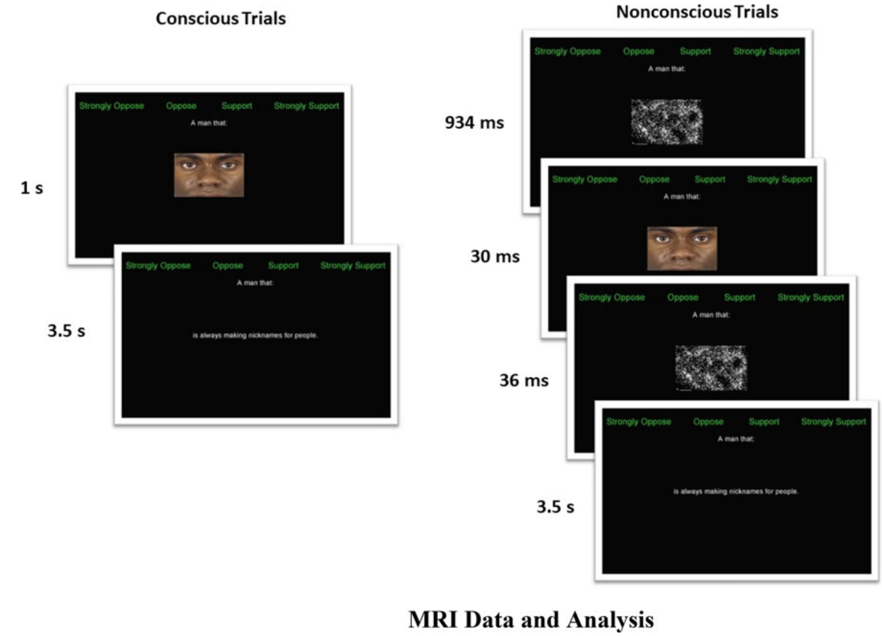

Figure 1. Illustration of trial procedure.

Smith et al. (2004), and so we recommend seeing that paper for additional details on how data were modeled. Generally, we have trials nested within blocks nested within participants, and so our analyses use multilevel modeling in conjunction with Bayesian priors to address several properties specific to fMRI data. General linear models are used to model the blood oxygen level-dependent (BOLD) signal in the brain across trials, and then more sophisticated mixed effects models were used for block- and participant-level analyses. Our main analyses involve modeling BOLD signal at the block-level, as our hypotheses are focused on conscious versus nonconscious racial evaluations rather than evaluations of Black versus White applicants, per se. In other words, as alluded to earlier, we are interested in whether race is explicitly or implicitly primed rather than whether the target is Black or White, and so we do not present analyses of brain activation across race of applicant at the trial level. BOLD signal differences were modeled as contrasts across the *conscious awareness* conditions (Conscious/Nonconscious). In other words, we modeled brain activation that was greater in conscious blocks relative to nonconscious blocks, as well as the reverse (Nonconscious > Conscious).

Importantly, we show the clusters of BOLD activation evident in both contrasts (Nonconscious > Conscious and Conscious > Nonconscious) separately. Unlike analyses political scientists may be accustomed to, in fMRI analyses, activation in a region of interest (ROI) can be associated with *both* high and low levels of something simultaneously due to the fact that multiple clusters can exist within the same region. So, one cluster may be associated with Conscious (relative to Nonconscious) trials and another with Nonconscious (relative to Conscious) trials. Cluster information is based on the clusters of contiguously connected voxels that

https://doi.org/10.1017/XPS.2024.12 Published online by Cambridge University Press

differed the most across conditions in that ROI. As such, cluster size indicates straightforwardly the number of voxels connected to one another in the ROI that showed statistically significant differences in BOLD activation across the two conditions being compared. There is no set rule for determining what a "substantially" sized cluster is, and a cluster being small does not necessarily indicate it is not meaningful, but we only look at activation of clusters that are at least 10 voxels in size and significant at $p < 0.05$. Z-scores and p-values are based on the peak difference in BOLD activation in a voxel across conditions in that cluster.

The subject-level analyses were then combined into group-level ROI analyses using an analytic technique called FLAME (see Smith et al., 2004), which uses multilevel time series analysis in conjunction with Bayesian priors. ROI analyses on left amygdala, right amygdala, bilateral insula, ACC, frontal orbital cortex (OFC), and dorsolateral prefrontal cortex (dlPFC) were masked prior to analysis using anatomical masks from the Harvard–Oxford Cortical and Subcortical Atlases provided with FSL. Our threshold for meaningful activation is a cluster size of at least 10 voxels and significance of $p < 0.05$. In a second group-level analysis, we add between-subjects variables of interest (i.e., CPBE, described in more detail below) to the multilevel model in FSL. Additional information on MRI acquisition, preprocessing, masks, and analysis are available in the Online Appendix.

Results

Behavioral task results

We start with analyses of participants' behavioral responses to the task. Multilevel models were run with trial as the unit of analysis nested within participant, and so with 25 participants, that gives us 100 blocks and 4,000 trials of data. See the Online Appendix for additional details on model specification and additional analyses used to test the ecological validity of the task.

The primary behavioral analyses concern applicant race and conscious awareness predicting level of support for monetary aid to the applicant. A main effects model indicated no significant main effect of conscious awareness on support, 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 ($\beta = -0.055$, $SE = 0.28$, $p < 0.05$). Next, an interaction was calculated between applicant race and conscious awareness. The interaction was marginally significant ($\beta = 0.103$, $SE = 0.055$, $p = 0.063$), and showed that although a pro-Black preference existed in Conscious blocks ($\beta = -0.093$, $SE = 0.043$, $p < 0.05$), no significant racial bias existed in Nonconscious blocks ($\beta = 0.010$, $SE = 0.042$, $p = 0.813$). Thus, although the lack of anti-Black bias in Nonconscious blocks was unexpected, more racially progressive evaluations were observed in Conscious blocks as hypothesized, which we return to in the Discussion.

fMRI results

First, BOLD signal was modeled as a function of conscious awareness (i.e. contrasts between Conscious and Nonconscious blocks) to see which brain regions exhibited significant activation for Conscious (compared to Nonconscious) trials as well as Nonconscious (compared to Conscious) trials. Table 1 lists significant clusters of activation in ROIs using the aforementioned thresholds regarding cluster size and significance. The columns in Table 1 list anatomical regions of interest, the contrast of interest, cluster size (number of contiguous voxels), Z-score for peak voxel of activation, p-value corresponding to the Z-value, and XYZ coordinates in MNI space for the peak voxel.

We have organized results according to the brain regions thought to reflect the distinct types of neural processing in our hypotheses (automatic processing, conflict detection, and controlled processing). As expected, and in line with the assumptions of existing literature, there were significant clusters of activation in brain regions associated with conflict detection, emotional integration, and controlled processing (insula, ACC, OFC, dlPFC) during Conscious (compared to Nonconscious) blocks of trials. Some activation in OFC and dlPFC was also associated with Nonconscious (compared to Conscious) trials, but these clusters were generally smaller in size and strength.

Contrary to expectations from the literature, we did not find evidence for the amygdala being more active in Nonconscious than Conscious trials (suggesting amygdala activation was similar in both). As an example of the controlled processing we found, Figure 2 illustrates the significant cluster of activation in dlPFC for Conscious (compared to Nonconscious) trials.

Beyond the main effects of *conscious awareness*, we also wanted to test our expectation that the observed brain activation differences across condition are associated with behavioral differences in racial biases in aid support. Since we could not examine *differences* in support for White versus Black applicants at the trial level (each trial involved only one applicant), we constructed a between-subjects CPBE variable to test how conflict detection, emotional integration, and controlled processes related to differences in racial biases in support for government assistance between Conscious and Nonconscious trials. The CPBE 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. Then, we subtracted the racial bias score for Nonconscious trials from the racial bias score for Conscious trials. As such, the CPBE score gauged the degree to which participants exhibited more *pro*-Black evaluations during Conscious trials than Nonconscious trials (or, *more* anti-Black bias during Nonconscious trials than Conscious trials).[5] Said differently, this variable reflected the degree to which the race priming literature's hypotheses regarding explicit racial evaluations were true for any given individual ($M = 0.103$, $SD = 0.428$, range $= -0.640$ to 1.059). Despite our small sample size at the subject

[5]Given our behavioral results showed Conscious trials to results in pro-Black evaluations in the aggregate (rather than just no differences by applicant race), we opt for phrasing that alludes to pro-Black evaluations in Conscious trials rather than anti-Black evaluations in Nonconscious trials.

Table 1. Clusters of bold activation in ROIs for conscious vs nonconscious trials

Region of interest		Contrast	Cluster size (# voxels)	Peak activation (*Z*-score)	*p*-value	*X*	*Y*	*Z*
Automatic processing	Left amygdala	Noncon. > Con.	—	—	—	—	—	—
		Con. > Noncon.	—	—	—	—	—	—
	Right amygdala	Noncon. > Con.	—	—	—	—	—	—
		Con. > Noncon.	10	2.273	0.023	26	−10	−10
Conflict detection & emotional integration	Insula	Noncon. > Con.	—	—	—	—	—	—
		Con. > Noncon.	395	2.678	0.007	−38	10	−8
	ACC	Noncon. > Con.	—	—	—	—	—	—
		Con. > Noncon.	360	2.987	0.003	−4	−4	32
Controlled processing	OFC	Noncon. > Con.	40	2.302	0.021	26	14	−26
		Con. > Noncon.	358	2.698	0.007	50	24	−6
	dlPFC	Noncon. > Con.	22	1.972	0.049	34	40	44
		Con. > Noncon.	1438	3.252	0.001	−44	38	18

Note: "Noncon. > Con." rows are clusters that were more active in Nonconscious than Conscious trials, and vice versa for "Con. > Noncon."; cluster size indicates the number of voxels in 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.

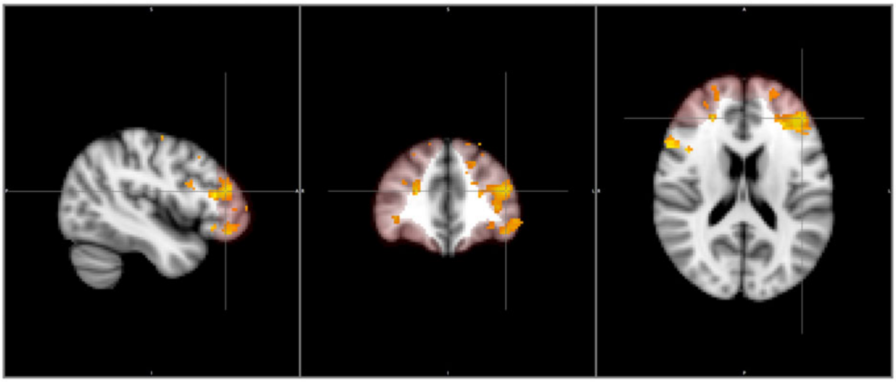

Figure 2. BOLD activation in dlPFC in response to conscious > nonconscious trials.
Note: **Activation in dorsolateral prefrontal cortex for conscious relative to nonconscious blocks of trials. Images were created by overlaying the thresholded *Z*-statistic image on a standard space template (MNI152). Images are centered on the peak voxel for the cluster from the ROI analyses ($X = -44$, $Y = 38$, $Z = 18$). Areas highlighted in red indicate the ROI for the analysis, and yellow/orange clusters indicate regions of significant activation.**

level, we had a decent range of CPBE scores. Eight participants had negative CPBE scores, 1 had a perfectly neutral CPBE score, and 16 had positive CPBE scores.

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. We modeled BOLD signal as a function of the interaction between conscious awareness and CPBE to identify which brain regions were more active in Conscious than Nonconscious trials among individuals who tended to make more pro-Black decisions in Conscious (compared to Nonconscious) trials.[6] Table 2 contains these results. We again show both contrasts to identify clusters of activation associated with Conscious trials among those high in CPBE as well as those associated with Conscious trials among those low in CPBE.

Activation in amygdala was associated with CPBE in Conscious trials. Participants who showed greater pro-Black behavior in conscious trials had greater activation in left amygdala. However, we also found activation in left amygdala associated with less pro-Black behavior. Substantial clusters of activation in insula, ACC, and OFC were associated with reduced racial biases in support for government assistance in Conscious (compared to Nonconscious) trials but not the reverse. Thus, in line with the expectations of the existing literature, individuals who exhibited more pro-Black evaluations in the Conscious (compared to Nonconscious) trials were more likely to show activation in brain regions associated with conflict detection, emotional integration, and controlled processing. Notably, though, this was not the case regarding dlPFC, and so CBPE seems driven more by regions associated with conflict detection and emotional integration (ACC,

[6]If we just looked at brain activation associated with CPBE generally, it would be telling us clusters of activation that occurred in both Conscious and Nonconscious blocks for those high (or low) in CPBE, and so it is necessary to specify the interaction term so that we are only identify which parts of the brain are more active in Conscious blocks for high (or low) CPBE individuals.

https://doi.org/10.1017/XPS.2024.12 Published online by Cambridge University Press

Table 2. Clusters of bold activation in ROIs for conscious awareness*conscious pro-black evaluations contrast

Region of interest		Contrast	Cluster size (# voxels)	Peak activation (*Z*-Score)	*p*-value	*X*	*Y*	*Z*
Automatic processing	Left amygdala	>CPBE	18	2.366	0.018	−34	−2	−20
		<CPBE	13	2.019	0.043	−16	−6	−20
	Right amygdala	>CPBE	—	—	—	—	—	—
		<CPBE	—	—	—	—	—	—
Conflict detection & emotional integration	Insula	>CPBE	209	2.455	0.014	40	18	−10
		<CPBE	—	—	—	—	—	—
	ACC	>CPBE	298	2.531	0.011	−4	44	4
		<CPBE	—	—	—	—	—	—
Controlled processing	OFC	>CPBE	75	2.455	0.014	40	18	−10
		<CPBE	14	2.488	0.013	−38	20	−26
	dlPFC	>CPBE	—	—	—	—	—	—
		<CPBE	—	—	—	—	—	—

Note: "> CPBE" rows are clusters that were more active among individuals who exhibited more pro-Black evaluations in Conscious than Nonconscious trials (i.e., were high in CPBE); "< CPBE" rows are the same but for individuals who exhibited less pro-Black evaluations in Conscious than Nonconscious trials (i.e., were low in CPBE); See Table 1 caption for additional details.

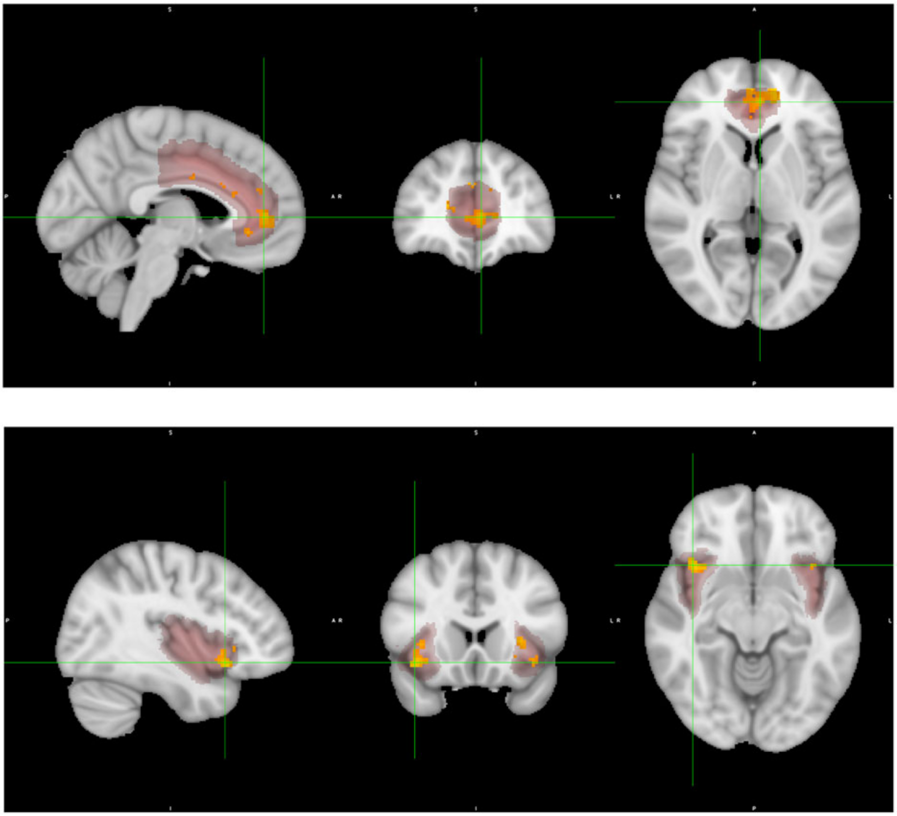

Figure 3. BOLD activation in ACC (Top) and Insula (Bottom) for individuals high in Conscious Pro-Black Evaluations (CPBE) for conscious > nonconscious trials.
Note: **See Figure 2 caption (coordinates for ACC: $X = -4$, $Y = 44$, $Z = 4$; coordinates for Insula: $X = 40$, $Y = 18$, $Z = -10$).**

insula) than by regions associated with executive function or controlled processing (dlPFC). Upon further inspection, it was also the case that the cluster of activation in OFC overlapped with the larger cluster in Insula, suggesting the cluster in OFC was related more to emotional integration than controlled processing of the PFC. Figure 3 shows the significant cluster of activation in ACC and Insula associated with increased pro-Black evaluations in Conscious over Nonconscious trials.

Overall, during evaluations regarding applicants for government assistance, brain regions associated with conflict detection and emotional integration (insula, ACC) and controlled processing (OFC, dlPFC) indeed seemed to be more active when race was conscious compared to when race was nonconscious. However, the converse – that regions associated with automatic, affective processes would be more active when race is nonconscious – was not supported. When it comes to the relationship between conflict detection/emotional integration/controlled processing and racial biases in support for government assistance, brain regions associated with conflict detection and emotional integration were especially active

https://doi.org/10.1017/XPS.2024.12 Published online by Cambridge University Press

during Conscious (compared to Nonconscious) trials among individuals who exhibited more pro-Black evaluations in terms of support for government assistance, suggesting the brain's signal that a conflict exists between one's automatic and controlled reactions and emotional responses to that conflict are key to actual changes in preferences.

Discussion: Inside the black box of race priming in America

A good deal of research in political science has been focused on examining how racial "dog whistles" influence political attitudes. The predominant thinking – the IE model – has been that anti-Black biases are most likely to influence political evaluations when racial cues are covert and that overt cues allow people to recognize the attempt at triggering a prejudiced reaction (conflict detection), integrate their automatic, emotional response (emotional integration), inhibit their racial biases (controlled processing), and express egalitarian views. We sought to test the neural mechanisms of this theoretical model.

Our results did not provide evidence for automatic neural processes being associated with nonconscious (relative to conscious) racial evaluations, but we also did not find anti-Black bias in the aggregate. The former result may be seen as going against the social neuroscience literature on race processing to some degree, but we want to be clear that our interests (and ultimately, our analyses) are fundamentally different than what is examined in most of the social neuroscience work on race. Rather than looking at Black versus White face primes, we examined when race, broadly, is primed in the context of political evaluations. As such, beyond possibly being due to our sample, it may be the case that automatic processes were present in both Conscious and Nonconscious blocks to roughly equal degrees. When race meets politics, automatic emotional processing may be more relevant even in the explicit setting because of the emotional and motivated nature of political decision-making. Indeed, we can see there were clusters of activation in the amygdala associated with both high CBPE and low CBPE. Further, finding no aggregate anti-Black bias could make sense given the "colorblind" nature of opposition to race-based policies (e.g., Kinder and Mendelberg 2000; Sniderman et al. 1996). Whereas other studies tend to examine racial bias in attitudes toward politicians, campaign ads, and covertly racial policies (e.g., crime, healthcare) when race is primed implicitly versus explicitly (Hurwitz and Peffley 2005; Mendelberg 2001; Valentino, Neuner, and Vandenbroek 2018), our study asked participants to repeatedly evaluate aid to specific individuals, and so "not seeing color" may be a more likely manifestation of anti-Black attitudes in this sort of task than participants overtly opposing aid to Black applicants.

Despite a lack of support for the role of automatic processing of conscious versus nonconscious primes, our findings regarding conscious racial evaluations and controlled processing/emotional integration supported the framework of the IE model. Conflict detection and emotional integration (insula and ACC) and controlled processing (OFC and dlPFC) were significantly more evident in Conscious than Nonconscious trials, as expected. Further, conflict detection and emotional integration were associated with exhibiting more pro-Black evaluations in terms of support for government assistance in Conscious (compared to Nonconscious) trials (i.e., being high in CPBE).

Several notable limitations exist within this study. Additional research might use other tasks to assess responses to implicit versus explicit race primes. Although the task used here allowed us to more clearly identify how conscious race was in the decision-making process, it lacked ecological validity and arguably did not represent the actual decision-making process people generally go through when considering policies. Relatedly, the task involved decisions about monetary aid, and the races primed were just White or Black, and so we cannot speak to how well these findings generalize to other issue domains and racial groups. There are many more issue domains (e.g., crime/the police, immigration, affirmative action) and racial or ethnic groups (e.g., Latinos, Muslims) that existing work suggests are also influenced by race priming (e.g., Hurwitz and Peffley, 2005; Reny, Valenzuela, and Collingwood, 2020). Finally, as is often the case with MRI studies, our sample size at the subject level was limited, which is particularly relevant to the between-subject analyses of CPBE. Finally, even with just 25 subjects in the CPBE analyses, there was decent variation in CPBE in our sample (see Results). Nonetheless, consideration of the low sample size here is warranted, and it would be useful to have greater variation in factors like ideology (given the lack of extreme conservatives in our sample).

Taken together, our findings provide the first direct evidence of the psychological mechanisms theorized to be at work by scholars studying race priming. That said, future research should still be done to identify the nuances of how conflict detection and controlled processing lead to more pro-Black evaluations. Neuroscience continues to shed light on the distinct implications of specific areas or processes within the ROIs we analyzed. For example, we know cortical regions such as the frontal lobe and ACC can be further divided into regions more related to "cognitive functions" and regions more related to "emotional functions" (see Bush, Luu, and Posner 2000). A cursory look at our own results for the ACC suggests activation associated with Conscious trials was indeed clustered mainly in what Bush, Luu, and Posner identify as the more "cognitive" region of the ACC whereas activation associated with greater CBPE was clustered in the more "emotional" region. Perhaps explicit race primes trigger an initial cognitive recognition of conflict between automatic responses and conscious goals reflected in the ACC, but the difference between those who subsequently make more pro-Black evaluations and those who do not lays in the emotional responses to perceived goal conflict associated with the ACC and insula. Future work might investigate important nuances like these more directly.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/XPS.2024.12

Data availability. Support for this research was provided by the National Science Foundation (Award no. 1560432). The data, code, and any additional materials required to replicate all analyses in this article are available at the *Journal of Experimental Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/1DFIJA (Gonzalez and Haas, 2024).

Acknowledgments. We would like to thank the National Science Foundation Division of Social and Economic Sciences (NSF SES) 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. This work also benefited tremendously from invaluable feedback from colleagues at the *Political Psychology Four Corners Network* conference at Arizona State University in February, 2023 and the *Science and Politics: Changing Attitudes and Emerging Methodologies* workshop at Western University in November, 2021.

Author contributions. Frank J. Gonzalez is responsible for the initial conception and design of the study. Both Frank J. Gonzalez and Ingrid J. Haas contributed to data analysis and interpretation, manuscript drafting and revision, and approval of the final version for submission.

Competing interests. The authors have no conflicts of interests related to the current study to disclose.

Ethics statement.

(A) This study was approved by the University of Nebraska – Lincoln Internal Review Board (IRB) (Protocol # 20150215036EP).
(B) The research presented here adheres to APSA's Principles and Guidance for Human Subjects Research (https://connect.apsanet.org/hsr/principles-and-guidance/)
(C) More information on the ethical practices related to this study can be found in the Online Appendix.

References

Amodio, D. M. 2010. "Coordinated Roles of Motivation and Perception in the Regulation of Intergroup Responses: Frontal Cortical Asymmetry Effects on the P2 Event-Related Potential and Behavior." *Journal of Cognitive Neuroscience 22*: 2609–17.

Amodio, D. M., P. G. Devine, and E. Harmon-Jones. 2007. "A Dynamic Model of Guilt: Implications for Motivation and Self-Regulation in the Context of Prejudice." *Psychological Science 18*: 524–30.

Amodio, D. M., P. G. Devine, and E. Harmon-Jones. 2008. "Individual Differences in the Regulation of Intergroup Bias: The Role of Conflict Monitoring and Neural Signals for Control." *Journal of Personality and Social Psychology 94*: 60–74.

Amodio, D. M., E. Harmon-Jones, P. G. Devine, J. J. Curtin, S. L. Hartley, and A. E. Covert. 2004. "Neural Signals for the Detection of Unintentional Race Bias." *Psychological Science 15*: 88–93.

Brewer, M. B. 2007. "The Social Psychology of Intergroup Relations: Social Categorization, Ingroup Bias, and Outgroup Prejudice." In *Social Psychology: Handbook of Basic Principles*, ed. A. W. Kruglanski and E. T. Higgins, 695–715. The Guilford Press.

Bush, G., P. Luu, and M. I. Posner. 2000. "Cognitive and Emotional Influences in Anterior Cingulate Cortex." *Trends in Cognitive Sciences 4*(6): 215–22.

Cunningham, W. A., M. K. Johnson, C. L. Raye, J. C. Gatenby, J. C. Gore, and M. R. Banaji. 2004. "Separable Neural Components in the Processing of Black and White Faces." *Psychological Science 15*(12): 806–13.

Cunningham, W. A., J. J. Van Bavel, and I. R. Johnsen. 2008. Affective Flexibility: Evaluative Processing Goals Shape Amygdala Activity." *Psychological Science 19*(2): 152–60.

Friston, K. J. 2009. "Modalities, Modes, and Models in Functional Neuroimaging." *Science 326*(5951): 399–403.

Gonzalez, Frank, and Ingrid Haas. 2024. "*Replication Data for: The Neural Mechanisms of Race Priming in American Politics*." https://doi.org/10.7910/DVN/1DFIJA, Harvard Dataverse, V1, UNF:6:/DWkcNa7csVTAIdv26k0Iw= = [fileUNF].

Haas, I. 2016. "Political Neuroscience." In *Neuroimaging Personality, Social Cognition, and Character*, ed. John R. Absher and Cloutier, Jasmin. Academic/Elsevier, 355–70.

Haas, I. J., M. N. Baker, and F. J. Gonzalez. 2017. "Who Can Deviate from the Party Line? Political Ideology Moderates Evaluation of Incongruent Policy Positions in Insula and Anterior Cingulate Cortex." *Social Justice Research 30*: 355–80.

https://doi.org/10.1017/XPS.2024.12 Published online by Cambridge University Press

Haas, I. J., M. N. Baker, and F. J. Gonzalez. 2021. "Political Uncertainty Moderates Neural Evaluation of Incongruent Policy Positions." *Philosophical Transactions of the Royal Society B 376*(1822): 20200138.

Hurwitz, J., and M. Peffley. 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.

Jost, J. T., H. H. Nam, D. M. Amodio, and J. J. Van Bavel. 2014. "Political Neuroscience: The Beginning of a Beautiful Friendship." *Political Psychology 35*(S1): 3–42.

Kinder, D. R. and T. Mendelberg. 2000. "Individualism Reconsidered: Principles and Prejudice in Contemporary American Opinion." In *Racialized Politics: The Debate about Racism in America*, ed. J. S. David, O. Sears and L. Bobo, 44–74. Chicago, IL: University of Chicago Press.

Ma, D. S., J. Correll and B. Wittenbrink. 2015. "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data." *Behavior Research Methods 47*(4): 1122–35.

Mendelberg, T. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton University Press.

Peirce, J. W. 2009. "Generating Stimuli for Neuroscience Using PsychoPy." *Frontiers in Neuroinformatics 2*: 10.

Reny, T. T., A. A. Valenzuela, and L. Collingwood. 2020. "No, You're Playing the Race Card": Testing the Effects of ANTI-BLACK, ANTI-LATINO, and Anti-Immigrant Appeals in the Post-Obama Era." *Political Psychology 41*(2): 283–302.

Richeson, J. A., A. A. Baird, H. L. Gordon, T. F. Heatherton, C. L. Wyland, S. Trawalter, and J. N. Shelton. 2003. "An fMRI Investigation of the Impact of Interracial Contact on Executive Function." *Nature Neuroscience 6*(12): 1323–28.

Rohr, M., J. Degner, and D. Wentura. 2015. "The "Emotion Misattribution" Procedure: Processing Beyond Good and Bad Under Masked and Unmasked Presentation Conditions." *Cognition and Emotion 29*(2): 196–219.

Shkurko, A. V. 2013. "Is Social Categorization Based on Relational Ingroup/Outgroup Opposition? A Meta-Analysis." *Social Cognitive and Affective Neuroscience 8*(8): 870–77.

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., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. 2004. "Advances in functional and structural MR image analysis and implementation as FSL." *NeuroImage 23*(S1): 208–19.

Sniderman, P. M., E. G. Carmines, G. C. Layman, and M. Carter 1996. "Beyond Race: Social Justice as a Race Neutral Ideal." *American Journal of Political Science 40*(1): 33–55.

Stanley, D., E. Phelps, and M. R. Banaji. 2008. "The Neural Basis of Implicit Attitudes." *Current Directions in Psychological Science 17*(2): 164–70.

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

Valentino, N. A., F. G. Neuner, and L. M. Vandenbroek. 2018. "The Changing Norms of Racial Political Rhetoric and the End of Racial Priming." *The Journal of Politics 80*(3): 757–71.

Cite this article: Gonzalez FJ and Haas IJ (2024). The neural mechanisms of race priming in American politics. *Journal of Experimental Political Science*. https://doi.org/10.1017/XPS.2024.12