doi:10.1093/scan/nss010 SCAN (2013) 8, 387–393

An fMRI investigation of racial paralysis

Michael I. Norton,[1] Malia F. Mason,[2] Joseph A. Vandello,[3] Andrew Biga,[3] and Rebecca Dyer[4]

[1]Harvard Business School, Harvard University, Boston, MA 02139, [2]Columbia Business School, Columbia University, New York, NY 10027, [3]Department of Psychology, University of South Florida, Tampa, FL 33620, and [4]Department of Psychology, Yale University, New Haven, CT 06520

We explore the existence and underlying neural mechanism of a new norm endorsed by both black and white Americans for managing interracial interactions: "racial paralysis', the tendency to opt out of decisions involving members of different races. We show that people are more willing to make choices—such as who is more intelligent, or who is more polite—between two white individuals (same-race decisions) than between a white and a black individual (cross-race decisions), a tendency which was evident more when judgments involved traits related to black stereotypes. We use functional magnetic resonance imaging to examine the mechanisms underlying racial paralysis, to examine the mechanisms underlying racial paralysis, revealing greater recruitment of brain regions implicated in socially appropriate behavior (ventromedial prefrontal cortex), conflict detection (anterior cingulate cortex), deliberative processing (dorsolateral prefrontal cortex), and inhibition (ventrolateral prefrontal cortex). We also discuss the impact of racial paralysis on the quality of interracial relations.

Keywords: race; prejudice; inhibition; choice

INTRODUCTION

Imagine stepping onto a crowded subway car, shopping bags in each hand, and finding two seats left, each next to a similarly dressed man: one white, the other black. Where would you sit? If you are white, choosing to sit next to the white passenger raises the concern that you will be seen as biased, while choosing to sit next to the black passenger raises the concern that you will be seen as—perhaps disingenuously—bowing to political correctness. Nor does being black solve the dilemma; even for a black passenger, either decision appears to constitute a choice made on the basis of race. What happens in these common situations, when individuals must decide whom to sit next to on a bus or stand next to in an elevator, or in even more consequential decisions, such as whom to hire or admit to college?

We suggest that the concern about appearing biased elicited by such situations creates conflict about the appropriate response. As a result, one popular—if sometimes suboptimal—solution is to opt out of the decision altogether in an effort to display racial neutrality. Despite the weight of their shopping bags, individuals may choose to forgo either seat and remain standing, rather than risk the appearance of bias. We suggest that similar solutions to such problems are representative of an emerging trend in interracial relations, which we term racial paralysis: the tendency for people to opt out of situations that require choices seemingly made on the basis of race.

Racial neutrality and racial paralysis

Decades of research in social psychology have explored the tension and negative affect that interracial interactions can engender, with heightened concerns about doing something 'wrong' or behaving inappropriately in such interactions (e.g. Stephan and Stephan, 1985; Vorauer *et al.*, 1998; Richeson and Shelton, 2003; Shelton, 2003; Vorauer and Turpie 2004; Shelton *et al.*, 2005). Most relevant to the present investigation, people seek to appear race-neutral when making decisions between members of different races. In research on aversive racism, for example, while whites continue to exhibit bias against blacks, they do so only when they are able to justify that behavior to themselves and others (Snyder *et al.*, 1979; Crandall and Eshleman, 2003; Dovidio and Gaertner, 2004). In fact, research demonstrates that once 'stuck' in situations in which their discrimination would be obvious—such as when the only bystander in view of a black person in need of help—whites can behave more positively towards blacks (Gaertner and Dovidio, 1986; Pearson *et al.*, 2009), though they are still likely to claim that race was not a factor in their decision (Hodson *et al.*, 2002; Norton et al., 2004). Whether allowing a poorly dressed black patron to enter a restaurant for fear of appearing biased or refusing to help a black person when one can justify it (Dutton, 1971; Dovidio and Gaertner, 1981), interracial situations evoke feelings of uncertainty stemming from a desire to appear unbiased. We suggest that this desire in some cases can lead people to wish to appear as though they have no preference at all.

This desire for racial neutrality has become increasingly prevalent in American culture, as reflected by the emergence of norms of colorblindness (Wolsko *et al.*, 2000; Richeson and Nussbaum, 2004; Pager and Quillian, 2005; Norton *et al.*, 2006; Apfelbaum *et al.*, 2008; Plaut *et al.*, 2009; Vorauer *et al.*, 2009). While norms of colorblindness likely arose from well-meaning intentions—'the best way to be egalitarian is to not even notice race'—the norms provide very little guidance in everyday situations, such as the subway situation with which we opened. If showing any preference in any situation can be construed as evidence of bias, how should a person in a diverse setting behave? We suggest that norms of racial neutrality can in some situations induce 'racial paralysis', where people's concern with appearing unbiased can inhibit both what they say and what they do—all in the direction of saying and doing nothing, but rather opting out of such situations altogether.

We use a paradigm that captures the most basic form of this dilemma: forgoing a choice between two individuals of different races solely on the basis of photographs of their faces. Such an unwillingness to judge faces would stand in stark contrast to people's skill at face perception (Chernoff, 1973; Zebrowitz, 1997; Smith *et al.*, 2005) and willingness to make judgments on that basis. For instance, people are quick to form judgments on the basis of facial attractiveness (Zebrowitz and McDonald, 1991; Willis and Todorov, 2006) and are also willing to choose *between* individuals on the basis of attractiveness

Received 25 July 2011; Accepted 17 January 2012
Advance Access publication 20 January 2012
Funding for these studies was provided by Columbia Business School and Harvard Business School.
The authors thank Dan Ariely, Benoît Monin and Sam Sommers for their helpful suggestions.
Correspondence should be addressed to Michael I. Norton, Harvard Business School, Soldiers Field Road, Boston, MA 02163, USA. Email: mnorton@hbs.edu

(Johansson *et al.*, 2005); indeed, people generally are comfortable making choices between people based on their faces on a variety of dimensions (Hassin and Trope, 2000).

In the example with which we opened, however, all these fine-tuned processes appear to come to a crashing halt. In particular, we suggest that choosing between two individuals from different racial groups—in theory employs many of the same processes as choosing between members of the same groups—is in practice something that people are loathe to do. It is not that judging people based on their race is inherently more difficult, since categorising people by their race is a relatively effortless task (Montepare and Opeyo, 2002; Ito and Urland, 2003), and people do draw inferences about members of other racial groups based on their photographs (Blair et al., 2002). Instead, we suggest that while choosing between two faces of the same race constitutes mere perceptual discrimination between those individuals, choosing between members of different races has greater significance—due to the concern that any decision may serve as an evidence of bias—and therefore induces greater decision conflict, leading individuals to opt out.

Goals of the experiments

We first wanted to establish that people are less willing to choose between members of different races than members of the same race. We predicted that emerging norms of racial neutrality make picking *either* a white person or a black person inappropriate, leading people not to favor members of one race over another, but instead to opt out of decisions altogether. Our second goal was to identify moderating factors and potential boundary conditions of racial paralysis. In particular, we explore whether all cross-race choices increase opting out, or if this response is specific to cross-race choices involving traits that are relevant to black stereotypes (e.g. 'intelligence'). We suggest that opting out is a strategic response that people adopt to avoid seeming racially biased. As with previous research suggesting people's sensitivity to factors that increase concerns of appearing biased (e.g. Blanchard et al., 1991; Apfelbaum et al. 2008), we predicted that opting out would be employed more frequently when cross-race choices involved a stereotype-relevant trait—when such judgments were more loaded with racial connotations.

Our third goal was to elucidate the mechanisms underlying racial paralysis by measuring brain activity while participants engaged in a series of same-race and cross-race judgments. In particular, we wished to demonstrate that cross-race decisions were associated with the recruitment of brain regions that detect and signal conflict as well as brain regions that mediate deliberative processing and implement cognitive control, a finding which would support our contention that cross-race decisions evoke both feelings of uncertainty and compel people to strategise on how to respond in a socially appropriate manner.

As a result, we expected cross-race judgments to be associated with increased recruitment of four particular brain regions. First, we predicted greater recruitment of the anterior cingulate cortex (ACC), which monitors for conflict and signals the need for controlled processing and further deliberation (Petersen *et al.*, 1988; Carter *et al.*, 1998; Botvinick *et al.*, 2001; Lieberman, 2003, 2007; Kerns *et al.*, 2004). Second, we expected greater activity in the dorsolateral prefrontal cortex (DLPFC), which supports efforts to collect, deliberate on and integrate information before choice (Waltz *et al.*, 1999; Christoff and Gabrieli, 2000; Goel and Dolan, 2000). Third, we anticipated that cross-race choices would be associated with increased activity in the ventrolateral prefrontal cortex (VLPFC), a region implicated in inhibiting preferred but contextually inappropriate responses (Kowalska *et al.*, 1991; Casey *et al.*, 1997). We have suggested that the tendency to avoid cross-race choices should be particularly acute when the decision is about traits related to black stereotypes. In light of its involvement in encoding and signaling the emotional value of decisions and behaviors, especially those that threaten normative and moral prescriptions (Damasio *et al.*, 1991; Greene *et al.*, 2001; Beer, Heerey *et al.*, 2003; Camille *et al.*, 2004; Koenigs *et al.*, 2007; Krajbich *et al.*, 2009), we expected to observe heightened ventromedial prefrontal cortex (VMPFC) activity during cross-race comparisons about traits related to black stereotypes.

After first establishing the hypothesised behavioral effect (Experiment 1)—increased opting out of cross-race choices involving traits that are relevant to black stereotypes—we measured cortical activity while participants made cross-race and same-race choices in a magnetic resonance imaging (MRI) scanner (Experiment 2).

EXPERIMENT 1

Method

Participants ($N = 46$; 36 Asian, 8 White, 1 Native American, 1 Hispanic; mean age = 22.5; 58.7% female) participated in exchange for monetary compensation. The experiment had a 2 (choice set: same-race, cross-race) × 2 (stereotype relevance: relevant, irrelevant) repeated-measures design.

Upon arrival in the laboratory, each participant was greeted by a female experimenter and directed to sit in front of a Dell PC computer. Participants were informed that they would see two faces on the screen at a time with a single characteristic listed at the top of the screen, and that their task was to indicate, via a key press, which person was more likely to exemplify the characteristic that was listed, or whether they had no gut feeling. For each trial, a fixation-cross appeared at the center of the screen for 3000 ms, then was replaced with a screen displaying two faces side by side and the phrase 'I have no gut feeling' between the faces. The target trait appeared at the top of the screen. This screen remained visible until a response was recorded and participants completed a total of 90 trials.

Stimuli comprised a total of 60 different male faces presented on a black background. Fifteen of the images presented black males, and 45 images presented white males, such that there were 15 cross-race (one white and one black) and 15 same-race (two white) choices. Each image was approximately 150 × 200 pixels. A total of 30 different characteristics—drawn in part from previous investigations of stereotypes against blacks (e.g. Devine, 1989; Devine and Elliot, 1995; Fiske *et al.*, 2002)—were used in the study, half of which were relevant to black stereotypes (e.g. intelligent, articulate) and half of which were irrelevant (e.g. restless, strict) based on pre-testing (Appendix A).

RESULTS AND DISCUSSION

A repeated measures analysis of variance (ANOVA) of 2 (choice set: same race, cross-race) × 2 (stereotype relevance: relevant, irrelevant) revealed that participants were less likely to make cross-race choices ($M = 0.80$, s.e. = 0.03) than same-race choices ($M = 0.83$, s.e. = 0.02), $F(1, 45) = 3.84$, $P = 0.056$, and were significantly less likely to make choices involving stereotype relevant ($M = 0.79$, s.e. = 0.03) than irrelevant traits ($M = 0.84$, s.e. = 0.02), $F(1, 45) = 11.41$, $P = 0.002$. These effects were qualified by a marginally significant interaction, $F(1, 45) = 3.46$, $P = 0.07$, which as predicted was driven by the fact that opt out rates were significantly higher when participants made cross-race decisions about stereotype-relevant traits relative to when they made choices in the other three contexts, $F(1, 180) = 4.54$, $P < 0.04$. Consistent with our hypothesis, participants were significantly less likely to make choices involving relevant ($M = 0.76$, s.e. = 0.03) than irrelevant traits ($M = 0.84$, s.e. = 0.02) when making

cross-race choices, $t(45) = 3.57$, $P = 0.001$, while choice rates for same-race choices did not depend on the relevance of the trait, $t < 1$, *ns*.

Among participants who did not opt out of cross-race choices, there was a slight preference for picking the white candidate for both relevant ($M = 0.45$) and irrelevant ($M = 0.46$) judgments, compared to choice rates for blacks ($Ms = 0.31$ and 0.38, respectively).

EXPERIMENT 2

Having established behaviorally our prediction that participants are most likely to opt out of cross-race judgments involving characteristics relevant to black stereotypes, we next conducted an imaging study-using the same 2 (choice set: same-race, cross-race) × 2 (stereotype relevance: relevant, irrelevant) repeated measures design, to explore brain regions associated with racial paralysis.

METHODS

Participants ($N = 18$; 12 females; mean age = 22.7; 9 Caucasians, 2 Hispanics, 7 Asians) completed the experiment for monetary compensation. All participants were strongly right-handed as measured by the Edinburgh handedness inventory (Raczkowski *et al.*, 1974), reported no significant abnormal neurological history, and had normal or corrected-to-normal visual acuity.

Stimuli comprised the same 60 faces and 20 of the traits—10 relevant, 10 irrelevant—used in Experiment 1 (Appendix A). Participants were given the same instructions as in Experiment 1—that they would see two faces on the screen and a trait and that their task was to indicate via response keys which person was more likely to exemplify the characteristic, or to indicate that they had no gut feeling. Each trial had the same format as Experiment 1; participants in Experiment 2 completed a total of 120 trials.

Participants were scanned in two event-related functional (EPI) runs. A total of 147 volumes were collected in each EPI run. Across the two runs, participants completed 30 of each trial type for a total of 120 trials. Each trial lasted for a duration of 1.5 TRs (the TR was 2 s). The remaining 57 EPI volumes were jittered catch trials (i.e. fixation symbols, '+') used to optimise estimation of the event-related BOLD response. The stimuli were presented using Presentation (Version 12.1) and back projected with a liquid crystal display (LCD) projector onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Pillow and foam cushions were placed within the head coil to minimize head movements. All images were collected using a GE scanner with standard head coil. T1-weighted anatomical images were collected using a 3D sequence (SPGR; 180 axial slices, TR = 19 ms, TE = 5 ms, flip angle = 20°, FOV = 25.6 cm, slice thickness = 1 mm, matrix = 256 × 256). Functional images were collected with a gradient echo EPI sequence (each volume comprised 27 slices; 4 mm thick, 0 mm skip; TR = 2000 ms, TE = 35 ms, FOV = 19.2 cm, 64 × 64 matrix; 84° flip angle).

fMRI analysis

Functional MRI data were analysed using Statistical Parametric Mapping software (SPM8, Welcome Department of Cognitive Neurology, London, UK; Friston *et al.*, 1995). For each functional run, data were preprocessed to remove sources of noise and artifact. Preprocessing included slice timing and motion correction, coregistration to each participant's anatomical data, normalisation to the ICBM 152 brain template (Montreal Neurological Institute), and spatial smoothing with an 8 mm (full-width-at-half-maximum) Gaussian kernel. Analyses took place at two levels: formation of statistical images and regional analysis of hemodynamic responses. For each participant, a general linear model with 30 regressors was specified. For each run, the model included regressors specifying the four conditions of interest (modeled with functions for the hemodynamic response), six motion-related regressors, a regressor for each of the first four brain volumes collected, and a regressor constant term that SPM automatically generates and includes in the model. The general linear model was used to compute parameter estimates (β) and *t*-contrast images for each comparison at each voxel. These individual contrast images were then submitted to a second-level, random-effects analysis to obtain mean *t*-images.

RESULTS AND DISCUSSION

We followed the same analytic strategy as Experiment 1. We first examined overall differences between cross-race and same-race choices, then turned to exploring interaction effects, and finally focused on the condition in which we expected racial paralysis to be most extreme, as confirmed by the behavioral data in Experiment 1: cross-race choices involving stereotype-relevant traits. Again as in Experiment 1, we conducted contrast analyses comparing these judgments to the other three types of judgments (same-race judgments involving relevant and irrelevant traits, and cross-race judgments involving irrelevant traits).

To determine which regions were more active when participants made cross-race relative to same-race choices, regardless of the stereotype-relevance of the trait, we computed the direct contrast 'cross-race choices > same-race choices', $P < 0.001$; $k = 10$. Consistent with the view that cross-race choices evoke conflict and feelings of uncertainty, the ACC (−6 33 24; BA32) was significantly more active during cross-race relative to same-race choices. Cross-race choices were also associated with greater recruitment of bilateral DLPFC (−15 42 31; 27 51 23; BA9), a brain area that supports explicit attempts by decision-makers to reflect, integrate and deliberate on information before choosing, and greater recruitment of bilateral VLPFC (−33 26 −11; 36 20 −18; BA47), a brain area that plays a central role in inhibiting instinctively preferred but contextually inappropriate responses. No brain regions exhibited significantly greater activity while participants made same- relative to cross-race choices at this threshold (Table 1; Figure 1).

We next explored brain areas that exhibited significant interaction effects. Consistent with our predictions, results revealed that the effect

Table 1 Peak coordinates of brain regions that where activity during cross-race choices was significantly greater than activity during same-race choices, $P < 0.001$, $k = 10$; k = contiguous voxels

k	Anatomical location	BA	Coordinates			*t*-value
			x	*y*	*z*	
Cross-race choices > Same-race choices						
30	R. DLPFC	9	27	51	23	5.23
50	L. DLPFC	9	−15	42	31	5.96
113	L. VLPFC	47	−33	26	−11	9.01
40	R. VLPFC	47	36	20	−18	4.75
15	L. ACC	32	−6	33	24	4.64
11	R. superior frontal	6	15	17	54	4.24
40	L. posterior cingulate	30	−18	−61	7	4.43
71	R. cuneus	18	15	−84	15	5.50
119	R. lingual	18	12	−58	7	4.47
14	R. middle temporal	21	60	−29	−5	4.90
11	L. middle temporal	21	−50	−7	−16	4.56
27	R. superior temporal	39	50	−54	25	4.52
11	L. superior temporal	22	−48	11	−4	4.07

The opposite contrast revealed no significant differences at this threshold. (L.) = Left; (R.) = Right; (BA) = Brodmann Area.

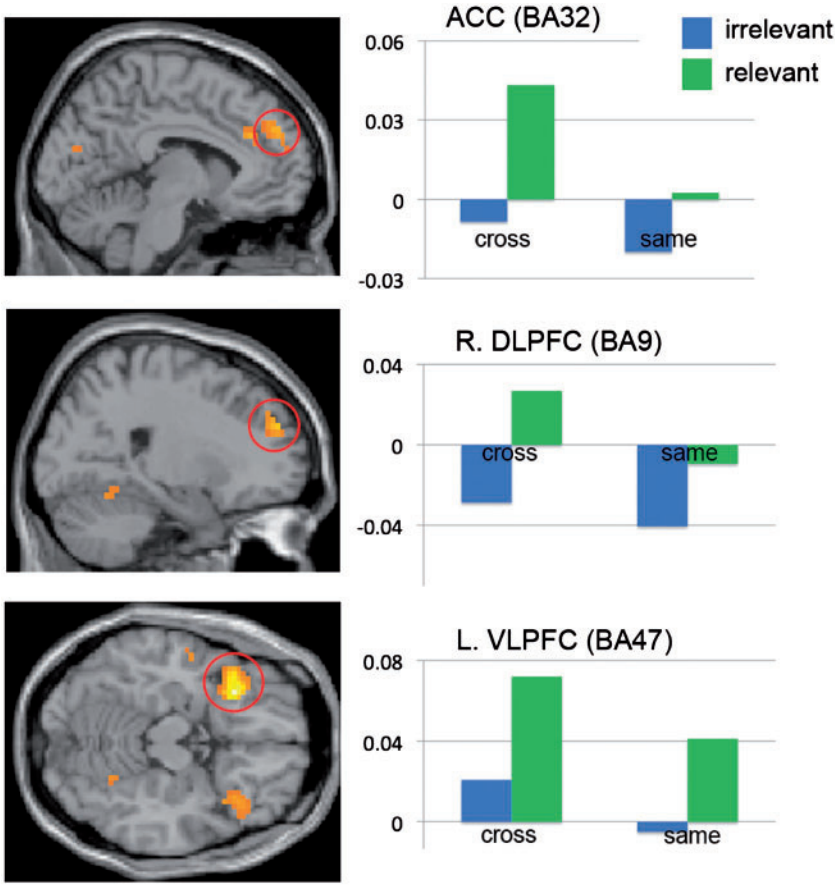

Fig. 1 Percent signal change by condition in regions that exhibited a significant main effect of choice set, $P < 0.001$, $k = 10$; k = contiguous voxels. (Top) is a cluster in the ACC (−6 33 24; BA32); (Middle) is a cluster in the right DLPFC (27 51 23; BA9); (Bottom) is a cluster in the left VLPFC (36 20 −18; BA10). Values were computed by dropping a 10 mm sphere at the cluster's peak, extracting the % signal change with the tools provided by the MarsBar interface (Brett et al., 2002), and then averaging across all participants.

Table 2 (Top) Peak coordinates of brain regions where the effect of relevance on activity was greater for cross- versus same-race comparisons, $P < 0.001$, $k = 10$; k = contiguous voxels (Bottom) Peak coordinates of brain regions where the effect of relevance on activity was greater for same- versus cross-race comparisons.

k	Anatomical location	BA	coordinates			*t*-value
			x	*y*	*z*	
Effect of relevance greater for cross versus same-race						
19	B. VMPFC	12	−6	37	−4	4.63
		12	2	45	−5	3.94
15	R. DLPFC	8/9	24	22	34	4.67
13	L. angular gyrus	22	−36	−72	31	4.26
15	R. precuneus	31	12	−34	31	4.40
Effect of relevance greater for same versus cross-race						
55	R. middle occiputal gyrus	18	24	−93	8	3.87
28	R. cerebellum		37	−69	−12	3.51

$P < 0.001$, k = 10 (B.) = Bilateral; (L.) = Left; (R.) = Right; (BA) = Brodmann Area.

of trait relevance was significantly greater for cross versus same-race judgments in an aspect of the VMPFC (−6, 37, −4; BA 12), $P < 0.001$; $k = 10$, a brain area that plays a central role in signaling the emotional value of decisions and behaviors. Stereotype relevance also moderated the effect of choice set in a region of the right DLPC (24 22 34; BA 8/9), the left angular gyrus (−36 −72 31; BA 39) and an aspect of the precuneus (12 −34 31; BA 31). The effect of trait relevance was significantly greater for same versus cross-race judgments in two regions—a cluster in the right middle occipital gyrus (24 −93 8; BA 18) and a cluster in the cerebellum, $P < 0.001$; $k = 10$ (Table 2).

Finally, we examined regions which were significantly more active during cross-race choices about relevant traits—the judgments we predicted would be most likely to elicit racial paralysis, and confirmed by data from Experiment 1—relative to the other three judgments. Consistent with our predictions, cross-race choices about relevant traits were associated with significantly greater VMPFC (0 50 −6; BA10) recruitment than the other three conditions, $P < 0.001$; $k = 10$, (Figure 2). Furthermore, stereotype relevance moderated the effect of choice set in the ACC (−12 27 21; BA32), bilateral DLPFC (30 19 32; −18 45 27; BA9) and bilateral VLPFC (−27 20 −14; 33 20 -14; BA47). No brain regions exhibited significantly less activity during cross-race choices about relevant traits compared to the other three choice contexts at this threshold (Table 3). These results demonstrate a role for a key moderator of the tendency to opt out of cross-race decisions: the relevance of the particular decision to stereotypes about black Americans. As we expected, and the behavioral data confirm, opting out of cross-race decisions was more pronounced for more sensitive judgments than more innocuous judgments. This increased sensitivity to stereotype-relevant judgments was accompanied by increased VMPFC recruitment, a brain region implicated in self-conscious emotions that plays a central role in the regulation of behaviors and judgments governed by strong social and moral norms. In addition, cross-race decisions—when compared with same-race decisions—were associated with increased activation of ACC, DLPFC and VLPFC, regions involved with conflict, deliberation and inhibition of

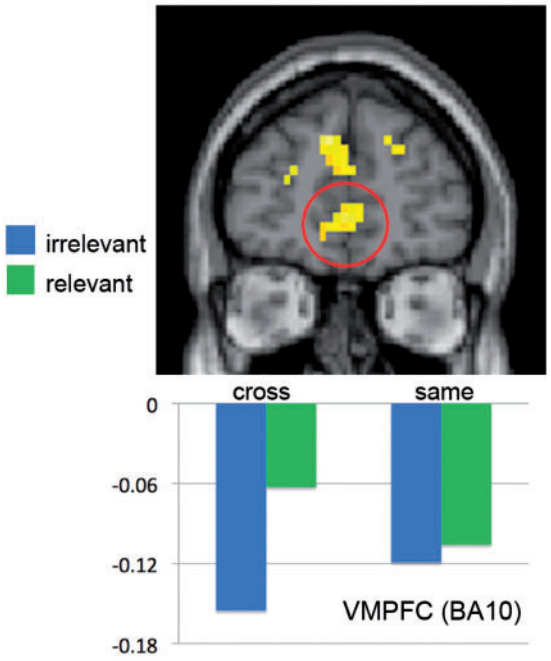

Fig. 2 Percent signal change by condition in a region of the VMPFC (0 49 −2; BA10) where the activity was significantly greater when participants made cross-race choices about stereotype-relevant traits relative to when participants made choices in the other three decision contexts, $P < 0.001$, $k = 10$; k = contiguous voxels. The values were computed by dropping a 10 mm sphere at the cluster's peak (−0 49 −2), extracting the average % signal change with the tools provided by the MarsBar interface (Brett *et al.*, 2002), and then averaging across all participants.

Table 3 Peak coordinates of brain regions where activity during cross-race comparisons involving relevant traits was significantly greater than the average of the other three choice context conditions, $P < 0.001$, $k = 10$; k = contiguous voxels

k	Anatomical location	BA	coordinates			*t*-value
			x	*y*	*z*	
Cross-race/relevant > Average of other three conditions						
35	B. VMPFC	10	0	50	−6	3.77
44	R. DLPFC	9	30	19	32	6.35
95	L. DLPFC	9	−18	45	27	6.32
31	R. DLPFC	9	21	48	27	4.62
44	L. VLPFC	47	−27	20	−14	5.95
44	R. VLPFC	47	33	20	−14	5.52
19	R. VLPFC	47	45	15	−1	5.15
60	L. ACC	32	−12	27	21	4.49
32	R. superior frontal	6	18	11	55	4.66
24	L. superior frontal	6	−9	17	51	4.82
24	R. superior temporal	39	52	−54	21	4.82
10	L. middle temporal	22	57	−41	2	4.41
16	R. hippocampus		24	−38	−2	4.81
10	R. precuneus	31	6	−63	25	3.45
14	R. thalamus		18	−26	1	3.45

The opposite contrast revealed no significant differences (B.) = Bilateral; (L.) = Left; (R.) = Right; (BA) = Brodmann Area.

responses, respectively. The implication of these regions in cross-race decisions offers support for our account that the fear of appearing biased evoked by such situations leads to conflict, greater reflection and a resulting tendency to opt out.

GENERAL DISCUSSION

We demonstrate that while people are willing to make choices between two members of the same group (two white males) on the basis of nothing more than their photographs, they experience racial paralysis when making judgments about members of different groups (a white and a black male), choosing to opt out of such decisions altogether. Somewhat ironically, people's efforts to honor racial neutrality by not choosing provides the very evidence that they do notice race; after all, if they truly did not notice race, they would be as likely to make choices in same-race and cross-race judgments. This tendency to opt out was most pronounced for judgments that were more relevant to stereotypes about blacks—and therefore more likely to elicit concerns about appearing biased—as reflected both in opt out rates and activation in brain regions related to emotionally guided choice. Thus, despite the extraordinary ability of humans to decode faces in order to facilitate judgments and decisions about others, changing the context seems to abruptly change these processes.

This is not to suggest that this reluctance is universal across all individuals and all choices. First, while individuals across the political spectrum are motivated to appear unbiased—reporting more positive attitudes than implicit measures reveal (Nosek *et al.*, 2002)—people's motivation to appear unbiased (Dunton and Fazio, 1997; Plant and Devine, 1998) may predict people's avoidance of choice. Second, situational factors, such as making choices more public or assuaging people's concern about appearing biased (Monin and Miller, 2001) would likely moderate our results. Finally, we have focused on the most salient judgment, between one white male and one black male, but the judgment tasks we use here could incorporate other ethnicities or social categories (e.g. gender or physical disability) to explore more generally the unwillingness to make choices between members of different social groups; indeed, the frequency with which people opt out of decisions between members of different social groups (for example, between an obese and normal weight person) could be used as a metric for concern about appearing biased towards those groups (Crandall *et al.*, 2002).

We conducted a follow-up study to address two alternative explanations. First, it is possible that whites might fail to make a choice in the cross-race condition not due to the different races but rather because the presence of any black face in the array makes choice suspect; our account, however, holds that refusal to choose occurs only when faces are of different races. Second, it is also possible that a failure to choose between a white and black face is due to whites' relative lack of familiarity with black faces (Malpass and Kravitz, 1969), making judgments about such faces more difficult; if this were the case, then judgments between two black faces would be particularly difficult, while our account suggests that these judgments are relatively easy. White participants ($N = 41$) were asked to choose which of two people they thought would perform better in college, and saw either two white faces, two black faces, or one white and one black face. As before, participants were quite willing choosing when the faces were both white (79%); most importantly, they were equally willing when the faces were both black (86%). Overall, then, 82% of participants expressed a preference when the faces were the same race; when presented with one black face and one white face, however, only 46% did so; $\chi^2\ (1) = 5.56$, $P < 0.02$.

Finally, we have defined same-race and cross-race choices as a choice *between* two individuals of the same or different races; of course, it is possible that the race of the decision-maker creates a different kind of same- versus cross-race comparison, where members of different racial groups are relatively more or less likely to experience racial paralysis when confronted with cross-race decisions. The results from our behavioral study suggest that Asian respondents are likely to experience racial paralysis, but would black respondents demonstrate racial paralysis when choosing between white and black targets? We recruited respondents ($N = 296$, $M_{age} = 44.6$, s.d. = 12.8) using an online survey research company to ensure that we had equal numbers

of white ($N = 151$) and black ($N = 145$) respondents. Respondents were randomly assigned to decide either who would perform better in college or who would be more likely to commit a violent crime; in this study, all respondents made cross-race choices (between a white and black candidate). For the college task, overall some 56% opted out, χ^2 (2) = 50.86, $P < 0$ .001; these results were strikingly similar for black and white respondents, with both blacks (57%) and whites (55%) opting out the majority of the time, χ^2s (2) > 21.88, P's < 0.001. Similarly for the crime task, 75% of respondents opted out, χ^2 (2) = 63.70, $P < 0$ .001; these results were again similar for black and white respondents, with both blacks (65%) and whites (82%) opting out the majority of the time, χ^2s (2) > 14.00, P's < 0.001. The overall similarity in responses from black and white respondents, together with the results for Asian participants in the behavioral study, offer some evidence that racial paralysis is more a function of the identities of the targets in a decision than the identity of the decision-maker.

Our task, which focuses on simple judgments between two faces of members of two different social groups, is a clear abstraction from the kinds of situations with which we opened the paper: choosing whom to sit next to on the subway, or talk to in an elevator. As we noted at the beginning, these innocuous real-world situations are in themselves less serious instantiations of more consequential real-world decisions, such as whom to hire, admit to college, or send to jail. Our results suggest that as the stakes of some choice get higher (when making a choice relevant to a racial stereotype feels more likely to indicate that one is biased) the incidence of racial paralysis increases. Our proxy for importance was the relevance of the judgment to some stereotype about blacks. We can only imagine the racial paralysis that might ensue during discussions of real-world impactful decisions, where speaking in favor of a white candidate over a black candidate makes one appear racist, whereas speaking in favor of a black candidate over a white candidate can make one appear as though one is trying too hard *not* to appear racist. We suspect that in such discussions, when people are forced to make some decision at the end of the day, decision makers rely on other strategies to avoid the appearance of bias, such as deferring to members of minority groups (Crosby *et al.*, 2008).

REFERENCES

Apfelbaum, E.P., Sommers, S.R., Norton, M.I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology*, *95*, 918–32.

Beer, J.S., Heerey, E.H., Keltner, D., Scabini, D., Knight, R.T. (2003). The regulatory function of self-conscious emotion: insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*, 594–604.

Blair, I.V., Judd, C.M., Sadler, M.S., Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. *Journal of Personality and Social Psychology*, *83*, 5–25.

Blanchard, F.A., Lilly, T., Vaughn, L.A. (1991). Reducing the expression of racial prejudice. *Psychological Science*, *2*, 101–105.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage*, *16*, abstract 497 (available on CD-ROM).

Camille, N., Coricelli, G., Sallet, J., Paradat-Diehl, P., Duhamel, J.R., Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, *304*, 1167–70.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D.C., Cohen, J.D. (1998). Anterior cingulated cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–9.

Casey, B.J., Castellanos, F.X., Giedd, J.N., et al. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 374–83.

Chernoff, H. (1973). Use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, *68*, 361–8.

Christoff, K., Gabrieli, J.D.E. (2000). The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168–86.

Crandall, C.S., Eshelman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, *129*, 414–46.

Crandall, C.S., Eshleman, A., O'Brien, L.T. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*, 359–78.

Crosby, J.R., Monin, B., Richardson, D. (2008). Where do we look during potentially offensive behavior? *Psychological Science*, *19*, 226–8.

Damasio, A.R., Tranel, D., Damasio, H. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. In: Levin, H.S., Eisenberg, H.M., Benton, A.L., editors. *Frontal Lobe Function and Dysfunction*. New York: Oxford University Press, pp. 217–29.

Devine, P.G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

Devine, P.G., Elliot, A.J. (1995). Are racial stereotypes really fading? The Princeton Trilogy revisited. *Personality and Social Psychology Bulletin*, *21*, 1139–50.

Dovidio, J.F., Gaertner, S.L. (1981). The effects of race, status, and ability on helping behavior. *Social Psychology Quarterly*, *44*, 192–203.

Dovidio, J.F., Gaertner, S.L. (2004). Aversive racism. In: Zanna, M.P., editor. *Advances in Experimental Social Psychology*, Vol. 36, San Diego, CA: Academic Press, pp. 1–51.

Dovidio, J.F., Kawakami, K., Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–8.

Dunton, B.C., Fazio, R.H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–26.

Dutton, D.G. (1971). Reactions of restaurateurs to blacks and whites violating restaurant dress regulations. *Canadian Journal of Behavioural Science*, *3*, 298.

Fazio, R.H., Jackson, J.R., Dunton, B.C., Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–27.

Fiske, S.T., Cuddy, A.J., Glick, P., Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiack, R.J.S. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2*, 189–210.

Gaertner, S.L., Dovidio, J.F. (1986). The aversive form of racism. In: Dovidio, J.F., Gaertner, S.L., editors. *Prejudice, Discrimination, and Racism*. Orlando, FL: Academic Press, pp. 61–89.

Goel, V., Dolan, R.J. (2000). Anatomical segregation of component processes in an inductive inference task. *Journal of Cognitive Neuroscience*, *12*, 110–9.

Greene, J.D., Sommerville, R.D., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–8.

Hassin, R., Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–52.

Hodson, G., Dovidio, J.F., Gaertner, S.L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, *28*, 460–71.

Ito, T.A., Urland, G.R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616–26.

Johansson, P., Hall, L., Sikstrom, S., Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116–9.

Jost, J.T., Rudman, L.A., Blair, I.V., et al. (2009). The existence of implicit bias is beyond scientific doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–6.

Koenigs, M., Young, L., Adolphs, R., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*, 908–11.

Kowalska, D.M., Bachevalier, J., Mishkin, M. (1991). The role of the inferior prefrontal convexity in performance of delayed nonmatching-to-sample. *Neuropsychologia*, *29*, 583–600.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N.L., Camerer, C.F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *The Journal of Neuroscience*, *29*, 2188–92.

Lieberman, M.D. (2003). Reflective and reflexive judgment processes: a social cognitive neuroscience approach. In: Forgas, J.P., Williams, K.R., von Hippel, W., editors. *Social Judgments: Implicit and Explicit Processes*. New York: Cambridge University Press, pp. 44–67.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 259–89.

Malpass, R.S., Kravitz, J. (1969). Recognition for faces of own- and other-race faces. *Journal of Personality and Social Psychology*, *13*, 330–4.

Mehl, M.R., Gosling, S.D., Pennebaker, J.W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862–77.

Monin, B., Miller, D.T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43.
Montepare, J.M., Opeyo, A. (2002). The relative salience of physiognomic cues in differentiating faces: a methodological tool. *Journal of Nonverbal Behavior*, *26*, 43–59.
Nisbett, R.E., Wilson, T.D. (1977). The halo effect: evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–6.
Norton, M.I., Sommers, S.R., Apfelbaum, E.P., Pura, N., Ariely, D. (2006). Colorblindness and interracial interaction: playing the political correctness game. *Psychological Science*, *17*, 949–53.
Norton, M.I., Vandello, J.A., Darley, J.M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, *87*, 817–31.
Pager, D., Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, *70*, 355–80.
Pearson, A.R., Dovidio, J.F., Gaertner, S.L. (2009). The nature of contemporary prejudice: insights from aversive racism. *Social and Personality Psychology Compass*, *3*, 314–38.
Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M. (1988). Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature*, *331*, 585–9.
Plant, E.A., Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *69*, 811–32.
Plaut, V.C., Thomas, K.M., Goren, M.J. (2009). Is multiculturalism or colorblindness better for minorities? *Psychological Science*, *20*, 444–6.
Raczkowski, D., Kalat, J.W., Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia*, *12*, 43–7.
Richeson, J.A., Nussbaum, R.J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–23.
Richeson, J.A., Shelton, J.N. (2003). When prejudice does not pay: effects of interracial contact on executive function. *Psychological Science*, *14*, 287–90.
Shelton, J.N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes and Intergroup Relations*, *6*, 171–85.
Shelton, J.N., Richeson, J.A., Salvatore, J., Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science*, *16*, 397–402.
Smith, M.L., Cottrell, G.W., Gosselin, F., Schyns, P.G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*, 184–9.
Snyder, M.L., Kleck, R.E., Strenta, A., Mentzer, S.J. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology*, *37*, 2297–306.
Snyder, M., Tanke, E.D., Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, *35*, 656–66.
Stephan, W.G., Stephan, C.W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*, 157–75.
Vorauer, J.D., Gagnon, A., Sasaki, S.J. (2009). Salient intergroup ideology and intergroup interaction. *Psychological Science*, *20*, 838–45.
Vorauer, J.D., Main, K.J., O'Connell, G.B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, *75*, 917–37.
Vorauer, J.D., Turpie, C.A. (2004). Disruptive effects of vigilance on dominant group members' treatment of outgroup members: choking versus shining under pressure. *Journal of Personality and Social Psychology*, *87*, 384–99.
Waltz, J.A., Knowlton, B.J., Holyoak, K.J., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*, 119–25.
Willis, J., Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science*, *17*, 592–8.
Wolsko, C., Park, B., Judd, C.M., Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, *78*, 635–54.
Zebrowitz, L.A. (1997). *Reading Faces: Window to the Soul?* Boulder, CO: Westview Press.
Zebrowitz, L.A., McDonald, S. (1991). The impact of litigants' babyfacedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*, 603–23.

APPENDIX A

Traits used in Experiment 1		Traits used in Experiment 2	
Relevant Traits	*Irrelevant Traits*	*Relevant Traits*	*Irrelevant Traits*
- Intelligent	- Outgoing	- Intelligent	- Outgoing
- Motivated	- Quiet	- Articulate	- Restless
- Articulate	- Restless	- Competent	- Impressionable
- Responsible	- Impressionable	- Polite	- Strict
- Competent	- Strict	- Agreeable	- Opinionated
- Honest	- Opinionated	- Hardworking	- Loyal
- Polite	- Loyal	- Conscientious	- Curious
- Agreeable	- Self-conscious	- Reliable	- Authoritative
- Hardworking	- Curious	- Patient	- Play the guitar
- Conscientious	- Artistic	- Math major	- Have a brother
- Reliable	- Authoritative		
- Patient	- Funny		
- Play golf	- From Canada		
- Cultured	- Play the guitar		
- Math major	- Have a brother		