Available online at www.sciencedirect.com
ScienceDirect
Journal of Experimental Social Psychology 44 (2008) 50–64
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Subliminal exposure to faces and racial attitudes: Exposure to Whites makes Whites like Blacks less ☆

Pamela K. Smith [a,*,1], Ap Dijksterhuis [b,1], Shelly Chaiken [c]

[a] *Leiden University, Netherlands*
[b] *University of Amsterdam, Netherlands*
[c] *New York University, USA*

Received 7 January 2006; revised 24 January 2007
Available online 4 February 2007
Communicated by Jeff Stone

Abstract

Despite recent social and political advances, most interracial contact is still superficial in nature, and White individuals interact mainly with other Whites. Based on recent mere exposure research, we propose that repeated exposure to Whites may actually increase prejudice. In a series of experiments, White participants were subliminally exposed to White faces or nothing (control) and then completed various explicit and implicit measures of racial attitudes. Exposure to White faces consistently led to more prejudice by making attitudes toward Blacks more negative, rather than by making attitudes toward Whites more positive. A final experiment demonstrated that the pattern of increased prejudice following exposure to Whites was moderated by the strength of participants' attitudes toward Whites. Only when White attitudes were strong did Black attitudes became more negative after exposure to White faces.

Keywords: Mere exposure; Subliminal; Prejudice; Attitudes toward Whites; Attitude strength

In real life, much of the intergroup contact that White people experience barely qualifies as contact at all. In stores, hallways, and classrooms, on buses, sidewalks, and subways, Whites are exposed to various members of outgroups and hardly ever speak to them, much less interact with them. When asked to think about their good friends in a 1998 poll, only 24% of White U.S. respondents reported having a Black friend (Smith, 2002). Most of the interpersonal contact Whites experience in their lives is instead with other Whites. After all, the average White person in the United States lives in an area that is almost 83% White and only 7% Black (Logan, 2001). In short, most White people are primarily exposed to other Whites.

What are the effects of this repeated intragroup contact on Whites' intergroup attitudes? The contact hypothesis holds that simple contact between groups, such as the Black–White contact described above, is insufficient to improve attitudes toward outgroups (e.g., Dovidio, Gaertner, & Kawakami, 2003; Pettigrew & Tropp, 2006), but it is silent regarding the effects of intragroup contact. However, almost four decades of mere exposure research suggest that even interactionless exposures to members of a particular group should improve attitudes toward that group (Bornstein, 1993; Zajonc, 1968). The mere exposure effect thus suggests an unexpected path to increased prejudice: repeated exposure to ingroup members should improve attitudes toward the ingroup, at the relative expense of the outgroup.

☆ We thank Ran Hassin, Ravit Nussinson, Russell Spears, and the Social Cognition lab group at the University of Amsterdam for their valuable suggestions and comments, and Caitlin Henry and Teun Meurs for their assistance with data collection. Special thanks to Brian Nosek, Thierry Devos, and Stephanie Goodwin for sharing their photographs of Black and White faces with us.
* Corresponding author. Fax: +31 24 361 2677.
E-mail address: p.smith@psych.ru.nl (P.K. Smith).
[1] Present address: Department of Social Psychology, Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.

0022-1031/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2007.01.006

In the present research, we explore how subliminal exposure to White faces affects the attitudes of White individuals toward Whites and Blacks. We focus in particular on prejudice, or negative attitudes toward an outgroup relative to an ingroup, because contact has a stronger effect on affective than on cognitive reactions to outgroups (Tropp & Pettigrew, 2005). Despite changing societal norms, Whites' attitudes toward Blacks remain ambivalent at best (e.g., Gaertner & Dovidio, 1986).

Mere exposure and racial prejudice

The mere exposure effect is one of the most robust findings in social psychology (see Bornstein, 1989, for a review). Just perceiving an object repeatedly—regardless of whether that object is an irregular polygon (e.g., Kunst-Wilson & Zajonc, 1980) or a person (Bornstein, Leone, & Galley, 1987; Moreland & Beach, 1992)—seems to improve a person's attitude toward that object. This effect does not depend on conscious awareness of the stimuli and may even be stronger when the stimuli are presented subliminally (Bornstein & D'Agostino, 1992).

As Bornstein (1993) has suggested, mere exposure is highly relevant for the prejudice problem. However, in over 200 studies of mere exposure, only four investigated how exposure could alter attitudes toward Blacks and Whites (Ball & Cantor, 1974; Cantor, 1972; Hamm, Baum, & Nikels, 1975; Perlman & Oskamp, 1971), and this research has limited ability to address our research question. The studies have a variety of methodological limitations, such as the use of pre-posttest designs that exposed participants to all stimuli at least once, and vague or nonexistent cover stories. They are also open to a wide array of demand effects due to their use of long supraliminal exposure times. Finally, in all these experiments, each participant was exposed to both Black and White faces, rather than just to faces of one race. To investigate the effects of intragroup contact, White participants must only be exposed to Whites.

For mere exposure to be of any importance for understanding prejudice, its effects must also extend beyond the particular stimuli that are presented. Like most prejudice researchers, we are interested in attitudes toward Blacks and Whites in general, not just attitudes toward four or five members of each group. A few mere exposure researchers have looked at generalization of liking to new, related stimuli from the same category (e.g., Gordon & Holyoak, 1983; Kramer & Parkinson, 2005; Rhodes, Halberstadt, & Brajkovich, 2001), finding that the mere exposure effect does seem to generalize within a category. For example, Gordon and Holyoak (1983) exposed participants to letter strings generated by an artificial grammar and found that they preferred new letter strings that also followed this grammar to new, "random" strings. Rhodes et al. (2001) found similar generalization with composite faces that represented averages of previously presented faces. These studies suggest that if stimuli can produce typical mere exposure effects—if repeated exposure to them results in greater liking of them—then this improved attitude should generalize within the particular category. It follows logically that exposure to White faces should lead to greater liking of Whites in general.

Traditional versus generalized mere exposure effects

The notion of within-category generalization makes the mere exposure effect more complex. In essence, there are two kinds of mere exposure effects. The first we will call the *traditional* mere exposure effect, as it is how mere exposure has been classically measured since the original Zajonc (1968) paper. Here participants like old stimuli more than new stimuli from the *same* category. This traditional conceptualization of the mere exposure effect is inextricably bound to the standard design of a mere exposure experiment: Stimuli in previous mere exposure experiment have always been drawn from a single category (Whittlesea & Price, 2001). With such a design, one can only test for the presence or absence of the traditional mere exposure effect. You cannot test for generalization without including other categories of stimuli in the rating phase.

In their second experiment, Monahan, Murphy, and Zajonc (2000) employed just such a design. They exposed some participants to one category of stimuli (Chinese ideographs), others to a different category of stimuli (polygons), and a third group to nothing. Then all participants rated both old and new ideographs and old and new polygons. In other words, participants in the two exposure conditions had to rate old and new stimuli from the category they were exposed to, as well as stimuli from the other, nonexposed category. Under these conditions, participants in the two exposure conditions did not show a traditional mere exposure effect. They liked previously exposed stimuli and new stimuli from the same category equally well. Instead, all stimuli, both old and new, from the exposed category were liked more than stimuli from the other, nonexposed category. Such a result may be seen as a stronger form of the generalization found in the mere exposure research mentioned previously (e.g., Gordon & Holyoak, 1983; Rhodes et al., 2001); in this case no within-category distinctions are made. We will thus refer to such an effect as a *generalized* mere exposure effect. Having participants evaluate stimuli from multiple categories, including nonexposed categories, not only allows one to observe a generalized mere exposure effect, but it also seems to weaken or eliminate a traditional mere exposure effect.

This type of multiple-category design seems well-suited to research on racial attitudes. Attitudes toward ingroups and outgroups are rarely discussed in isolation. Researchers often compare the two, using attitudes toward the ingroup as a benchmark to determine whether attitudes toward an outgroup are positive, negative, or neutral (e.g., Fiske, 1998). There is a long tradition, beginning at least with Allport (1981/1954), of treating relative preferences as prejudicial. Hence, most popular modern

measures of prejudice involve comparisons between attitudes toward Whites and attitudes toward Blacks (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998). A person's degree of prejudice is then indicated by how much more the person likes Whites relative to Blacks. If we truly want to examine the effects of exposure on prejudice, our participants need to evaluate both Whites and Blacks.

Based on this body of mere exposure research, we can make two major predictions about the effects of subliminal exposure to White faces on the racial attitudes of Whites. First, if participants rate faces of both races, we will not find traditional mere exposure effects. Participants exposed to faces of one race will express the same attitude toward old and new faces of that race. Second, we do expect to find generalized mere exposure effects: Exposing White participants to White faces should improve their attitudes toward Whites relative to their attitudes toward Blacks.

Overview of experiments

We ran a series of five experiments to explore how subliminal exposure to White faces affects the racial attitudes of Whites. In all experiments, participants were randomly to assigned to one of two exposure conditions. The White-exposure condition was subliminally exposed to White faces. We used subliminal exposure to avoid rousing the suspicion of our participants. Additionally, both traditional and generalized mere exposure effects have been found with subliminal exposures (Bornstein & D'Agostino, 1992; Monahan et al., 2000).

The control condition completed the same exposure task, but they were exposed to a blank screen instead of faces. We included this control group to provide a baseline to which we could compare the White-exposed condition. After all, if White-exposed participants reported liking Whites more than Blacks, this would not be sufficient evidence of increased prejudice due to exposure: The stimuli themselves might lead anyone to express such bias. Thus, our hypothesis involved a significant Race by Exposure Condition interaction: White-exposed participants should demonstrate more prejudice *relative to* control participants.

In Experiments 1a and 1b, prejudice was measured via explicit liking ratings of Black and White faces. That is, all participants saw a series of Black and White faces and evaluated them. For White-exposed participants, half of the White faces they rated were old (previously exposed), half of the White faces were new (nonexposed), and all of the Black faces were new (nonexposed). For control (no exposure) participants, all Black and White faces were new (nonexposed). We expected that White-exposed participants would not distinguish between faces of the exposed race that were old versus new, and instead would rate all faces of the exposed race as equally positive. We also predicted that participants exposed to White faces would show greater prejudice—greater liking for Whites relative to Blacks—as compared to control participants.

We next explored potential boundary conditions of our effects. In Experiment 2, we used a fully between-participant design so that each participant only expressed attitudes toward one race, to test whether our effects were dependent on each participant evaluating both Whites and Blacks. Experiment 2 also tested whether our effects generalized beyond simple liking by using a different measure of prejudice, namely ratings of defendant guilt. Experiment 3 further extended our findings by using an implicit measure of prejudice. Here participants completed a sequential priming task that measured whether White versus Black faces affected the accessibility of positive and negative concepts (Fazio et al., 1995).

Finally, Experiment 4 addressed a consistent but unexpected component of our results, using the strength of participants' initial attitudes as a moderator. Details will be discussed later in the article.

Experiment 1a

Method

Participants and design. White undergraduate students from New York University ($N = 73$) took part in the experiment as part of a course requirement. Five participants were dropped from the analyses: four because they did not follow instructions, and one due to a computer problem. Thus, 68 participants (23 males, 45 females) were included in the final analyses. Average age was 19.29 years ($SD = 0.83$).

Participants were randomly assigned to one of two exposure conditions: exposure to White faces or no exposure (control). The major comparisons involved a 2 (Race of Rated Photo: Black face vs. White face) $\times$ 2 (Exposure Condition: White-exposed vs. control) mixed-model factorial, with the last factor between subjects. Additionally, within the White-exposed condition, some White faces in the rating task were old (i.e., had been presented subliminally in the exposure task), and others were new.

Photographic stimuli. Photographs of Black and White students were selected from recent high school yearbooks and piloted on 14 White undergraduate students. These students rated the photographs on 7-point scales on a series of questions, including four target questions on liking of the person, attractiveness of the person, viewing the person favorably, and pleasantness of the photograph. Two sets of six photographs of Black faces (3 male and 3 female), and two sets of six photographs of White faces (3 male and 3 female) were selected (Sets 1 and 2). Average ratings of these sets across the four target questions did not differ reliably, $F(1,13) = 1.54$, $p = .24$, for Black sets, $F < 1$ for White sets. Across sets, Black and White faces were also rated similarly, $F < 1$. The sets were all given moderate ratings (Ms = 3.61 to 3.80) on the target questions. All photographs were grayscale and approximately 220 pixels wide and 240 pixels tall.

Faces of male and female Asian students were also obtained from high school yearbooks but were not normed. These faces served as fillers in the explicit attitude measure.

Procedure. Groups of up to five participants took part in the experiment at one time, with each participant seated in front of a computer in a separate room. The experimenter explained that she was looking at how rapidly people are able to process visual information, and that they would complete a few tasks on computer and fill out some questionnaires.

First, participants completed the exposure task on computer. Labeled as the "attention task," it was described as testing how quickly people process visual information with different levels of distraction. In each trial, a fixation cross appeared for 500 ms, followed by a forward mask for 13 ms, then a photograph (White-exposure condition) or a blank screen (control condition) for 13 ms, then a backward mask for 26 ms, and finally a gray square with colored dots on it. The forward and backward masks were black and white houndstooth patterns of the same size and shape as the gray square. Participants were told to press the space bar as soon as this gray square appeared. The gray square stayed on the screen until participants pressed the space bar. After 1 s, the next trial began.

There were 10 blocks of six trials each. In the White-exposure condition, the same set of three male photos and three female photos (either Sets 1 or 2) was shown in each block. Thus, these participants were exposed 10 times to six different photos of White faces. (The number of stimulus presentations was chosen based on the Bornstein (1989) meta-analysis.) In the control condition, participants were repeatedly exposed to a blank screen. Afterwards all participants completed a filler questionnaire apparently examining their verbal versus visuospatial ability to bolster the cover story.

A computerized photograph rating task followed, which served as the measure of participants' explicit attitudes toward Blacks and Whites. Each photograph was presented with the question, "How much do you like this person?" and a 7-point rating scale ($1 = \textit{not at all}$, $7 = \textit{very much}$) beneath it. The photographs in this task included the photographs that had been subliminally shown to participants in the White-exposed condition, as well as new photographs. Participants were told these photographs were going to be used in a new version of the attention task, but that norms first had to be collected on them so they could be standardized. The instructions emphasized that there were no right or wrong answers, and that participants should respond based on their initial impression of the person.

Participants completed three blocks of 12 photographs. Each block consisted of 1 Black female, 1 Black male, 1 White female, and 1 White male photograph from Sets 1 and 2, as well as 2 Asian female and 2 Asian male photographs (image order was randomized within block). The Asian photos were included to decrease the likelihood of participants guessing our interest in their attitudes toward Blacks and Whites.

For White-exposed participants, half of the White photographs in the explicit attitude measure were the set of photographs that they had been exposed to earlier. The remaining six White photographs were the other, nonexposed set. The Black photographs were all new. For control participants, all photographs were new.

Finally, participants were taken through a funnel debriefing procedure (Bargh & Chartrand, 2000) to probe for suspicion and for perception of the subliminal photographs. Then they were debriefed and thanked for their participation.

Results and discussion

The computer program recorded how long participants took to respond to each liking and attractiveness question. All responses that took less than 150 ms were dropped (0.46% of responses), as they were most likely due to mistaken key presses.

Traditional mere exposure effects on liking. Based on Monahan et al. (2000), we did not expect any traditional mere exposure effects (i.e., we did not expect more positive attitudes toward old versus new faces of the same race). Two new variables were created for White-exposed participants only: one indicating their liking for previously exposed faces, and one indicating their liking for new faces of the same race. A 2 (Type of Photo: previously exposed vs. same-race new) repeated-measures ANOVA run on liking ratings for these participants was not significant, $F < 1$. As predicted, people exposed to faces of a particular race held the same explicit attitude toward these old, familiar faces and new, same-race faces. Thus, in the remaining analyses, we collapsed across this factor (previously exposed vs. same-race new).

Generalized mere exposure effects on liking. Did exposure to photographs of people of a particular race affect participants' explicit liking of Black versus White faces in general? A 2 (Race of Rated Photo: Black face vs. White face) × 2 (Exposure Condition: White exposed vs. control) mixed-model ANOVA was run on liking ratings, with race of photo varied within subjects. A significant main effect of Photo Race, $F(1,66) = 12.82$, $p = .001$, $\eta_p^2 = .16$, indicates that participants generally liked Black faces more than White faces. However, this effect was moderated by a significant Photo Race by Condition interaction, $F(1,66) = 9.24$, $p = .003$, $\eta_p^2 = .12$. The means for this interaction are listed in Table 1.

Control participants liked Black faces significantly more than White faces, $Fs(1,66) = 19.10$, $p < .001$, $\eta_p^2 = .22$. However, participants who had been exposed to White faces liked Black and White faces equally well, $F < 1$. In other words, exposure to White faces eliminated the preference for Blacks relative to Whites shown by control participants.

Table 1
Ratings of black and white targets by exposure condition, Experiments 1–3

Race of targets	Exposure condition	
	White-exposed	Control
Experiment 1a—liking		
Black	4.31 (0.54)	4.64 (0.51)
White	4.28 (0.57)	4.23 (0.55)
Experiment 1b—liking		
Black	4.24 (1.38)	5.14 (1.02)
White	4.64 (1.24)	5.04 (0.77)
Experiment 2—guilt		
Black	3.40 (1.64)	2.18 (1.64)
White	2.52 (1.94)	3.22 (1.69)
Experiment 3—implicit attitude		
Black	−20.7 (62.6)	10.7 (72.4)
White	−3.9 (56.7)	5.6 (67.6)

Note. In Experiments 1a, 1b, and 3, race of targets was a within-subjects factor. Implicit attitudes for Experiment 3 are difference scores (mean response facilitation in milliseconds of positive adjectives—mean response facilitation in milliseconds of negative adjectives). Standard deviations are listed in parentheses.

As can be seen from the means, this effect was driven by a decrease in liking for Blacks, not by an increase in liking for Whites. White-exposed participants liked Black faces significantly less than control participants, $F(1,66) = 6.37$, $p = .01$, $\eta_p^2 = .09$, but the two groups did not differ in their liking of White faces, $F < 1$.

As expected, repeated exposure to White faces increased prejudice, as compared to the control condition. However, contrary to a mere exposure explanation, this increased ingroup bias was driven by decreased liking for Blacks, rather than increased liking for the exposed group, Whites.

At first glace, the results in the control condition may seem surprising: Control participants actually liked Black faces more than White faces. However, many theories of modern prejudice (e.g., Dovidio & Gaertner, 1998) hypothesize that most Americans today avoid expressing prejudice when it would be obvious to themselves and others. Recent research comparing various experimental locations and setups indicates that in standard laboratory settings using explicit measures, American participants often express more positive attitudes toward Blacks than White targets (e.g., Evans, Garcia, Garcia, & Baron, 2003; Henderson-King & Nisbett, 1996; Ito, Thompson, & Cacioppo, 2004; Towles-Schwen & Fazio, 2003; Vanman, Paul, Ito, & Miller, 1997). We thus sought to replicate these initial findings with a new subject population, one that might be less concerned with appearing politically correct. With such a population we would be more likely to produce what is traditionally seen as prejudice—more positive attitudes toward Whites than Blacks—as opposed to merely eliminating a bias in favor of Blacks.

Experiment 1b

Method

Participants and design. White undergraduate students from the University of Amsterdam ($N = 65$; 20 males, 45 females) took part in the experiment as part of a course requirement, or for €7 (approximately US$6 at the time the experiment was run). Average age was 21.91 years ($SD = 3.66$). All materials were presented to participants in Dutch. The overall design of the experiment was identical to that of Experiment 1a.

Photographic stimuli. Because Experiment 1b took place in a different country than Experiment 1a, new sets of photographs were normed. Photographs of Black and White males were gathered from various sources and piloted on 9 White Dutch graduate students.[2] (Only male photographs were used due to the difficulty of finding stimuli.) These students rated the photographs on 10-point scales on a series of questions, including two target questions on the person's attractiveness and his race/ethnicity. Two sets of four photographs of Black males (Sets 1 and 2) and two sets of four photographs of White males (Sets 1 and 2) were selected. Average ratings of the attractiveness of the two sets did not differ reliably, Fs < 1. Across sets, Black and White faces were roughly equivalently attractive, $t(8) = 1.34$, $p = .22$. All photographs were grayscale and approximately 150 pixels wide and 200 pixels tall.

Procedure. The general procedure was the same as in Experiment 1a. However, due to the different computers and photographs used in this experiment, some aspects of the exposure task were changed. In each trial, a fixation cross appeared for 500 ms, followed by a forward mask for 11 ms, then a photograph (White-exposure condition) or a blank screen (control condition) for 11 ms, then a backward mask for 22 ms, and finally a gray rectangle with colored dots on it. The forward mask was a black and white houndstooth pattern of the same size and shape as the gray rectangle. The backward mask was the same black and white houndstooth pattern covered with several overlapping rectangles and circles of various sizes and shapes, each containing different black and white patterns. As before, the gray rectangle stayed on the screen until participants pressed a key. The time between trials varied randomly from 1 to 3.5 s. Participants in the White-exposure condition were exposed 10 times to four different photos

[2] Photographs of Turkish and Moroccan males were also normed and used in the photograph rating task in Experiment 1b, to test if mere exposure effects would extend to other outgroups. The pilot data indicated the Turkish/Moroccan faces were roughly as attractive as the Black faces, and even somewhat more attractive than the White faces. However, across all experimental conditions, the Turkish/Moroccan faces were rated as far less likable than the other faces, suggesting that these photographs were not comparable to the others. Thus, responses to the Turkish/Moroccan faces are not included in the analyses, but are available from the first author.

of White faces. In the control condition, participants were repeatedly exposed to a blank screen.

Results and discussion

Traditional mere exposure effects on liking. As in Experiment 1a, a 2 (Type of Photo: previously exposed vs. same-race new) repeated-measures ANOVA run on liking ratings for White-exposed participants only was not significant, $F < 1$. Again people exposed to faces of a particular race held the same explicit attitude toward these old, familiar faces and new, same-race faces. We collapsed across this factor for further analyses.

Generalized mere exposure effects on liking. A 2 (Race of Rated Photo: Black face vs. White face) × 2 (Exposure Condition: White exposed vs. control) mixed-model ANOVA was run on liking ratings, with type of photo varied within subjects. A main effect of Exposure Condition, $F(1,63) = 6.32$, $p = .01$, $\eta_p^2 = .09$, indicates that in general, White-exposed participants liked all faces less than control participants. However, these effects were moderated by a Photo Race by Condition interaction, $F(1,63) = 6.03$, $p = .02$, $\eta_p^2 = .09$. The means for this interaction are listed in Table 1.

Here control participants liked Black and White faces equally well, $F < 1$. However, participants who had been exposed to White faces liked Black faces significantly less than White faces, $F(1,63) = 7.77$, $p = .007$, $\eta_p^2 = .11$. Exposure to White faces lead to a preference for Whites relative to Blacks that was not present in the control condition.

Again this effect was driven by a decrease in liking for Blacks. As in Experiment 1a, White-exposed participants liked Black faces significantly less than control participants, $F(1,63) = 9.04$, $p = .004$, $\eta_p^2 = .13$. If anything, exposure to White faces tended to make participants also like White faces less relative to controls, though this was a marginal trend, $F(1,63) = 2.50$, $p = .12$, $\eta_p^2 = .04$.

Experiment 1b replicated the main findings of Experiment 1a: no within-category mere exposure effects were found, and repeated exposure to White faces reduced liking of Blacks relative to Whites. The major difference between these two experiments was the baseline rating of the photographs. In the present experiment, participants generally liked the Black and White faces to the same degree.

In these first two experiments, we found that repeated exposure to White faces made White participants express more prejudice, as measured by liking for Blacks relative to Whites, compared to a baseline control condition. This result fits our hypotheses, except that this effect was driven by a decrease in liking for Blacks. A mere exposure explanation would predict that White-exposed participants would have shown an increase in liking for Whites instead.

We ran two additional studies to test two potential explanations of these effects, mood and racial identity, that we will describe briefly.[3] Since repeated exposure to any stimulus seems to generate diffuse positivity that elevates mood (Monahan et al., 2000), exposure to White faces might elevate mood. Positive mood often leads to more negative judgments of outgroup members (e.g., Bodenhausen, Kramer, & Suesser, 1994), so this improvement in mood could explain why exposing White participants to White faces led to them liking Blacks less. However, when participants completed a standard mood measure (the Positive Affect Negative Affect Scale or PANAS: Watson, Clark, & Tellegen, 1988) after the exposure task, White-exposed and control participants did not differ, Fs < 1.

Another interpretation of our White-exposure condition is that it served as a prime of White identity. General ingroup verbal labels, such as "we" or "us," have been shown to prime social identity in general (Brewer & Gardner, 1996). It would make sense, then, that exemplars of a specific ingroup would prime that specific ingroup identity. Social identity and self-categorization theory would then predict that this heightened ingroup identification would lead to greater ingroup bias, as we found with White-exposed participants (Brewer, 1997). However, when participants completed a racial identification IAT immediately after the exposure task to assess their implicit identification with White versus Blacks (Knowles & Peng, 2005), White-exposed and control participants identified equally highly with Whites, $F < 1$. They also did not respond differently to an explicit measure of White identification, an adaptation of the Inclusion of Other in the Self Scale (Aron, Aron, & Smollan, 1992), $F(1,48) = 1.59$, $p = .21$, $\eta_p^2 = .03$.

Thus, it appears that neither mood nor racial identity priming can explain our effects. However, methodological aspects of our experiments may have contributed to these results. Up to this point, all participants in our experiments have indicated their liking for both Black and White faces. Such a within-subjects measure increases statistical power, but it may have alerted participants to the real purpose of these studies. Once participants saw they were ratings faces of different races, they may have guessed that the experiment was about prejudice. They may then have altered their responses accordingly to ensure they appeared non-prejudiced. Furthermore, our effects may be dependent on having a salient racial context. To make the element of racial bias less obvious, in the next two experiments we had each participant only rate faces of one race. If we continue to find that exposure to Whites faces produces greater prejudice, even in a fully between-subject design, this will testify to the robustness of our effect.

We made one additional change to our methods for the remaining experiments. We used only new, nonexposed faces for the dependent measures since the first two experiments also already confirmed our prediction of no traditional exposure effects. We were also concerned that having both old and new faces in the dependent measure contributed to our effects somehow. The presence of old, familiar White faces may have made the category Whites seem particularly positive, and Blacks much less so in comparison, for White-exposed participants.

[3] The details of these two experiments are available upon request from the first author.

Experiment 2

This experiment investigated whether exposing participants to Whites would still lead to greater prejudice, and particularly make attitudes toward Black more negative, with a fully between participants design. That is, half of our participants expressed their attitude toward a White person and the other half expressed their attitude toward a Black person.

Additionally, in contrast to the previous experiments, which focused on liking judgments, here we employed a new measure of prejudice, judgments of guilt in court cases in a jury decision-making task (borrowed from Bodenhausen et al., 1994). We assumed that if participants viewed a particular race more negatively, they would feel that members of that race were more likely to be guilty of a crime. As a between-subjects design was used, participants rated only one court case, which involved either a White or a Black defendant.

Method

Participants and design. White undergraduate students from the University of Amsterdam ($N = 59$; 12 males, 47 females) took part in the experiment as part of a course requirement, or for €7 (approximately US$9 at the time the experiment was run). Two participants were excluded from the analyses: one due to a computer error and one due to reading difficulties that prevented her from completing the experiment. Thus, 57 participants (12 males, 45 females) were included in the final analyses. Average age was 21.65 years ($SD = 2.57$). All materials were presented to participants in Dutch. This experiment was a 2 (Race of Defendant: Black vs. White) × 2 (Type of Crime: breaking and entering vs. robbery) × 2 (Exposure Condition: White-exposed vs. control) fully between-subjects design.

Procedure. The setup and cover story were the same as in Experiment 1b. A few important procedural changes were made to the photographic stimuli and exposure task. The same sets of photographs were used in the exposure and rating tasks as in Experiment 1b, but no previously shown faces were used in the photographic rating task. Because Experiments 1b did not find differences between Sets 1 and 2 of the White and Black faces, the use of the sets was not counterbalanced. Instead, only Set 1 of White faces was used in the exposure task for White-exposed participants.

Next participants completed a jury decision-making task. One court case was presented to each participant, either a breaking and entering or a robbery (see Appendix). Half of the participants read a case involving a White Dutch male, and the other half read a case involving an African male. Race was manipulated via the name of the defendant (African (Mbamba) or White Dutch (Henk)). After reading the case, participants answered one question on a 9-point scale: "how likely is it that the person is guilty" (1 = *not at all likely*, 9 = *highly likely*). The time participants spent reading the case was also measured.

Results and discussion

Generalized mere exposure effects on guilt. A 2 (Race of Defendant: Black vs. White) × 2 (Type of Crime: breaking and entering vs. robbery) × 2 (Exposure Condition: White-exposed vs. control) between-subjects ANOVA was run on ratings of how likely the defendant was guilty. A significant Defendant Race by Condition interaction emerged, $F(1,49) = 4.34$, $p = .04$, $\eta_p^2 = .08$. The means for this interaction are listed in Table 1.

Control participants tended to think that the Black defendant was less likely to be guilty than the White defendant, $F(1,49) = 3.02$, $p = .09$, $\eta_p^2 = .06$. However, guilt judgments made by White-exposed participants were not affected by the race of the defendant, $F(1,49) = 1.59$, $p = .21$, $\eta_p^2 = .03$. Control participants thought the defendant was more likely to be guilty when he was White than when he was Black, but this tendency was eliminated and even slightly reversed in participants exposed to White faces. This effect was driven by differences in rated guilt of the Black defendant: White-exposed participants thought he was more guilty than control participants, $F(1,49) = 3.89$, $p = .05$, $\eta_p^2 = .07$. The conditions did not differ in their guilt ratings of the White defendant, $F = 1$.

Exposure effects on reading time. If exposing participants to White faces makes their attitudes toward Blacks more negative, then these participants should have been predisposed to view a Black defendant more negatively. White-exposed participants then may not have read the case as carefully, or thought about it as thoroughly, when the defendant was Black because of these negative expectations. To test this possibility, we looked at reading time (i.e., the number of seconds between when the case appeared on the screen and when the participant pressed a key to answer questions about the defendant). There was indeed a significant Exposure Condition by Defendant Race interaction, $F(1,49) = 4.64$, $p = .04$, $\eta_p^2 = .09$. However, it was the control participants who spent more time reading the case when the defendant was White ($M = 35.02$ s, $SD = 11.86$) than when he was Black ($M = 27.86$ s, $SD = 5.69$), $F(1,49) = 3.66$, $p = .06$, $\eta_p^2 = .07$. White-exposed participants spent the same amount of time reading the case with a Black defendant ($M = 35.15$ s, $SD = 9.67$) versus a White defendant ($M = 32.01$ s, $SD = 16.43$), $F(1,49) = 1.45$, $p = .23$, $\eta_p^2 = .03$. Across conditions, reading time was uncorrelated with guilt ratings, $r(55) = .13$, $p = .33$, and covarying reading time did not significantly reduce the previously reported effects. The greater guilt White-exposed participants ascribed to the Black defendant does not appear to be due to cursory consideration of the case.

Experiment 2 replicated the results of the previous experiment with a new measure of prejudice, a jury decision-making task than involved verbal, rather than visual,

stimuli. Here White-exposed participants thought a Black defendant was more likely to be guilty than a White defendant, whereas the reverse was true for control participants. Again the effects were focused on the Black target: White-exposed participants thought he was more likely to be guilty than control participants.

Experiment 3

Will exposure to White faces affect racial attitudes even if those attitudes are assessed implicitly? Though implicitly measured attitudes do not necessarily represent "truer" or more "real" attitudes than explicitly measured ones, they are indeed less suspect to demand characteristics and social desirability effects (Nosek, 2005). They allow one to measure prejudice more subtly. White exposure may have led to greater prejudice in our previous experiments by somehow lifting the norms against explicit expressions of prejudice. That is, being exposed to Whites may have made our participants feel somehow more justified in expressing more negative attitudes toward Blacks. Our effects would then be based on changes in social desirability or self-presentational concerns, rather than actual changes in attitudes. Using an implicit measure of racial attitudes allows us to rule out this possibility.

An affective priming task (Fazio et al., 1995) was used as an implicit measure of attitudes toward Whites and Blacks. In contrast to the IAT, normal samples often do not show a bias against Blacks with the affective priming task (e.g., Fazio & Dunton, 1997; Fazio & Hilden, 2001; Towles-Schwen & Fazio, 2003).

Method

Participants and design. Eighty-nine White undergraduate students from New York University took part in the experiment as part of a course requirement. Ten participants were not native speakers of English, so they were dropped from further analyses. Four additional participants were excluded for suspicion about the purpose of the experiment (e.g., that they had been "primed"). Thus, 75 participants (24 males, 51 females) were included in the final analyses. Average age was 19.57 years ($SD = 1.50$). The overall design of the experiment was a 2 (Race of Prime Photo: Black face vs. White face) × 2 (Valence of Target Word: positive vs. negative) × 2 (Exposure Condition: White exposed vs. control) mixed-model factorial, with the last factor between subjects.

Procedure. The setup, cover story, and exposure task were the same as in Experiment 1a. Next participants completed a computerized affective priming procedure (Fazio et al., 1995) to measure their implicit attitudes toward Blacks and Whites. Participants were told the task examined the degree to which the judgment of word meaning is an automatic skill. The experimenter explained that the task had two parts. In the first part, participants would see a series of words and indicate whether each one was a good or a bad word (i.e., whether it was positive or negative in meaning) by pressing a key labeled "G" (the *K* key) or a key labeled "B" (the *D* key). In the second part, participants would again decide if words were good or bad, but this time a picture would be flashed before each word. Participants were told to pay special attention to each picture because a recognition test would follow later in the experiment.

In the first part of the task, a trial began with fixation crosses in the center of the screen for 500 ms. Then a word appeared and remained on the screen until the participant pressed a key. After 2.5 s, the next trial began. Participants first completed 8 practice trials. In the real trials, participants completed 2 blocks of 24 words each, 12 positive (*appealing*, *delightful*, *favorable*, *likable*, *pleasant*, *wonderful*, *fascinating*, *attractive*, *magnificent*, *satisfying*, *enjoyable*, *beautiful*) and 12 negative (*awful*, *horrible*, *offensive*, *repulsive*, *frightful*, *sickening*, *terrible*, *hideous*, *disgusting*, *painful*, *miserable*, *annoying*). Thus, two response times were obtained for each word presented alone to serve as baseline latencies.

In the second part of the task, participants again responded to the words, but a photograph flashed briefly before each word. In each trial, a photograph appeared for 315 ms, followed by a blank screen for 135 ms, followed by the target adjective. The adjective remained on the screen until the participant pressed a key, or until 1750 ms passed. After 2.5 ms, the next trial began. Participants first completed 8 practice trials. In the actual task, 48 photographs were shown: 6 male and 6 female photographs of White, Black, Hispanic, and Asian faces. The White and Black photographs were the same as in Experiment 1a. The Asian and Hispanic faces were fillers, included to decrease the likelihood of participants guessing our interest in their attitudes toward Blacks and Whites (e.g., Fazio et al., 1995), and thus were not normed. There were 4 blocks, with 48 trials in each block. In each block each photograph was shown once, so that across all 4 blocks, each photo was paired with two positive and two negative adjectives.

Finally, participants were taken through a funnel debriefing procedure (Bargh & Chartrand, 2000) to probe for suspicion and for perception of the subliminal photographs. Then they were debriefed and thanked for their participation.

Results and discussion

The data were analyzed following the technique of Livingston and Brewer (2002). First, RTs for incorrect responses and RTs below 200 ms in either part of the implicit attitude task were deleted, eliminating 3.09% of the trials. There were no significant differences between experimental conditions in terms of incorrect responses or outliers. Then the raw response time for each target word when it was preceded by a face was subtracted by the baseline RT for that word (as measured in the first part of the implicit attitude task) to arrive at a facilitation score.

Positive values indicate that prime faces facilitated the target words (compared to baseline), and negative values indicate inhibition. There was no evidence of skewness in the distribution of these difference scores.

Generalized mere exposure effects on implicit attitudes. A 2 (Race of Prime Photo: Black face vs. White face) × 2 (Valence of Target Word: positive vs. negative) × 2 (Exposure Condition: White exposed vs. control) mixed-model ANOVA was run on these facilitation scores, with race of photo and target word valence varied within subjects. The only effect that emerged was a marginal three-way Photo Race by Valence by Condition interaction, $F(1,73) = 3.67$, $p = .06$, $\eta_p^2 = .05$.

As the previous experiments did not measure positive and negative components of participants' attitudes separately, we felt it would be inappropriate to look separately at how each type of face facilitated responses to positive versus negative words in this experiment. To allow for easier comparison of effects across experiments, we thus created a single implicit Black attitude index and a single implicit White attitude index for each participant. We also felt comfortable doing this because no lower-order effects involving valence were significant, $ps > .14$. The single implicit Black attitude index was calculated by subtracting the mean facilitation score for negative words paired with Black faces from the mean facilitation score for positive words paired with Black faces. The single implicit White attitude index was calculated similarly by subtracting the mean facilitation score for negative words paired with White faces from the mean facilitation score for positive words paired with White faces.[4] These implicit attitude indices are listed for each exposure condition in Table 1.

Did different exposure conditions change participants' liking of Blacks relative to Whites? To answer this question, single-factor (Race of Prime Photo) ANOVAs were run on the implicit attitude indices separately for control and White-exposed participants. Control participants showed the same implicit attitude toward Black faces as toward White faces, $F < 1$. However, participants who had been exposed to White faces had more negative attitudes toward Black faces than toward White faces, $F(1,37) = 4.53$, $p = .04$, $\eta_p^2 = .11$. Exposure to White faces lead to an implicit preference for Whites relative to Blacks that was not present in the control condition.

Implicit attitudes toward Black versus White faces were also examined separately. Replicating the pattern found in the previous experiments, exposure condition significantly affected participants' implicit attitudes toward Blacks, $F(1,73) = 4.03$, $p < .05$, $\eta_p^2 = .05$, but did not influence participants' implicit attitudes toward Whites, $F < 1$. As compared to the no exposure group, exposure to White faces made participants' implicit attitudes toward Blacks more negative, but did not alter implicit attitudes toward Whites.

Because the overall three-way interaction was only marginally significant, we must interpret the results of Experiment 3 with some caution. Exposure to White faces may have weaker effects on implicit attitudes than on explicit attitudes. More research using a variety of implicit measures needs to be done before solid conclusions may be drawn. However, we believe it is notable that this experiment showed the same specific pattern of results as the previous experiments while employing an implicit attitude measure. Again, exposure to White faces led to significantly more negative attitudes toward Blacks relative to a control group, but did not affect attitudes toward Whites.

Across our first 3 experiments, exposure to White faces led to participants expressing more explicit prejudice relative to a control group. In particular, exposure led to more negative explicit attitudes toward Blacks, rather than more positive attitudes toward Whites. A fourth experiment found weaker but similar effects on implicit attitudes. Here the overall pattern did not reach standard levels of significance, but the simple effects did: Exposure to White faces made implicit attitudes toward Blacks significantly more negative without affecting implicit White attitudes. In short, so far we have found extensive evidence supporting our hypothesis that exposure to White faces leads to greater prejudice among White individuals. However, the particular pattern of this increased prejudice was unexpected. Why would exposure to White faces affect ratings of Blacks rather than Whites?

We propose that exposure to White faces did lead to more positive attitudes toward Whites, but due to differences in malleability between attitudes toward Whites and Blacks, this effect emerged as less positive attitudes toward Blacks. Whites and Blacks are often compared. In fact, this idea that "many socially significant categories form complementary pairs" (Greenwald & Farnham, 2000, p. 1023), partially drove the development of one of the most popular modern prejudice measures, the IAT.

Judging a target in comparison to another stimulus involves a different process than judging a target on its own. When a target stimulus is judged by comparing it to a second stimulus, a contrast effect generally results (see Stapel & Koomen, 2001, for a recent review). When the second stimulus is used as a comparison standard, judgments of a target shift or contrast away from that second stimulus. We propose such a mechanism is operating in our experiments: When our participants rated Blacks, they rated them in comparison to Whites. Because Whites and Blacks are so commonly viewed as a "complementary pair," we think such a comparison process occurs even when Blacks are viewed in isolation. That is, even when our participants only responded to Black stimuli (as in

[4] Such indexes of implicit attitudes are often calculated to allow for ease of comprehension and to focus on overall patterns of effects rather than single facilitation scores (e.g., Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995; Livingston & Brewer, 2002; Wittenbrink, Judd, & Park, 1997). In our case, where we did not have specific predictions as to whether our effects were driven by increasing negativity or decreasing positivity toward Blacks, such indices seem particularly appropriate. It should also be noted that when the White-exposed and control conditions were examined separately, neither one showed a main effect of valence, $ps > .17$.

Experiment 2), they still rated Blacks in comparison to Whites. Thus, if Whites were viewed more positively after mere exposure to Whites, as we predicted, this higher comparison standard would have led Blacks to be judged more negatively as a result.

However, this explanation implies that we should have found more positive attitudes toward Whites and more negative attitudes toward Blacks. Our data only provide evidence for the latter. Why did attitudes toward Whites not change? Being White is often experienced as the "norm" or cognitive default (Knowles & Peng, 2005; Smith & Zárate, 1992; Stroessner, 1996; Zárate & Smith, 1990), and for our participants White was also their ingroup. Given that Whites are both the norm and the ingroup, attitudes toward Whites should be higher in ego involvement and commitment and be based on a broader and more extensive array of knowledge—in short, they should be very strong (Eagly & Chaiken, 1998). Strong attitudes are more resistant to change (Eagly & Chaiken, 1995). Hence, we would expect that for White individuals, attitudes toward Blacks would be easier to move around than attitudes toward Whites. Indeed, across all four experiments, attitudes toward Whites never significantly differed between exposure conditions, whereas attitudes toward Blacks always did. The greater stability of attitudes toward Whites means that the positivity engendered by mere exposure could only cause a slight, nonsignificant change in these attitudes, and that this positivity could also lead, via contrast, to a larger, significant change in attitudes toward Blacks.

To test this strength explanation, in Experiment 4 we included a premeasure of the strength of participants' attitudes toward Whites. If we are correct, White-exposed participants should again show greater prejudice relative to control participants, but the specific pattern of this effect should vary depending on the strength of participants' attitude toward Whites. White-exposed participants whose White attitudes are strong should show more negative attitudes toward Blacks relative to control participants, as their White attitudes are too strong to be significantly altered. In contrast, White-exposed participants whose White attitudes are weaker should not show this effect. If anything, because their White attitudes are more malleable, they should report more positive attitudes toward Whites relative to control participants.

Experiment 4

This experiment investigated whether White attitude strength moderated the effect of exposing participants to Whites on attitudes toward Whites and Blacks. As attitude strength is generally considered to be multi-dimensional (e.g., Bassili, 1996; Krosnick, Boninger, Chuang, Berent, & Carnot, 1993; Pomerantz, Chaiken, & Tordesillas, 1995), it is important to use multiple measures of this construct. Evidence for the dimensions of centrality and commitment is fairly strong (e.g., Holland, Verplanken, & van Knippenberg, 2003; Pomerantz et al., 1995), so we chose to use these two dimensions to operationalize White attitude strength in the present experiment. Centrality refers to how much a given attitude is rooted in people's sense of identity, whereas commitment refers to the clarity of an attitude, such as how certain an attitude is and how likely it is to change.

It is important to note that we measured two components of attitude strength only to tap the construct thoroughly. We did not have any *a priori* expectations regarding how one component would behave relative to the other. In short, we viewed certainty and commitment as two measures of the same latent construct.

Experiment 4 was also designed to directly assess attitude change by measuring participants' attitudes toward Whites and Blacks both before and after the exposure task. Our predictions have always explicitly been about how exposure to White attitudes changes attitudes, but our previous experiments did not truly measure attitude change. Instead, they only assessed attitudes at one point in time. Experiment 4 allows for a true assessment of attitude change by including both pre and post attitude measures.

Furthermore, our strength hypothesis predicts different directions of effects for participants low versus high in White attitude strength. To properly assess these effects, we have to adjust for any initial differences in attitudes toward Whites or Blacks between these groups. Measuring these attitudes at two points in time allows us to account for these differences.

Method

Participants and design. Eighty-four White undergraduate students from the University of Amsterdam took part in the experiment as part of a course requirement or for €7 (approximately US$7 at that time). Six participants were excluded from analyses: three due to a computer error, two for not taking the experiment seriously (e.g., they finished the experiment much more quickly than average), and one for suspicion about the purpose of the experiment (e.g., that they had been "primed"). Thus, 78 participants (28 males, 50 females) were included in the final analyses. Average age was 21.71 years ($SD = 3.98$). All materials were presented to participants in Dutch.

Only two experimental conditions were used: White-exposed and control. Furthermore, none of the photographs used in the exposure task were used in the rating task, and each participant rated faces of only one race. Thus, this experiment was a 2 (Race of Rated Photo: Black face vs. White face) × 2 (Exposure Condition: White-exposed vs. control) between-subjects design, with White attitude strength included as a continuous factor.

Procedure. At the start of the experimental session, participants completed the attitude strength measures. They were told that the researchers were interested not only in people's attitudes toward a variety of social groups, but also in how strong and important those attitudes are for

different people. It was emphasized that it was important for participants to answer these items honestly. Participants first indicated their attitude toward 11 different social groups, each on an 11-point scale ($-5 = very\ negative$, $+5 = very\ positive$). The critical groups were Whites and Blacks, but a variety of other groups (e.g., politicians, Spaniards, women) were included as fillers. The groups were presented in a fixed order. Next, participants answered several questions to measure the strength of their attitudes toward Whites, as well as a number of filler groups. They answered two questions to tap attitude commitment: "How certain are you about your attitude toward X?" ($1 = very\ uncertain$, $7 = very\ certain$) and "How likely is it that you will change your attitude toward X in the near future?" ($1 = very\ unlikely$, $7 = very\ likely$). Finally, they answered four questions to measure attitude certainty: "How important is the issue of X to you personally?" ($1 = very\ unimportant$, $7 = very\ important$), "To what extent is your attitude toward X related to your self-concept?" ($1 = not\ at\ all$, $7 = very\ much$), "To what extent does your attitude toward X provide a good description of yourself?" ($1 = not\ at\ all$, $7 = very\ much$), and "To what extent is your attitude toward X related to your important values?" ($1 = not\ at\ all$, $7 = very\ much$). All items were taken from Pomerantz et al. (1995).

After doing 40 min of unrelated tasks, participants completed the exposure and photograph ratings tasks. The setup and cover story were the same as in Experiment 1b. The same sets of photographs were used in these tasks as in Experiment 1a, but no previously shown faces were used in the rating task. Set 1 of White faces was used in the exposure task for White-exposed participants, and Set 2 of White faces and Set 1 of Black faces were used for the rating task. In the rating task, first participants rated how much they liked each of the four people ($0 = not\ at\ all$, $8 = very\ much$). Then they rated how positive or negative their impression was of each of the four people ($0 = very\ negative$, $8 = very\ positive$).

Results and discussion

Attitude strength measures. The White attitude commitment items were not reliable as a separate measure ($\alpha = .45$), so the commitment and centrality items were averaged together and analyzed as a single attitude strength measure ($\alpha = .78$).

Generalized mere exposure effects on liking and impression judgments. Liking and impression ratings were highly correlated, $r(76) = .78$, $p < .001$, so they were averaged together to create a single post-exposure-task attitude measure. Since attitudes were measured on different scales at the beginning of the session (i.e., pre) than after the exposure task (i.e., post), the pre- and post-attitude measures were next standardized. Then pre-attitudes were subtracted from post-attitudes for separate attitude change indexes for Blacks and Whites.

A 2 (Race: Black vs. White) × 2 (Exposure Condition: White exposed vs. control) mixed-model ANOVA was run on these attitude change scores, with race varied within subjects, and with White attitude strength included as a continuous factor. There was a significant main effect of race, indicating that participants' attitudes toward Blacks generally became more negative after the exposure task than their attitudes toward Whites, $F(1,74) = 5.15$, $p = .03$, $\eta_p^2 = .07$. There was also a significant Race by White attitude strength interaction, $F(1,74) = 5.14$, $p = .03$, $\eta_p^2 = .06$. However, as predicted, this was moderated by a three-way interaction of Race, Condition, and White attitude strength, $F(1,74) = 3.98$, $p < .05$, $\eta_p^2 = .05$.

To explore this interaction, we categorized participants as high versus low in White attitude strength based on a median split, and then ran separate 2 (Race: Black vs. White) × 2 (Exposure Condition: White exposed vs. control) mixed-model ANOVAs on attitude change scores for the two groups. Looking first at participants high in White attitude strength, the Race by Exposure Condition interaction was highly significant, $F(1,37) = 29.21$, $p < .001$, $\eta_p^2 = .44$. The pattern of means here is exactly the same as in the previous experiments, as can be seen in the bottom half of Table 2. When White attitude strength was high, exposure to White faces made attitudes toward Blacks more negative, $F(1,37) = 4.40$, $p = .04$, $\eta_p^2 = .11$, but it had no effect on attitudes toward Whites, $F < 1$.

The Race by Exposure Condition interaction was marginally significant for participants low in White attitude strength, $F(1,37) = 3.79$, $p = .06$, $\eta_p^2 = .09$. The pattern of means here is the opposite of the pattern for the high-strength participants, as can be seen in the top half of Table 2. When White attitude strength was low, exposure to White faces had no effect on attitudes toward Blacks, $F = 0$. It did tend to improve attitudes toward Whites, though this effect was not significant, $F(1,37) = 1.63$, $p = .21$, $\eta_p^2 = .04$.

Thus, it appears that our previous finding that exposure to White faces mainly affected attitudes toward Blacks may be explained by the moderating influence of White attitude strength. When participants had relatively strong attitudes toward Whites, we replicated this pattern, found in the previous four experiments. When participants had relatively

Table 2
Attitude change by race of target, white attitude strength, and exposure condition, Experiment 4

Race of target	Exposure condition	
	White-exposed	Control
Low white attitude strength		
Black	−0.09	−0.11
White	0.32	−0.26
High white attitude strength		
Black	−0.40	0.49
White	0.02	−0.17

weak attitudes toward Whites, however, a different pattern emerged. Here exposure to Whites had no effect on attitude change.

Our predictions for strong-White-attitude participants were fully supported, but our predictions for weak-White-attitude participants were not. Participants with relatively weak attitudes did not show improved attitudes toward Whites when they were exposed to Whites. The means are in the appropriate direction, but they are not significantly different.

General discussion

Across our first three experiments we demonstrated that repeated subliminal exposure to White faces made White participants express greater racial prejudice (as evidenced in their attitudes toward Blacks relative to Whites). Exposure to White faces was expected to produce greater prejudice based on recent work by Monahan and colleagues (2000), where repeated exposure to exemplars of one category led to more positive attitudes toward all exemplars of that category, compared to another, unexposed category. In short, we predicted that exposure to White faces would exacerbate racial prejudice by further improving attitudes toward Whites. However, the increase in prejudice that we repeatedly found actually resulted from more negative attitudes toward Blacks. We changed various aspects of the way we measured prejudice, but none of these eliminated the effects of White exposure. When an implicit measure of racial attitudes was used in Experiment 3, the effects were weaker, but the same pattern of data appeared: exposure to White faces led to more negative implicit attitudes toward Blacks, relative to control participants, but did not affect implicit attitudes toward Whites.

In our final experiment, we tested a possible explanation for the fact that White exposure affected attitudes toward Blacks but not attitudes toward Whites. We proposed that White attitudes were generally too strong among our White participants to be changed easily. Instead, Whites served as a comparison standard for judgments of Blacks, leading to a contrast effect in which Blacks were viewed more negatively. Indeed, in Experiment 4 White attitude strength moderated the Race by Exposure Condition interaction. White-exposed participants with relatively strong White attitudes showed a decrease in liking for Blacks, the same pattern we found in the previous experiments.

Such an effect did not occur for participants with relatively weak White attitudes. In fact, these participants were generally unaffected by White exposure, showing no significant change in their attitudes toward Whites or Blacks. However, we had predicted that weak-attitude participants would show a classic mere-exposure pattern, with more positive White attitudes after White exposure. This pattern is apparent in the means, but statistically is nonsignificant.

In short, we found support for the main component of our strength hypothesis, that exposure to White faces makes attitudes toward Blacks more negative among individuals with relatively strong White attitudes. We did not find support for the corollary, that exposure to White faces should make attitudes toward Whites more positive among individuals with relatively weak White attitudes.

What, then, does our attitude strength effect signify? One possibility is that our hypotheses were both correct, but that we did not have enough participants with truly weak White attitudes to test the latter prediction. Few participants in the sample for Experiment 4 had truly weak White attitudes. White attitude strength scores could range from 0 to 8, and almost 85% of our participants had a 3 or higher, leaving only 12 participants with a lower score. This is not surprising. As previously discussed, we had expected that our participants had moderate to strong attitudes toward Whites, since they were all White themselves. Unfortunately, this also suggests we were statistically underpowered to test our predictions for weak White attitudes.

It is also possible that even with more participants with low White attitude strength, we would still not find support for our prediction regarding weak attitudes. In that case, we see at least two additional explanations for our effects. One is that the strength effect reflects the difference between attitude change and attitude formation processes. Having a strong attitude toward Whites means that one has a fully formed attitude that is difficult to change. Thus, mere exposure to White faces is unlikely to alter that attitude, leading to the effects on Black attitudes that we saw. In contrast, a person with a relatively weak attitude toward Whites has an attitude that is not fully formed and is open to change. Exposing this person to White faces can then affect two components of the attitude: the attitude itself (i.e., how positively or negatively the person views Whites), and also the strength of the attitude.

We have already reviewed how repeated exposure to a stimulus should affect the evaluative component of an attitude via the mere exposure effect (Zajonc, 1968). It may be less clear why exposure to an object would also affect the strength of one's attitude toward that object. A person's attitude toward an object may be spontaneously activated upon perceiving that object (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Such automatic activation of a person's attitudes is the basis for many implicit attitude measures, including the affective priming task (Fazio et al., 1995) used in Experiment 3. If attitudes are automatically activated when an attitude object is presented, then repeated exposure to that object would also lead to repeated activation of the attitudes themselves, increasing the accessibility of these attitudes. Accessibility is a key component of attitude strength (Fazio, 1995), and increasing the accessibility of an attitude increases commitment, another component of attitude strength (Holland et al., 2003). In this way, repeated exposure to an attitude object might increase the strength of a person's preexisting attitude toward that object. If our participants' attitudes toward Whites were spontaneously activated each time they were exposed to a

White face, then after repeated exposures to White faces these preexisting attitudes would have become stronger.

This could explain why participants with relatively weak White attitudes did not show a statistically significant improvement in their White attitudes after repeated exposure to Whites. Repeated exposure to White faces may have strengthened these participants' preexisting attitudes toward Whites, making them more resistant to change. Even though this exposure should have also improved these attitudes, the fact that the original attitudes simultaneously became stronger could have dampened or reduced any mere-exposure-induced attitude change. In short, because of two competing processes—strengthening of old attitudes and formation of a new, more positive attitude—occurring simultaneously and interfering with each other, exposure to White faces could only lead to a minimal improvement in White attitudes for participants with relatively weak White attitudes.

A second possibility invokes the moderating role of social identity processes (Tajfel & Turner, 1986). The strength of a person's White attitude may reflect the strength of their identification with Whites. That is, participants higher in White attitude strength may have also identified more strongly with Whites as an ingroup. Ingroup identification is at best weakly related to ingroup bias under neutral circumstances, but in salient intergroup contexts stronger ingroup identification is associated with greater ingroup bias (e.g., Brown, 2000; McGarty, 2001). It is possible that exposing some participants to White faces primed their racial identity—not in the sense of strengthening it, as was ruled out in the study described after Experiment 1b, but in the sense of simply heightening its accessibility. Thus, the White-exposure condition may have served to make racial identification more salient, thus creating a salient intergroup context. In such a salient intergroup context, participants with a strong White identity (i.e., those with strong White attitudes) would be motivated to make their ingroup more positively distinct from other groups. Participants with a relatively weak White identity (i.e., those with relatively weak White attitudes) would not have such motivation, due to their weaker ties to the ingroup. In short, this social identity approach predicts that exposure to White attitudes would lead to ingroup bias only among strongly identified Whites, as we indeed found. However, the social identity literature cannot clearly explain why this bias took the form of outgroup derogation and not ingroup favoritism in our experiments (Brewer, 1999). To provide a stronger test of this idea, future research should measure explicit social identity directly and also measure attitudes toward members of other racial outgroups.

This set of experiments represents a first stringent test of the application of mere exposure to racial attitudes, and any conclusions must be tentative at best. Future research needs to apply this paradigm to other pairs of ingroups and outgroups, preferably crossing them within the same experiment. However, the present data implies some subtle but chilling effects of the highly racially segregated character of most of the United States and many other White-dominated countries. When White people eat, sleep, work, and shop in predominantly White areas, they may not be merely depriving themselves of diverse viewpoints and experiences. They may actually be inadvertently contributing to their own racial biases.

Appendix

Cases used in jury decision-making task in Experiment 2

Name of defendant: X
City: Rotterdam
Age: 25
Accused of: Breaking and Entering

X is accused of breaking into an apartment and stealing electronics. Neighbors reported that they saw someone of X's height and build break the front window of the apartment and leave with the stereo and television. No eyewitnesses saw the thief's face. X owns a jacket similar to the one worn by the thief. X could not provide proof of his whereabouts at the time of the crime. No clothes fragments that could be linked to X were found at the scene. Police searched X's apartment and did not find the stolen stereo or TV.

Name of defendant: X
City: Rotterdam
Age: 26
Accused of: Robbery

X is accused of robbing a nightshop. According to the manager, X entered the store, approached the cash register, and demanded money. The manager claims X threatened to "hurt him," but the manager never saw a weapon. There were no other witnesses to the crime. The manager did not see the robber clearly, but X fit the manager's general verbal description. A woman who works with X reported that he recently complained about money problems and gambling debts. X's girlfriend claims he was at home at the time of the robbery. No fingerprints were found at the scene that matched those of X.

Note: Each participant received only one of the above cases. The letter X was replaced with the appropriate defendant's name in the experiment.

References

Allport, G. W. (1981/1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596–612.

Ball, P. M., & Cantor, G. N. (1974). White boys' ratings of pictures of whites and blacks as related to amount of familiarization. *Perceptual and Motor Skills, 39*, 883–890.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and*

personality psychology (pp. 253–285). New York, NY: Cambridge University Press.

Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology, 71*, 637–653.

Bodenhausen, G. V., Kramer, G. P., & Suesser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66*, 621–632.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106*, 265–289.

Bornstein, R. F. (1993). Mere exposure effects with outgroup stimuli. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 195–211). San Diego, CA: Academic Press.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology, 63*, 545–552.

Bornstein, R. F., Leone, D. R., & Galley, D. J. (1987). The generalizability of subliminal mere exposure effects: Influence of stimuli perceived without awareness on social behavior. *Journal of Personality and Social Psychology, 53*, 1070–1079.

Brewer, M. B. (1997). The social psychology of intergroup relations: Can research inform practice? *Journal of Social Issues, 53*, 197–211.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429–444.

Brewer, M. B., & Gardner, W. (1996). Who is this "we"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*, 83–93.

Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology, 30*, 745–778.

Cantor, G. N. (1972). Effects of familiarization on children's ratings of pictures of whites and blacks. *Child Development, 43*, 1219–1229.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 3–32). Thousand Oaks, CA: Sage Publications.

Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup contact: The past, present, and the future. *Group Processes and Intergroup Relations, 6*, 5–20.

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. (1995). Attitude strength, attitude structure, and resistance to change. In R. E. Petty & J. A. Krosnick (Eds.). *Attitude strength: Antecedents and consequences. Ohio state university series on attitudes and persuasion* (vol. 4, pp. 413–432). Mahwah, NJ: Lawrence Erlbaum Associates.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function (4th ed.. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (Vol. 2, pp. 269–322). New York: McGraw Hill.

Evans, D. C., Garcia, D. G., Garcia, D. M., & Baron, R. S. (2003). In the privacy of their own homes: Using the internet to assess racial bias. *Personality and Social Psychology Bulletin, 29*, 273–284.

Fazio, R. H. (1995). Attitudes as object–evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33*, 451–470.

Fazio, R. H., & Hilden, L. E. (2001). Emotional reactions to a seemingly prejudiced response: The role of automatically activated racial attitudes and motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 27*, 538–549.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination (4th ed.. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (Vol. 2, pp. 357–411). Boston, MA: McGraw-Hill.

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

Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the "mere exposure" effect. *Journal of Personality and Social Psychology, 45*, 492–500.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Hamm, N. H., Baum, M. R., & Nikels, K. W. (1975). Effects of race and exposure on judgments of interpersonal favorability. *Journal of Experimental Social Psychology, 11*, 14–24.

Henderson-King, E. I., & Nisbett, R. E. (1996). Anti-black prejudice as a function of exposure to the negative behavior of a single black person. *Journal of Personality and Social Psychology, 71*, 654–664.

Holland, R. W., Verplanken, B., & van Knippenberg, A. (2003). From repetition to conviction: Attitude accessibility as a determinant of attitude certainty. *Journal of Experimental Social Psychology, 39*, 594–601.

Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin, 30*, 1267–1280.

Knowles, E. D., & Peng, K. (2005). White selves: Conceptualizing and measuring a dominant-group identity. *Journal of Personality and Social Psychology, 89*, 223–241.

Kramer, R. S. S., & Parkinson, B. (2005). Generalization of mere exposure to faces viewed from different horizontal angles. *Social Cognition, 23*, 125–136.

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology, 65*, 1132–1151.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 557–558.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.

Logan, J. (2001). *Ethnic diversity grows, neighborhood integration lags behind*. Albany, NY: State University of New York at Albany, Lewis Mumford Center.

McGarty, C. (2001). Social identity theory does not maintain that identification produces bias, and self-categorization theory does not maintain that salience is identification: Two comments on Mummendey, Klink and Brown. *British Journal of Social Psychology, 40*, 173–176.

Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science, 11*, 462–466.

Moreland, R. L., & Beach, S. R. (1992). Exposure effects in the classroom: The development of affinity among students. *Journal of Experimental Social Psychology, 28*, 255–276.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565–584.

Perlman, D., & Oskamp, S. (1971). The effects of picture content and exposure frequency on evaluations of Negroes and Whites. *Journal of Experimental Social Psychology, 7*, 503–514.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751–783.

Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology, 69*, 408–419.

Rhodes, G., Halberstadt, J., & Brajkovich, G. (2001). Generalization of mere exposure effects to averaged composite faces. *Social Cognition, 19*, 57–70.

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99*, 3–21.

Smith, T. W. (2002). Measuring inter-racial friendships. *Social Science Research, 31*, 576–593.

Stapel, D. A., & Koomen, W. (2001). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the legacy and future of social cognition* (pp. 229–246). Mahwah, NJ: Lawrence Erlbaum Associates.

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition, 14*, 247–276.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed.). Chicago: Nelson-Hall Publishers.

Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin, 29*, 170–182.

Tropp, L. R., & Pettigrew, T. F. (2005). Differential relationships between intergroup contact and affective and cognitive dimensions of prejudice. *Personality and Social Psychology Bulletin, 31*, 1145–1158.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73*, 941–959.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.

Whittlesea, B. W. A., & Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition, 29*, 234–246.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8*, 161–185.