

Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation

Elizabeth A. Phelps

New York University

Kevin J. O'Connor

Massachusetts Institute of Technology

William A. Cunningham and E. Sumie Funayama

Yale University

J. Christopher Gatenby and John C. Gore

Yale University Medical School

Mahzarin R. Banaji

Yale University

Abstract

■ We used fMRI to explore the neural substrates involved in the unconscious evaluation of Black and White social groups. Specifically, we focused on the amygdala, a subcortical structure known to play a role in emotional learning and evaluation. In Experiment 1, White American subjects observed faces of unfamiliar Black and White males. The strength of amygdala activation to Black-versus-White faces was correlated with two indirect (unconscious) measures of race evaluation

(Implicit Association Test [IAT] and potentiated startle), but not with the direct (conscious) expression of race attitudes. In Experiment 2, these patterns were not obtained when the stimulus faces belonged to familiar and positively regarded Black and White individuals. Together, these results suggest that amygdala and behavioral responses to Black-versus-White faces in White subjects reflect cultural evaluations of social groups modified by individual experience. ■

INTRODUCTION

Over the last several decades, research has shown that expressions of prejudicial attitudes toward Black and White social groups, as measured by self-report, have declined steadily (Biernat & Crandall, 1999; Schuman, Steeh, & Bobo, 1997). In spite of this decline, robust evidence of negative evaluations has been observed on indirect measures that bypass access to conscious awareness and conscious control (Banaji, in press; Cunningham, Nezlek, & Banaji, 2000; Nosek, Cunningham, Banaji, & Greenwald, 2000; Bargh & Chen, 1997; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Devine, 1989). Studies such as these have shown, time and again, negative indirect (automatic) evaluations of and behavior toward Black compared with White Americans. Understanding the nature of these unconscious evaluations of social groups is regarded to be a

primary achievement of the field of social cognition. The present investigation expands research on social evaluation by measuring brain activity, in addition to behavior, with two primary goals: (1) to examine the neural correlates of responses to racial groups, and (2) to examine the relation between individual variability in conscious and unconscious social evaluation and brain activity.

Although the neural systems involved in the evaluation of social groups are likely to be extensive and complex, in the present study, we chose to focus on the amygdala, a subcortical structure known to be involved in emotional learning, memory, and evaluation. The amygdala is critically involved in emotional learning as measured by fear conditioning, a task in which a neutral stimulus comes to acquire emotional properties through direct association with an aversive stimulus (Davis, 1997; LeDoux, 1996; Kapp, Pascoe, &

Bixler, 1984). In humans, the amygdala's role extends beyond fear conditioning to the expression of learned emotional responses that have been acquired without direct aversive experience (Funayama, Grillon, Davis, & Phelps, in press; Phelps, LaBar, et al., 1998). The amygdala has also been shown to play a role in the evaluation of social stimuli in both humans and nonhumans (Adolphs, 1998; Adolphs et al., 1999; Kling & Brothers, 1992). In addition, patients with amygdala damage show deficits in the evaluation of fearful faces, suggesting that it is necessary for learning responses to social and emotional signals (Phelps & Anderson, 1997). In normal adults, the amygdala's involvement in perceiving emotional faces is demonstrated by its preferential response to fearful faces as measured by fMRI (Breiter et al., 1996), even if such faces are presented subliminally (Whalen, 1998).

The amygdala has been shown to be important in numerous forms of emotional learning and evaluation. In humans, however, its role is often limited to the indirect expression of the learned emotional response (Bechara et al., 1995). For example, a classic finding is that the startle reflex response is enhanced or potentiated in the presence of negative stimuli (Grillon, Ameli, Woods, Merikangus, & Davis, 1991; Lang, Bradley, & Cuthbert, 1990). The startle potentiation is used to indirectly indicate the emotional evaluation of the stimulus. Patients with amygdala damage, in contrast to controls, do not exhibit this startle potentiation in the presence of negative stimuli (Angrilli et al., 1996). Interestingly, these patients explicitly rate these stimuli as equally arousing and negative as control subjects (Funayama et al., in press). Given the amygdala's involvement in the indirect expression of learned emotional responses coupled with the importance of learning and memory in social evaluation (see Eagly & Chaiken, 1993) suggests that the amygdala is an obvious starting point to investigate the neural systems underlying the indirect evaluation of social groups.

Using fMRI, we investigated amygdala activity in White American subjects in response to Black and White male faces with neutral expressions. In Experiment 1, the faces presented belonged to individuals who were unfamiliar to the subjects. In Experiment 2, the faces belonged to famous and positively regarded Black and White individuals. In each experiment, we also measured conscious and unconscious evaluations of racial groups. Previous research using behavioral measures with White American samples have shown stronger unconscious negative reactions to Black compared to White social groups (Dasgupta, McGhee, Greenwald, & Banaji, in press; Greenwald, McGhee, & Schwartz, 1998). To the extent that Black faces evoke greater negative emotional evaluations, we should observe greater activity in the amygdala. In particular, we measured the correlation between amygdala activity and measures of unconscious and conscious evaluation.

RESULTS

Experiment 1

White American subjects first participated in the fMRI portion of the experiment that was described as a study about memory for faces. During image acquisition, subjects viewed pictures of Black and White unfamiliar male faces with neutral expressions taken from a college yearbook. For each face, subjects indicated if the face was the same or different than the one immediately preceding it, using a button press. After scanning, we obtained three behavioral responses, one of which was a direct (i.e., conscious) measure of racial evaluation and two of which were indirect (i.e., unconscious) measures of racial evaluation.

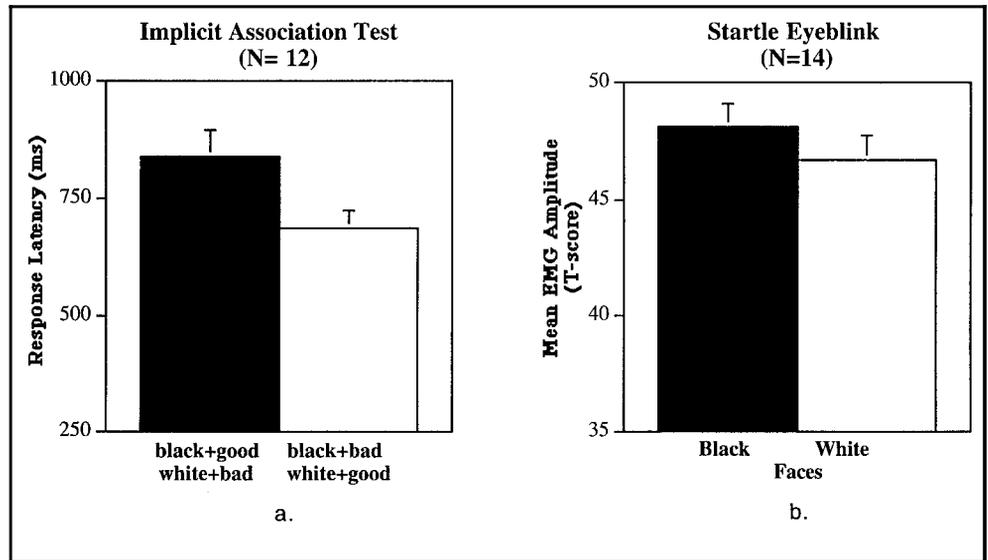
First, subjects took a version of the Implicit Association Test (IAT) (Greenwald et al., 1998) to indirectly measure race bias. The term "bias" in this context refers to the presence of an indirect or noncontrollable behavioral response that exhibits preference for one group over another. It is distinguished from the colloquial use of the term "racial bias" that often implies a purposeful and conscious action of discrimination. The IAT measures the degree to which social groups (e.g., Black vs. White, old vs. young) are automatically associated with positive and negative evaluations (for a demonstration of selected IAT procedures visit www.yale.edu/implicit). Subjects categorized the same faces viewed during imaging as "Black" or "White," while simultaneously categorizing words as "good" (joy, love, peace) or "bad" (cancer, bomb, devil). The difference in response latencies to the Black + good/White + bad pairing compared to the Black + bad/White + good pairing provided the indirect measure of group evaluation. Several studies have now robustly shown negative evaluation among White Americans in the form of faster responding in the Black + bad/White + good pairings (Banaji, in press; Dasgupta et al., in press; Cunningham, Preacher, & Banaji, 2000; Greenwald et al., 1998). The IAT was followed by the Modern Racism Scale, a commonly used measure of conscious, self-reported beliefs, and attitudes toward Black Americans (McConahay, 1986).

Approximately 1 week after the IAT and Modern Racism assessments, we measured the magnitude of the eyeblink startle response to the same Black and White faces as another measure of indirect racial bias. Previous studies examining the startle response suggest that it is potentiated in the presence of negative or fearful stimuli (Lang et al., 1990) and this potentiation has been shown to be related to amygdala function (Davis, 1992; Funayama et al., in press; Angrilli et al., 1996).

Results

Performance on the IAT revealed significantly slower responses to Black + good/White + bad pairings

Figure 1. (a) Mean IAT response latency to Black + good/White + bad and Black + bad/White + good. (b) Mean startle eyeblink response (EMG amplitude) to Black and White faces.



compared to Black + bad/White + good, $t(11) = 7.26$, $p < .001$ (see Figure 1a). This result is consistent with previous studies using White subjects and suggests an unconscious negative anti-Black or pro-White evaluation (Banaji, in press; Dasgupta et al., 2000; McConahay, 1986). There was a trend towards greater or potentiated startle eyeblink when viewing Black compared to White faces, $t(13) = 1.33$, $p = .10$, one-tailed (see Figure 1b). This pro-White race bias on the indirect measures (IAT and eyeblink startle) was in contrast to responses on the Modern Racism Scale where subjects consciously expressed pro-Black beliefs and attitudes. The average score for subjects was 1.89 (with 6 being strongly anti-Black and 1 strongly pro-Black) with a standard deviation of .66 and an effect size of -2.44 (Cohen's d).

With the imaging data, we were primarily interested in activity of the amygdala and limited our image acquisition to slices covering this region. Several previous studies assessing amygdala activity using fMRI in human subjects (Buchel, Morris, Dolan, & Friston, 1998; LaBar, Gatenby, Gore, & Phelps, 1998; Phelps, O'Connor, et al., 1998; Whalen et al., 1998; Breiter et al., 1996) and electrophysiology in nonhuman animals (Quirk, Armony, & LeDoux, 1997; Maeda, Morimoto, & Yanagimoto, 1993; Pascoe & Kapp, 1985) have found that this region responds maximally to the onset and early presentations of a stimulus with emotional significance, including emotional faces. In light of these results, we compared early responses of the amygdala to Black and White faces. To localize responses in the amygdala we used a region-of-interest (ROI) analysis. With this ROI

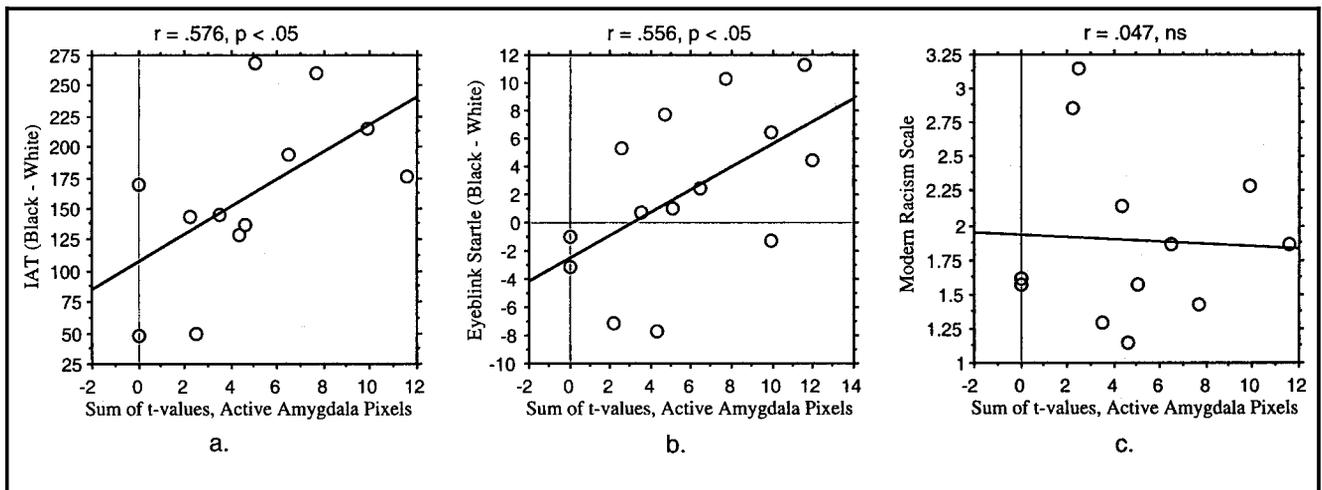


Figure 2. Correlations between the magnitude of amygdala activation to Black-versus-White faces as measured by the sum of the t values for active amygdala pixels and behavioral measures: (a) IAT response latency for Black-White, (b) difference in eyeblink startle response to Black-White faces, and (c) score on the Modern Racism Scale. Similar results were obtained when magnitude of amygdala activation was assessed by counting the number of active amygdala pixels: IAT $r = .52$, $p < .08$; startle eyeblink: $r = .54$, $p < .07$; Modern Racism Scale, ns.

analysis, we assessed the strength of amygdala activity for each individual subject. This ROI analysis revealed that the majority of White subjects showed greater amygdala activation when viewing unfamiliar Black compared to White faces. These data also showed that the extent of amygdala activation to Black-versus-White faces varied across subjects. Eight subjects showed greater amygdala activation in the Black-versus-White comparison than the White-versus-Black comparison. An additional four subjects showed some evidence of significant amygdala activation in the Black-versus-White comparison. As a result of this variability, a group composite analysis did not show significant amygdala activity.

We were particularly interested in this variability because it offered an opportunity to assess the relationship between activity in the amygdala and behavioral measures of race bias. There was a significant correlation between bias in response time on the IAT and strength of the amygdala activation to Black-versus-White faces (see Figure 2a). There was also a significant correlation between amygdala activity and the potentiation of the eyeblink startle response to Black faces (see Figure 2b). There was no correlation between amygdala activity and the conscious measure of racial attitudes assessed with the Modern Racism Scale (see Figure 2c). In addition, we correlated all three behavioral measures with the strength of amygdala activation in the White-versus-Black comparison and no tests reached statistical significance.

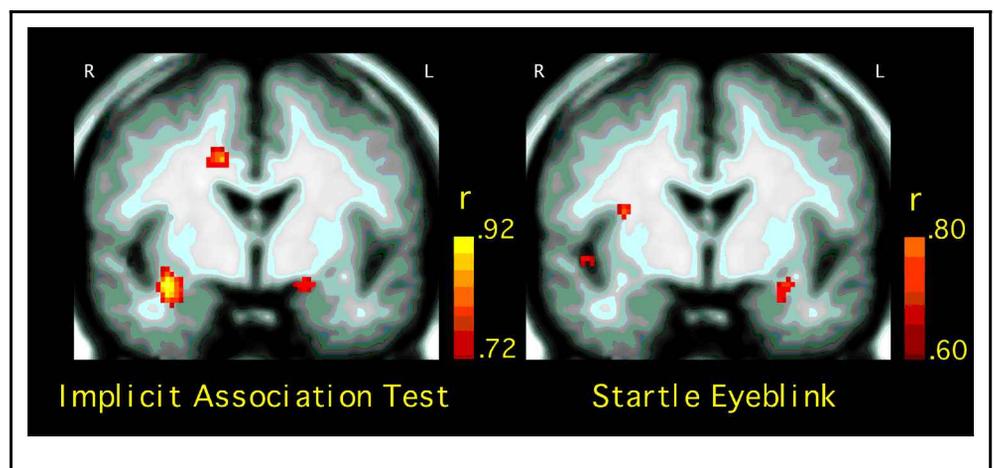
In order to determine the precise brain regions whose activity were most strongly related to performance on indirect measures of racial bias, composite correlation maps were generated. This technique, of generating composite images portraying regions where variability in brain activity is correlated with variability on a behavioral measure, has been previously used in PET (Hamann, Ely, Grafton, & Clinton, 1999; Cahill et al., 1996) and is conceptually similar to the selection of event

types based on behavior used in event-related fMRI studies (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998). On the composite image, individual subjects' activation in a region was correlated with (1) the magnitude of the IAT effect, (2) the magnitude of the startle eyeblink potentiation to Black-versus-White faces, and (3) the score on the Modern Racism Scale. The resulting correlation values (r) are plotted on the composite anatomical image, displaying regions where the strength of activation to Black-versus-White faces is correlated with the magnitude of the behavioral response.

As can be seen in Figure 3, activation in the left amygdala is correlated with both the IAT reaction-time bias (Figure 3a) and the startle eyeblink potentiation bias (Figure 3b, see figure caption for all Talairach & Tournoux, 1998 coordinates). Consistent with the results from the ROI analysis, there was no region within the amygdala where activation was correlated with performance on the Modern Racism Scale. As can be seen in Figure 3a, there are two additional regions that were correlated with IAT reaction time bias. There was a large region of activation that extends from the right amygdala to the inferior insular cortex. The insular cortex has numerous reciprocal connections with the amygdala and is often active in tasks assessing emotional responses (Shi & Davis, 1999; Ploghaus et al., 1999). In addition, a region in the anterior cingulate was correlated with IAT performance. The anterior cingulate is thought to play a role in attentional processes and is often active in tasks where there is response competition, such as the Stroop task (Pardo, Pardo, Janer, & Raichle, 1990). As seen in Figure 3b, there were also two small regions within the more superior region of the insular cortex that were correlated with the magnitude of eyeblink startle potentiation to Black faces. However, the only common region of activity on the two correlation maps examining both indirect behavioral measures of racial bias was the left-superior amygdala.

Figure 3. Composite correlation maps displaying regions where magnitude of activation to Black-versus-White faces is correlated with indirect behavioral measures.

(a) IAT response latency Black-White, regions of significant correlation include: left-superior amygdala (Talairach & Tournoux coordinates: $-17.6, -5, -10.8$), right amygdala extending to the inferior insula ($31.7, -5, 12.2$), and right anterior cingulate ($14.1, -5, 36.1$). (b) Eyeblink startle difference Black-White, regions of significant correlation include: left-superior amygdala ($-22.1, -5, -11.7$) and two small regions in the right insular cortex ($31.8, -5, 17.1$; $41.4, -5, -2.4$).



Discussion

These data show for the first time that representations of social groups that differ in race evoke differential amygdala activity and that such activation is related to unconscious social evaluation. Notably, variability in amygdala activation among White subjects is correlated with negative indirect responses to Black compared to White faces on behavioral measures. Additionally, this relationship between amygdala activity and indirect measures of racial bias is in contrast to the lack of such a relationship with the direct or conscious measure of racial attitudes.

For both indirect measures, the region in the amygdala most strongly correlated with negative evaluation was the left-superior amygdala. This is of interest because this region is also known to be activated when viewing (supraliminally presented) faces with fearful versus neutral expressions (Breiter et al., 1996).

Experiment 2

We propose that the variability in the response of the amygdala to Black compared with White faces in Experiment 1 is likely to be a reflection of culturally acquired knowledge about social groups filtered through individual experience. We expected that the greater amygdala activity in response to unfamiliar Black faces is the result of a general learned negative evaluation of Black Americans (Adolphs, 1998; Fiske, 1998). If the results of Experiment 1 reflect a learned negative evaluation of the group, the pattern of results from Experiment 1 should disappear when presenting exemplars of Black Americans who are as familiar and well liked as White Americans.

To demonstrate a boundary condition on the results obtained in Experiment 1, we conducted a similar fMRI study and used the same behavioral measures with an independent group of White American subjects. However, the faces in both categorization tasks (scanning and IAT) belonged to famous and positively regarded Black and White males. The Black and White individuals portrayed were roughly equivalent in degree of fame, age, and domain of achievement.

Results

The data analysis for Experiment 2 was identical to Experiment 1. Consistent with Experiment 1, subjects consciously expressed pro-Black attitudes on the Modern Racism Scale. The average score was 1.92 with a standard deviation of .67 and an effect size of -2.35 (Cohen's d). However, unlike Experiment 1, the two indirect measures yielded different patterns of results. The IAT continued to show positive evaluation of famous White compared to famous Black faces, $t(12) = 3.61$, $p < .01$, although there was a reduction com-

pared to Experiment 1 in the mean difference in response time between the Black + good/White + bad and Black + bad/White + good conditions. In contrast, there was no evidence of eyeblink startle potentiation to the Black famous faces. In fact, the mean eyeblink startle response was slightly greater to the White famous faces. The continued observation of race bias on the IAT is likely to be a result of the emphasis that task places on attention to Black and White group labels in categorization, unlike the other behavioral tasks that did not require identification of the race of the face stimuli. It is clear from other research using the IAT that the particular labels that are used are critical in the evaluative effect that is produced (Mitchell, Nosek, & Banaji, 2000). By using labels that allow categorization on dimensions other than race, it is possible to elicit positive evaluations of familiar and positively regarded Black individuals. If such a task were used in the present study, we would expect no differences in evaluation of White and Black faces.

The imaging data revealed no consistent pattern of amygdala activity when White subjects viewed well-known Black and White faces. Although some subjects showed evidence of amygdala activation, this was observed equally often in the Black-versus-White and White-versus-Black comparisons. More importantly, there were no significant correlations between the indirect or direct measures of racial bias and the strength of amygdala activation for either the Black-versus-White or White-versus-Black comparisons.

Discussion

These results suggest that the amygdala's response to Black faces in White subjects is not observed when they are familiar and positively regarded. With this data, however, we cannot distinguish responses due purely to familiarity from those due to positive evaluation, independent of familiarity. In fact, these two factors may be difficult to untangle because of the documented role of familiarity in producing positive evaluation (Kunst-Wilson & Zajonc, 1980). However, there is evidence to suggest that race bias cannot be explained merely in terms of familiarity. After controlling for the effects of familiarity on the IAT test, a preference for White over Black still remains (Dasgupta et al., 2000; Ottaway, Hayden, & Oakes, in press). In addition, studies examining responses to familiar emotional scenes suggest that the presentation of familiar, negative stimuli will result in the potentiation of the startle response (Funayama et al., in press; Angrilli et al., 1996). Although we cannot rule out a role for familiarity in the differences observed between Experiments 1 and 2, previous behavioral studies suggest that the positive evaluation of the famous individuals is also a significant factor. Finally, we express caution in interpreting this null result, especially as it stands in comparison to the significant findings of

Experiment 1. These were, after all, independent experiments and future research ought to test the main variable of interest in a single study.

GENERAL DISCUSSION

These studies have shown for the first time that members of Black and White social groups can evoke differential amygdala activity and that this activity is related to unconscious social evaluation. In Experiment 1, the strength of amygdala activation to Black-versus-White faces was correlated with two indirect (unconscious) measures of race evaluation (IAT and potentiated startle), but not with the direct (conscious) expression of race attitudes. In Experiment 2, these patterns were not obtained when the faces observed belonged to familiar and positively regarded Black and White individuals. Together, these results suggest that the amygdala response to Black-versus-White faces in White subjects is a function of culturally acquired information about social groups, modified by individual knowledge and experience.

Although the present studies found that activity in the amygdala to unfamiliar Black-versus-White faces is related to indirect measures of race bias, these results do not specify a particular behavioral function for the observed amygdala activation. The neural systems underlying racial evaluation are most likely extensive and expand beyond the amygdala. Results from previous studies showing that the amygdala is not necessary for conscious learning about the emotional nature of stimuli (Phelps, LaBar, et al., 1998; Bechara et al., 1995; LaBar, Ledoux, Spencer, & Phelps, 1995) suggest that it is not likely to play a role in the formation of conscious attitudes toward social groups. Our results indicate that the amygdala may be specifically involved in indirect or nonconscious responses to racial groups. One possible mechanism by which the amygdala may affect racial responses is suggested by studies showing its involvement in nonconsciously signaling the presence of stimuli that have acquired an emotional significance based on previous experience (Whalen, 1998; Whalen et al., 1998).

Because the combination of procedures used in these studies are novel, several caveats are in order. It is important to reaffirm that although we have obtained significant correlations between amygdala activity and indirect behavioral measures of racial bias, these data cannot speak to the issue of causality. Our own interpretation is that both amygdala activation as well as behavioral responses of race bias are reflections of social learning within a specific culture at a particular moment in the history of relations between social groups. Specifically, the influences that predict such responses (both brain and behavior) may include knowledge of cultural evaluations of social groups, personal experience with social group members, and one's own group membership.

An obvious question regarding one's own group membership as a determinant of race bias concerns the likely performance of Black subjects in these studies. Although we have yet to collect such data, we do offer a speculation. From preliminary IAT data it appears that Black Americans show more favorable evaluations of Blacks than do White Americans. However, they do not show as strong an in-group preference as White subjects. We take this finding to show that Black Americans' own indirect evaluations may be determined in part by the larger culture's negative evaluation of their group (Banaji, in press). Based on the results of the present study, we expect to see less amygdala activation for Black subjects in response to Black-versus-White faces. Importantly, we expect the correlations in the data of Black subjects to mimic the data observed here. That is, the extent to which individual Black subjects show overall greater favorability toward their own group on the IAT, we expect greater relative amygdala activity to unfamiliar White compared to Black faces.

For a century, psychologists have measured attitudes toward social groups as if they existed solely in conscious form. Recent research in social cognition has shown that unconscious social evaluations, however dissociated they may be from their conscious counterparts, are robust and reliable (Fiske, 1998; Greenwald & Banaji, 1995). Unless one is socially isolated, it is not possible to avoid acquiring evaluations of social groups, just as it is not possible to avoid learning other types of general world knowledge. Having acquired such knowledge, however, does not require its conscious endorsement. Yet such evaluations can affect behavior in subtle and often unintentional ways (see Bargh & Chen, 1997; Fazio et al., 1995).

In the present studies, we have for the first time related indirect behavioral measures of social evaluation to neuronal activity. Understanding the mechanisms underlying these indirect responses can initiate discovery of the means by which they are learned and modulated. Besides the finding itself of a relationship between brain activity and behavioral race bias, such data have the potential to shift orthodox thinking about the separation of social, mental, and physical spheres. They start to reveal how social learning and evaluation are rooted in the ordinary mechanics of mind and brain.

METHOD

Experiment 1

Subjects

Fourteen subjects (seven female, seven male) were submitted for final analysis. A total of 20 subjects were run. All subjects gave informed consent. Six subjects were excluded because center of mass motion during scan exceeded our criterion of 0.33 pixels in any direc-

tion. There were no systematic differences observed between male and female subjects.

Stimuli

Subjects were presented photographs of nine Black and nine White male faces, with neutral facial expressions. The photographs were taken from college yearbooks. All photographs were in black and white and depicted men with short hair, no facial hair, and no distinctive clothing.

fMRI: Procedure and Parameters

Prior to image acquisition, the anterior and posterior commissures (AC and PC, respectively) were localized for slice orientation. Whole-brain sagittal T1-weighted anatomical images were acquired using a spin echo pulse sequence (5 mm contiguous slices, TE = 12 msec, TR = 600 msec, matrix size = 256×192 , in-plane resolution = 1.56×1.56 mm, and FOV = 40×40 cm). Five 6-mm coronal slices (slice skip = 2 mm) were then prescribed perpendicular to the AC–PC line, with the middle slice centered on the amygdala. Amygdala localization was accomplished by placing the middle (third) slice 4–5 mm posterior to the AC in the midsagittal view and assessing the position of the amygdala in the subsequent coronal sections using anatomical landmarks and a standardized atlas (Talairach & Tournoux, 1998). During the study, echoplanar functional images were acquired using an asymmetric spin echo pulse sequence (TE = 30 msec, echo offset = 30 msec, TR = 1.5 sec, in-plane resolution = 3.125×3.125 mm, matrix size = 128×64 , and FOV = 40×20 cm). The experimental paradigm was a standard block design consisting of six blocks of each stimulus condition, Black and White, resulting in 12 trials. During image acquisition subjects were presented pictures of nine Black and nine White unfamiliar male faces. The pictures were presented as six blocks each of Black and White faces. Each face was presented for 2 sec and each block lasted 18 sec. The face presented on any given trial was randomly selected from among the set of nine for that racial group. For each face, subjects indicated with a button press if the face was the same or different than the one immediately preceding it. During each trial, 12 images were acquired over 18 sec (TR = 1.5 sec). In the described analysis, subjects' functional activation was averaged across the first six images of the first three trials of each condition.

fMRI Analysis

A *t* test analysis was conducted on the individual subject data. Resultant *t* maps were generated by subtraction to reveal differential activation between conditions. Pixels

showing significant differential activation ($p < .05$, uncorrected) were used in subsequent ROI and correlation analyses. For each subject, ROI analyses were performed on the amygdala. This region was first outlined on anatomical images. The functional maps of Black versus White were then superimposed on the anatomical images to identify active pixels within these regions. To determine if an individual subject showed greater amygdala activation to Black-versus-White faces, a count of significantly active amygdala pixels was conducted (*t* value, $p < .05$, cluster value = 1) for the Black-versus-White and the White-versus-Black comparisons. To calculate the correlation between a behavioral response and activation, the behavioral measure of interest was regressed on the magnitude of amygdala activity. The magnitude of amygdala activity was calculated using the sum of *t* values for significantly active pixels within each ROI and also counting the number of significantly active pixels (*t* value, $p < .05$, cluster value = 0, see Constable et al., 1998 for more details). Each measure of amygdala activity was used in separate regression analyses and virtually identical results were obtained. Additional regression analyses were performed examining the White-versus-Black activation and no significant correlations were obtained.

Two types of group analyses were conducted. The first examined the overall group effect for the Black-minus-White comparison. Activation maps were generated by in-plane transformation of the individual subjects' statistical parametric maps (SPMs) and the anatomic images into a proportional three-dimensional grid defined by Talairach and Tournoux (1998). The individual activation maps were smoothed using a Gaussian filter (FWHM = 5.2 mm). In order to obtain *p* values for significantly active pixels across subjects, a contrast composite map was generated using a randomization test to create a distribution of task-related *t* values for each pixel (Bullmore et al., 1998), from which *p* values were calculated. The *p* value for each pixel was overlaid upon a mean anatomic image.

The second group analysis generated composite correlation maps designed to determine the pixels for which Black–White activation magnitude was correlated with behavioral performance. The SPMs and the anatomic images for the 12 individual subjects who had valid IAT results were transformed by in-plane transformation into a proportional three-dimensional grid defined by Talairach and Tournoux (1998). The individual activation maps were smoothed using a gaussian filter (FWHM = 5.2 mm). For each pixel, the subjects' Black-minus-White *t* values were correlated with the (1) difference in response latency on the IAT, (2) eyeblink startle difference to Black-versus-White faces, and (3) score on the Modern Racism Scale. The *r* value for each pixel was overlaid upon a mean anatomic image. Only significantly correlated pixels are displayed.

Subjects were asked to categorize the same faces they viewed during imaging as Black or White, while simultaneously categorizing words as good (joy, love, peace) or bad (cancer, bomb, devil). For half of the trials, subjects were asked to press a right button if the stimulus was either a White face or a good word and a left button if the stimulus was either a Black face or a bad word. For the remaining half of the trials, the pairings were reversed. The two conditions were counterbalanced. The difference in speed to respond to the Black + good/White + bad pairing compared to the Black + bad/White + good pairing provided the indirect measure of group evaluation. Some subjects were given the IAT immediately after scanning and others were given the test a week later. There was no difference in IAT performance related to the timing of the test. Two subjects were dropped from the IAT analyses; the first because of an unusually high error rate in categorization (28%), the second because of an error in the IAT program that resulted in the subject receiving only stimuli corresponding to the left key in the second half of the Black + good categorization condition.

Eyeblink Startle

The startle response is a defensive reflex, one component of which is an eyeblink response (Lang et al., 1990; Knorkski, 1967). Using electromyogram (EMG) to measure responses of the muscles below the eye, we assessed the magnitude of the eyeblink response as an indication of startle in the presence of the Black and White faces. The startle stimulus was a 50-msec burst of 100-dB white noise that was delivered through headphones. There were six habituation trials to the startle stimulus alone. All 18 faces were presented for 6 sec each. Startle eyeblink was assessed during the presentation of six White and six Black faces. The startle probes occurred 2 to 4 sec following stimulus onset. The eyeblink component of subjects' startle response was measured by EMG (BioPac Systems) and stored off-line for later analysis. Two Ag-AgCl electrodes were placed on the skin over the orbicularis oculi muscle under subjects' left eye. A reference electrode was placed behind subjects' left ear. Prior to analysis, the raw EMG signal was fully rectified, followed by a 10-Hz, two-pole, low-pass filter. The signal was fully integrated and a running value of the area under the curve was calculated. An eyeblink was defined as the difference between the preblink baseline, taken as the mean EMG activity in the 50-msec prior to the startle probe, and the peak amplitude occurring 120 msec following the startle probe. Subjects' EMG amplitudes were standardized (t scores = $z(10) + 50$) before analysis due to large between-subject differences in baseline eyeblink amplitude.

The Modern Racism Scale is a commonly used measure of conscious, self-reported beliefs and attitudes toward Black Americans. Examples of items are: "Discrimination against Blacks is no longer a problem in the United States"; "It is easy to understand the anger of Black people in America." Scores on a six-point scale asking for agreement or disagreement with items were computed, with lower scores representing pro-Black and larger scores representing anti-Black beliefs and attitudes. This scale is a standard measure of attitudes and beliefs about the current status and rights of Black Americans and does not tap purely evaluative responses toward the group.

Experiment 2

Subjects

Thirteen subjects (six female, seven male) were submitted to final analysis. A total of 26 subjects were run. All subjects gave informed consent. Thirteen subjects were excluded because center of mass motion during scanning exceeded our criterion of 0.33 pixels in any direction. There were no systematic differences observed between the male and female subjects and these data were combined.

Stimuli

The Black individuals whose faces were portrayed were Muhammad Ali, Arsenio Hall, Bill Cosby, Magic Johnson, Michael Jordan, Martin Luther King Jr., Colin Powell, Will Smith, and Denzel Washington. The White individuals were Larry Bird, Conan O'Brian, Tom Cruise, Harrison Ford, John F. Kennedy, Mark McGwire, Joe Namath, Norman Schwartzkopf, and Jerry Seinfeld. The photographs were taken from published books and the Internet. All photos were in black and white. All of the photographs were of the face and neck only with a frontal view and neutral facial expressions.

Procedures

The procedures and analysis were identical to Experiment 1.

Acknowledgments

We thank David Armor, R. Bhaskar, John Bargh, Siri Carpenter, Geoffrey Cohen, Michael Davis, Thierry Devos, Anthony Greenwald, Richard Hackman, Kristin Lane, Kristi Lemm, Joseph LeDoux, Jeansok Kim, Kevin LaBar, Matthew Lieberman, Brian Nosek, Kevin Ochsner, and Scott Yancey for helpful comments. We especially thank Adam Anderson for suggesting the fMRI composite analysis technique and Pawel Skudlarski for developing the techniques. This research was supported by grants from the McDonnell Foundation (97-26 to EAP), the National Science Foundation (SBR 97099324 to MRB), the

National Institute of Mental Health (MH57672 to MRB), and National Institute of Health (NS3332 to JCG).

For main correspondence and reprints, address requests to Elizabeth A. Phelps, Department of Psychology, New York University, 6 Washington Place, 8th Floor, New York, NY 10003, USA or liz.phelps@nyu.edu. For questions concerning social cognition or the IAT, specifically address correspondence to Mahzarin R. Banaji, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520, USA or mahzarin.banaji@yale.edu.

REFERENCES

- Adolphs, R. A. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.
- Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*, 1111–1117.
- Angrilli, A., Mauri, A., Palomba, D., Flor, H., Birbaumer, N., Sartori, G., & di Paola, F. (1996). Startle reflex and emotion modulation impairment after a right amygdala lesion. *Brain*, *119*, 1991–2000.
- Banaji, M. R. (in press). Implicit attitudes can be measured. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC: American Psychological Association.
- Bargh, J. A., & Chen, M. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*, 541–560.
- Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*, 1115–1118.
- Biernat, M., & Crandall, C. S. (1999). Racial attitudes. In J. P. Robinson, P. H. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 291–412). San Diego: Academic Press.
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S., & Rosen, B. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*, 875–887.
- Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, *281*, 1185–1187.
- Buchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron*, *20*, 947–957.
- Bullmore, E., Brammer, M., Williams, S. C., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., & Sham, P. (1998). Statistical methods of estimation and interference for functional MR image analysis. *Magnetic Resonance in Medicine*, *35*, 261–277.
- Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J., & McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 8016–8021.
- Constable, R. T., Skudlarski, P., Mencl, E., Pugh, K. R., Fulbright, R. K., Lacadie, C., Shaywitz, S. E., & Shaywitz, B. A. (1998). Quantifying and comparing region-of-interest activation patterns in functional brain imaging: Methodology considerations. *Magnetic Resonance Imaging*, *16*, 289–300.
- Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2000). *Conscious and unconscious ethnocentrism: Revisiting the ideologies of prejudice*. Unpublished manuscript: Yale University.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2000). *Psychometric properties of implicit attitude measures: Inter-item consistency, stability, and convergent validity*. Unpublished manuscript: Yale University.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White-Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316–328.
- Davis, M. (1992). The role of the amygdala in conditioned fear. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (225–306). New York: Wiley.
- Davis, M. (1997). Neurobiology of fear responses: The role of the amygdala. *Journal of Neuropsychiatry and Clinical Neurology*, *9*, 382–402.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 680–690.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt.
- 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*, 1013–1027.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.) *The handbook of social psychology* (vol. 2, pp. 357–411). New York: Oxford University Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition* (2nd ed.). New York: McGraw-Hill.
- Funayama, E. S., Grillon, C. G., Davis, M., & Phelps, E. A. (in press). A double dissociation in the affective modulation of startle in humans: Effects of unilateral temporal lobectomy. *Journal of Cognitive Neuroscience*.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
- Greenwald, A. G., McGhee, J. L., & Schwartz, J. L. (1998). Measuring individual differences in social cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Grillon, C., Ameli, R., Woods, S. W., Merikangus, K., & Davis, M. (1991). Fear-potentiated startle in humans: Effects of anticipatory anxiety on the acoustic blink reflex. *Psychophysiology*, *28*, 588–595.
- Hamann, S. B., Ely, T. D., Grafton, S. T., & Clinton, D. K. (1999). Amygdala activity related of enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*, 289–293.
- Kapp, B. S., Pascoe, J. P., & Bixler, M. A. (1984). The amygdala: A neuroanatomical systems approach to its contribution to aversive conditioning. In N. Butters & L. S. Squire (Eds.), *The neuropsychology of memory* (473–488). New York: Guilford Press.
- Kling, A. S., & Brothers, L. A. (1992). The amygdala and social behavior. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 255–306). New York: Wiley.

- Knorowski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago: University of Chicago Press.
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*, 557–558.
- LaBar, K. S., Gatenby, C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Amygdalo-cortical activation during conditioned fear acquisition and extinction: A mixed trial fMRI study. *Neuron*, *20*, 937–945.
- LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *Journal of Neuroscience*, *15*, 6846–6855.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*, 377–395.
- LeDoux, J. E. (1996). *The emotional brain*. New York: Simon & Schuster.
- Maeda, H., Morimoto, H., & Yanagimoto, K. (1993). Response characteristics of amygdaloid neurons provoked by emotionally significant environmental stimuli in cats, with special reference to response durations. *Canadian Journal of Physiology and Pharmacology*, *7*, 374–378.
- McConahay, J. P. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando: Academic Press.
- Mitchell, J., Nosek, B. A., & Banaji, M. R. (2000). *Category salience determines implicit attitudes toward black females and white male targets*. Nashville, TN: Society of Personality and Social Psychology.
- Nosek, B. A., Cunningham, W. A., Banaji, M. R., & Greenwald, A. G. (2000). *Measuring implicit attitudes on the Internet*. Nashville, TN: Society for Personality and Social Psychology.
- Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (in press). Implicit attitudes and racism: The effects of word familiarity and frequency in the Implicit Association Test. *Social Cognition*.
- Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, U.S.A.*, *87*, 256–259.
- Pascoe, J. P., & Kapp, B. S. (1985). Electrophysiological characteristics of amygdaloid central nucleus neurons during Pavlovian conditioning in the rabbit. *Behavioral Brain Research*, *16*, 117–133.
- Phelps, E. A., & Anderson, A. K. (1997). Emotional memory: What does the amygdala do? *Current Biology*, *7*, 311–314.
- Phelps, E. A., LaBar, K. S., Anderson, A. K., O'Connor, K. J., Fulbright, R. K., & Spencer, D. S. (1998). Specifying the contributions of the human amygdala to emotional memory: A case study. *Neurocase*, *4*, 527–540.
- Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Anderson, A. K., Grillon, C., Davis, M., & Gore, J. C. (1998). Activation of the human amygdala by a cognitive representation of fear. *Society for Neuroscience Abstracts*, *24*, 1524.
- Ploghaus, A., Tracey, I., Gati, J. S., Clare, S., Menon, R. S., Matthews, P. M., Rawlins, J., & Nicholas, P. (1999). Dissociating pain from its anticipation in the human brain. *Science*, *284*, 1979–1981.
- Quirk, G. J., Armony, J. L., & LeDoux, J. E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron*, *19*, 1029–1039.
- Schuman, H., Steeh, C., & Bobo, L. (1997). *Racial attitudes in America: Trends and interpretations*. Cambridge: Harvard University Press.
- Shi, C. J., & Davis, M. (1999). Pain pathways involved in fear conditioning measured with fear-potentiated startle: Lesion studies. *Journal of Neuroscience*, *19*, 420–430.
- Talairach, J., & Tournoux, P. (1998). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. New York: Thieme.
- Wagner, A. D., Schacter, D. L., Rotte, N., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191.
- Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*, 177–188.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*, 411–418.