NOT ALWAYS BLACK AND WHITE: THE EFFECT OF RACE AND EMOTIONAL EXPRESSION ON IMPLICIT ATTITUDES

Jennifer R. Steele, Meghan George, Margaret K. Cease, Tracy L. Fabri, and Jacob Schlosser
York University

Across three studies we examined people's implicit attitudes toward Black and White targets who differed systematically by emotional expression. In Study 1, both the race and the emotional expression of primes affected people's attitudes as measured by the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005). In Study 2, participants completed an Implicit Association Test (IAT; Greenwald, Nosek, & Banaji, 2003) containing smiling Black and neutral White faces. When categorizing by race, participants implicitly preferred neutral White, over smiling Black, faces. By contrast, when categorizing by emotional expression, participants showed an implicit preference for smiling Black faces. In Study 3, participants spontaneously categorized these faces by race or emotional expression. Implicit biases again reflected participants' social categorization, however the majority of non-Black participants spontaneously categorized by race. Taken together, these results suggest that how we categorize multiply categorizable others directly affects our spontaneous affective responses.

Keywords: implicit attitudes, emotion, racial bias, Affect Misattribution Procedure, Implicit Association Test

"The human mind must think with the aid of categories…This principle holds even though we often make mistakes in fitting events to categories and thus get ourselves into trouble."
—Allport (1954, p. 21), *The Nature of Prejudice*

In *The Nature of Prejudice* (1954), Gordon Allport made the case that people have a mental need to think categorically about objects, animals, and perhaps most consequentially, people. Support for his basic idea that categorization, an inevitable and necessary cognitive process, can lead us to make mistakes and "thus get ourselves into trouble" abound both in the social psychological literature and in the daily news. While evidence of intentional categorization and overt discrimination are often found throughout society, prejudice can also take much more subtle and even unintentional forms.

Over the last two decades, researchers examining racial prejudice have often aimed to better understand people's implicit attitudes, which are "unintentional, resource-independent, unconscious, or uncontrollable" evaluative responses (Gawronski & De Houwer, 2014, p. 284). One robust finding to emerge from this literature is that non-Black adults typically show a preference favoring White, as opposed to Black, racial group members, despite generally expressing egalitarian views (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Dovidio, Gaertner, Kawakami, & Hodson, 2002; Greenwald, Banaji, & Nosek, 2015; Lai et al., 2014). This implicit pro-White (versus Black) bias has been found at adult-like levels in White American children as young as 6 years of age (Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2007; Williams & Steele, 2017). Among adults, implicit racial preferences are a better predictor than explicit attitudes of negative nonverbal behavior during interracial interactions (Dovidio, Kawakami, & Gaertner, 2002; Greenwald, Poehlman, Uhlmann, & Banaji, 2009), as well as increased amygdala activation and startle eye-blink responses when viewing Black versus White faces (Cunningham et al., 2004; Phelps et al., 2000). Taken together, these findings suggest that there are both theoretical and practical benefits to better understanding the conditions under which implicit racial biases are attenuated or even reversed.

A substantial body of research examining implicit racial biases has focused on attitudes toward Black and White targets who differ only by race. However, people who we encounter in everyday life are multifaceted, with race being only one of many dimensions along which they might be categorized. How might other perceptually salient social categories and attributes affect people's implicit attitudes toward Black and White targets? The goal of the present research was to address this question. Across three studies we examined the effect that emotional expression, a dynamic social category (Freeman & Ambady, 2014), can have on the attitudes that are spontaneously evoked by Black and White target faces. Specifically, we examined people's implicit attitudes toward smiling Black and neutral White faces, with the goal of better understanding the conditions under which implicit racial biases are expressed.

MULTIPLE SOCIAL CATEGORIES

There is now ample evidence from the impression formation literature to suggest that people's multiple identities, and the stereotypes associated with these identities, can influence the impressions of others that we form. Both theory and re-

search suggest that when we are faced with a myriad of social categories and attributes, some with competing associations, categorization can play a central role in informing our impressions. How a person is categorized, be it by race, gender, age, or emotional expression, will affect which stereotypes are subsequently activated (Brewer, 1988; Fiske & Neuberg, 1990; Freeman & Ambady, 2014; Kunda & Thagard, 1996; Pittinsky, Shih, & Ambady, 1999). This can, in turn, have downstream consequences for impression formation.

The effect of categorization on stereotype activation has been demonstrated in a series of studies in which participants viewed an Asian-American woman (Macrae, Bodenhausen, & Milne, 1995). In one of these studies, participants were first primed with either Asian- or female-related words. They were then asked to view a brief video of an Asian-American woman, and subsequently completed a seemingly unrelated lexical decision task. Participants who had first been primed with Asian-related words showed an increased activation of Asian stereotypes, in the form of faster identification of Chinese-stereotypical words, and an inhibition of female stereotypes, in the form of slower identification of female-stereotypical words, on the lexical decision task. Participants who had first been primed with female-related words showed the reverse pattern of results. The activation and inhibition of competing social identities depended on which of these two identities— "female" or "Asian"—was made contextually salient prior to viewing the video (Macrae et al., 1995; cf. Gilbert & Hixon, 1991). Once categorized by race or gender, the relevant stereotypes were more accessible to the perceiver, while stereotypes associated with the other identity were inhibited. This finding provides evidence that the stereotypes activated by an identical target can differ quite dramatically depending on which of her multiple identities are salient and thus used as a basis for categorization during person perception.

In related research by Sinclair and Kunda (1999, Study 3), participants who received positive feedback from a Black doctor subsequently showed activation of doctor stereotypes, and inhibition of Black racial stereotypes on a lexical decision task. By contrast, participants who had been criticized by a Black doctor showed activation of Black racial stereotypes and inhibition of doctor stereotypes. In addition, participants who were evaluated negatively by a Black, as opposed to a White, manager not only showed greater activation of Black racial stereotypes, but also evaluated the Black manager more negatively (Sinclair & Kunda, 1999, Study 1). These differences in evaluations by manager race did not emerge when a Black, versus White, manager evaluated participants positively. Together, these studies provide additional evidence that social categories, including race, gender, and even occupation, do not always result in the spontaneous activation of relevant stereotypes (see also Gilbert & Hixon, 1991). Instead, contextual factors, as well as motivational processes, can affect how we categorize others, the ensuing stereotypes that are activated, and ultimately the impressions of others that we form (Brewer, 1988; Fiske & Neuberg, 1990; Freeman & Ambady, 2014; Kunda & Thagard, 1996; Pittinsky et al., 1999).

Interestingly, some research further suggests the bi-directionality of these processes. Specifically, it has been shown that perceiving stereotype-consistent emo-

tional cues can also affect the racial categorization process, at least for those higher in implicit prejudice (Hugenberg & Bodenhausen, 2004). Across two studies, people high in implicit prejudice were more likely to categorize racially ambiguous targets as Black, as opposed to White, when the targets had an angry facial expression (see also Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011). This finding is consistent with more recent theorizing which suggests a dynamic interactive model of person construal, in which top-down social cognitive processes interact with bottom-up perceptual processes in an interactive manner to influence both categorization and corresponding stereotype activation (Freeman & Ambady, 2011). It seems not only that social categorization (e.g., categorizing a person as Black) can influence stereotype activation (e.g., "aggressive"; Sinclair & Kunda, 1999), but that stereotypes activated by perceptual cues, including emotional expression (e.g., "angry"), can affect social categorization (e.g., categorizing a racially ambiguous person as "Black"; Hugenberg & Bodenhausen, 2004).

SOCIAL CATEGORIZATION AND IMPLICIT ATTITUDES

Although there is a great deal of research demonstrating that cognitively based stereotypes about multiple social categories, in particular gender, race, age, and emotion (Freeman & Ambady, 2014), can affect impression formation and person perception, much less is known about how these multiple social categories affect implicit attitudes. Research examining implicit biases has been integral in enhancing our broad understanding of social cognition as well as more nuanced aspects of racial prejudice. However, the vast majority of this research has focused on attitudes toward targets that differ along a single social category, often race. When we consider that people can be categorized in a multitude of ways, an important question arises: How do these multiple social categories influence the attitudes that targets spontaneously evoke?

There is some initial research to suggest that how we categorize others can influence our spontaneously activated affective responses. Wittenbrink, Judd, and Park (2001) examined people's implicit racial attitudes following exposure to videos of Black men in specific social contexts. Participants first watched video clips of Black targets in a positive (family barbeque) or negative (potentially gang-related) scene. Those who saw the positive clip subsequently demonstrated a significant decrease in implicit pro-White bias relative to baseline, whereas those who saw the negative clip did not (see also Barden, Maddux, Petty, & Brewer, 2004; Dasgupta & Greenwald, 2001; Livingston & Brewer, 2002; Lowery, Hardin, & Sinclair, 2001). Even when categorized by race, bias was attenuated when information provided prior to the implicit measure presented stereotype inconsistent social category information (e.g., family man).

Of greatest relevance to the current research, Mitchell, Nosek, and Banaji (2003, Study 1) assessed people's implicit attitudes by using familiar, liked Black athletes and disliked White politicians as target stimuli in an Implicit Association Test (IAT; Greenwald et al., 2003). These researchers demonstrated that how targets are cat-

egorized can affect implicit attitudes. When asked to categorize the targets based on their occupation, participants showed greater implicit positivity toward liked Black athletes as compared to disliked White politicians. However, when asked to categorize targets based on race, racial biases emerged, with participants showing an implicit bias favoring the disliked White politicians over the liked Black athletes.

In a subsequent study (Study 2), the authors demonstrated that this shift was not exclusively because attitudes toward different categories (politician versus athlete in one condition; Black and White racial groups in other) were being assessed. Participants in this study were asked to select either (a) three liked Black and three disliked White famous Americans, or (b) three disliked Black and three liked White famous Americans. These selections were used as the racial category exemplars in a subsequent IAT. Participants in both conditions then categorized the targets by race while completing a race-attitude IAT. Those who categorized disliked Black and liked White exemplars showed a pro-White bias, whereas those who categorized liked Black and disliked White targets did not, suggesting that both social categories and competing associations with specific known exemplars informed implicit attitudes.

PRESENT RESEARCH

The goal of the present research was to build on these findings by examining how a dynamic attribute, specifically emotional expression, can affect people's implicit attitudes toward Black and White targets. Across three studies, we examined participants' implicit attitudes toward targets who differed both by race (Black or White) and by emotional expression (smiling or neutral). In two of these studies (Studies 2 and 3) we focused exclusively on attitudes toward smiling Black versus neutral White targets.

We chose to examine the impact of emotional expression for several reasons. First, research suggests that emotional expressions can be used as a basis for social categorization in a way that is similar to other stable social categories (Cañadas, Lupiáñez, Kawakami, Niedenthal, & Rodríguez-Bailón, 2016; Freeman & Ambady, 2011). Second, although emotional expressions may be dynamic and fluid, these cues provide critical information when making social inferences, particularly during social interactions (for a review see Keltner & Haidt, 1999). We may be more likely to interpret a smiling face as someone who is willing to interact, and a neutral face as someone who is either unfriendly or not interested in communicating (Kraut & Johnston, 1979). As such, we felt that emotional expressions would provide dynamic social information that participants would spontaneously attend to. Finally, previous research has found that people show implicit biases favoring smiling faces (Donges, Kersting, & Suslow, 2012; Rohr, Degner, & Wentura, 2012), and this is true even in young children (Williams, Steele, & Lipman, 2016). By providing participants with smiling Black and neutral White faces, we could determine whether emotional expressions that have been shown to evoke consistent

and predictable implicit preferences would influence implicit attitudes toward racially prototypical Black and White targets, and whether this would depend on how attitudes were assessed.

In Study 1, we made use of the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) to examine participants' implicit attitudes toward targets that differed by both race (i.e., Black and White targets) and emotional expression (i.e., smiling and neutral expressions), on a measure that does not require categorization. We tested the possibility that both race and emotional expression would have independent effects on people's attitudes, with people showing greater positivity toward both White and smiling faces.

In Study 2, we examined people's implicit attitudes toward smiling Black and neutral White targets using the Implicit Association Test (IAT; Greenwald, Nosek, & Banaji, 2003). We tested the hypothesis that non-Black participants' implicit attitudes would reflect their basis of categorization. Based on the findings of Mitchell and colleagues (2003), we anticipated that participants would show greater implicit positivity toward smiling Black targets, relative to neutral White targets, when categorizing by emotional expression as opposed to race. In Study 3, we aimed to replicate and extend the results of Study 2 by including an ambiguous categorization condition. In this condition, participants were not instructed how to categorize smiling Black and neutral White faces, and were therefore able to attend primarily to race, emotional expression, or both. Across each of these studies, we tested the hypothesis that emotional expressions, which are inherently dynamic, can act as a competing social cue and influence participants' spontaneously activated affective responses relative to a more stable social category, specifically race. For each study, we report all measures, conditions, and data exclusions.

STUDY 1

In Study 1 we examined people's implicit attitudes toward targets who differed by both race (i.e., Black and White) and emotional expression (i.e., smiling and neutral) using an Affect Misattribution Procedure (AMP; Payne et al., 2005). In the AMP, participants view multiple trials in which a prime image is shown briefly (for 75 ms) followed by a blank screen (125 ms), an affect-neutral stimulus (100 ms), and then a mask that remains on the screen until the participant provides a response. The participant's task on each trial is to rate whether the affect-neutral stimulus is relatively more pleasant or unpleasant than average, while disregarding the prime image. The premise underlying this task is that if affect is elicited by the prime, it will be misattributed to the ambiguous stimulus, resulting in a rating that reflects affect elicited by the prime.

Priming measures have been found to be sensitive to perceptual features of evaluative targets (see, for example, Livingston & Brewer, 2002), and the AMP has been shown to have predictive validity and to be a reliable implicit measure (Cameron, Brown-Iannuzzi, & Payne, 2012; Payne & Lundberg, 2014). Results from previous studies have shown that positively versus negatively valenced primes can lead

to priming effects on the AMP in the anticipated direction, with both adults and children rating ambiguous stimuli that followed positive primes (e.g., bunnies, a baby) more favorably than ambiguous stimuli that followed negative primes (e.g., snakes, a spider; Payne et al., 2005; Williams et al., 2016). As such, it was expected that smiling primes would elicit more positivity than neutral primes.

In addition, White participants have been found to show a pro-White (versus Black) bias on the AMP (Payne et al., 2005). We therefore aimed to identify whether racial biases would still emerge when other relevant information (i.e., emotional expression) was available. Previous research using priming measures suggests that some racial targets might not elicit racial biases. For example, Black faces low in racial prototypicality did not consistently elicit racial biases on priming measures (Livingston & Brewer, 2002) nor did prototypical Black faces when the context suggested that they were athletes or lawyers (Barden et al., 2004; see also Remedios, Chasteen, Rule, & Plaks, 2011). In the current study we tested whether smiling Black faces would similarly fail to elicit racial biases on this priming measure, using a 2 (Race of Prime: Black or White) × 2 (Emotion of Prime: Smiling or Neutral) within-subject design.

METHOD

Participants

One hundred and sixty-six non-Black undergraduates participated for course credit. Data from 7 participants were excluded from our analyses because they exhibited patterned or controlled responding ($n = 5$)[1] or told the experimenter explicitly that they had rushed through the study, which was confirmed by their reaction times ($n = 2$). This led to a final sample of 159 participants (105 female, M_{age} = 19 years, SD = 4.36), comprised of South Asian (n = 42 [30 female], M_{age} = 19 years, SD = 5.28), White (n = 41 [26 female], M_{age} = 21 years, SD = 5.33), Middle-Eastern (n = 30 [20 female], M_{age} = 18 years, SD = 1.77), East/Southeast Asian (n = 28 [15 female], M_{age} = 18 years, SD = 1.31), Hispanic (n = 9 [6 female], M_{age} = 18 years, SD = 1.13), and non-Black multiracial/other race (n = 9 [8 female], M_{age} = 22 years, SD = 6.91) participants.[2]

---

1. Patterned responding was evidenced by the fact that participants either used only one of the two keys when responding to the AMP or they used alternating keys. These exclusions were based on an a priori decision to remove participants who identified between 90 and 100% of the inkblots shown after a smiling face as pleasant and 0 to 10% of the inkblots shown after a neutral face as unpleasant. Although such responding may have been the result of priming effects, we removed these participants out of concern that they may have been purposefully controlling their responses. Inclusion or exclusion of these participants did not substantially alter the results.

2. In each of the studies, Black participants who signed up were provided with the opportunity to complete the study, but were excluded from our data prior to any analyses. In Study 1, this included 36 Black participants. For Study 2, this included 10 Black participants as well as 7 participants who did not provide their race. For Study 3, this included 44 Black participants, 7 biracial/multiracial (Black) participants, as well as 6 participants who did not provide their race.

Measures

Implicit Racial Attitudes. Implicit attitudes were measured using the Affect Misattribution Procedure (AMP; Payne et al., 2005). The AMP consisted of two blocks, each containing 40 trials. In each trial, a prime was shown on the screen for 75 ms, followed by a blank screen for 125 ms, a novel inkblot for 100 ms, and finally a mask. The mask remained on the screen until the participant indicated, using one of two designated computer keys, whether the inkblot was more or less pleasant than an average inkblot. Unlike the original AMP, which asked participants to rate Chinese characters, in the present study we used inkblots to reduce the likelihood that participants from our racially diverse sample would recognize the characters being used. A total of 40 novel inkblots were used, and participants were instructed to ignore the faces that flashed before the inkblots, which participants were told were simply a warning that the inkblot was about to be revealed. The primes included 10 smiling Black men, 10 neutral Black men, 10 smiling White men, and 10 neutral White men, matched for age and attractiveness that were shown individually and in random order.

Procedure

Participants were taken to individual testing rooms where they provided written consent and then completed the AMP.[3] Upon completion of the study, participants were debriefed and thanked for their participation.

RESULTS AND DISCUSSION

A 2 (Race of Prime: Black or White) × 2 (Emotion of Prime: Smiling or Neutral) within-subjects ANOVA was conducted using participants' proportion of positive inkblot ratings following each prime as the dependant variable. A main effect of Race of Prime emerged, $F(1, 158) = 4.66$, $p = .03$, $\eta_p^2 = .03$. Participants evaluated inkblots more favorably when they were preceded by a White ($M = .54$, $SD = .13$) as opposed to a Black ($M = .51$, $SD = .14$) prime. A significant main effect of Emotion of Prime also emerged, $F(1, 158) = 42.48$, $p < .001$, $\eta_p^2 = .21$. Participants showed greater positivity toward inkblots that followed a smiling ($M = .59$, $SD = .19$) as opposed to neutral ($M = .45$, $SD = .16$) face. No significant interaction effect emerged, $F(1, 158) = .04$, $p = .84$, $\eta_p^2 < .001$.

As can be seen in Figure 1, participants were more likely than chance to rate the inkblots as pleasant when they followed smiling primes for both White ($M = .60$, $SD = .20$), $t(158) = 6.51$, $p < .001$, 95% CIs [.07, .14], $d = 1.04$, and Black ($M = .58$,

---

3. Participants in each study also completed a series of explicit measures following the completion of the implicit measure, including Feeling Thermometers (Nelson, 2008), the Motivation to Respond Without Prejudice Scale (Plant & Devine, 1998), the Modern Racism Scale (McConahay, Hardee, & Batts, 1981), as well as a Friends Questionnaire. These data were collected for exploratory purposes and results were therefore not analyzed as part of the current studies.
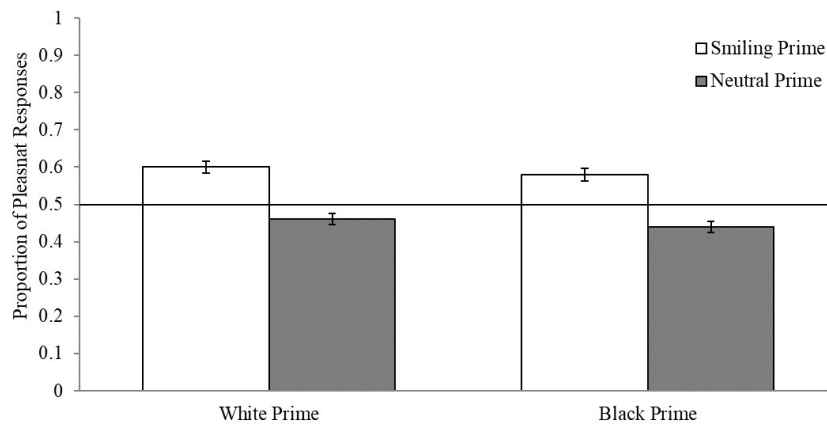
FIGURE 1. Study 1 proportion of AMP trials in which participants indicated that inkblots following smiling and neutral White and Black faces were more pleasant than average. Error bars represent the standard error for each column.

SD = .22), $t(158)$ = 4.67, p < .001, 95% CIs [.05, .12], d = .74, faces. Participants were also more likely than chance to evaluate the inkblots as unpleasant when they followed neutral White face primes (M = .46, SD = .19), $t(158)$ = -2.37, p = .02, d = -.38, or neutral Black face primes (M = .44, SD = .19), $t(158)$ = -4.10, p < .001, d = -.65. These findings suggest that emotion and race each had an independent effect on people's implicit attitudes when people were not required to categorize faces within the task.

In Study 1, participants showed more positive attitudes toward White (versus Black) target faces as well as smiling (versus neutral) targets faces. In the current research both stable (race) and dynamic (emotional expression) aspects of the targets had independent effects on people's implicit attitudes when these targets were viewed as exemplars, however the effects were of different magnitudes. The emotional expression of the prime had a greater impact, with the effect size for emotional expression being seven times larger than the effect of race. It seems possible that this was due largely to the structure of this priming task, a point that we return to in the General Discussion.

In Studies 2 and 3 we aimed to extend this research by examining people's implicit attitudes toward targets whose race and emotional expression typically elicit different affective responses. Specifically, we focused on people's implicit attitudes toward smiling Black and neutral White targets. In addition, across these two studies we examined people's implicit attitudes using the Implicit Association Test (IAT; Greenwald et al., 2003), a task that requires categorization. As such, we were able to test our hypothesis that the basis of categorization (i.e., by race or emotional expression) would affect people's implicit attitudes, with more positive attitudes being elicited by smiling Black targets, relative to neutral White targets, when they

were categorized by emotional expression as opposed to race. Although both the race and the emotional expression of targets had an effect on people's attitudes in Study 1, in Study 2 we examined whether associations with one social category (i.e., race) could be overridden by associations with other aspects of the target (i.e., emotional expression).

STUDY 2

In Study 2 we sought to extend the findings of Study 1 by examining participants' implicit attitudes toward smiling Black and neutral White targets when participants were asked to categorize these targets by race (Race-Categorization condition) or emotional expression (Emotion-Categorization condition). In addition to these two experimental conditions, we included two control conditions—one in which participants were asked to categorize faces that differed only by race (Race-Control condition) and another in which participants were asked to categorize same-race faces that differed only by emotional expression (Emotion-Control condition). This allowed us to determine whether the additional social category (i.e., race or emotional expression) would influence implicit biases. Specifically, when categorizing targets by race, would the magnitude of racial bias be equivalent regardless of whether the faces differed by emotional expression (i.e., smiling Black faces and neutral White faces) or not (all smiling faces)? Similarly, would an implicit preference for smiling faces be attenuated when the target smiling faces were racially prototypical Black faces, as opposed to racially prototypical White faces?[4]

As such, the study had a 2 (Categorization Task: Race or Emotion) × 2 (Condition: Experimental or Control) between-subjects design.

METHOD

Participants

One hundred and sixty non-Black undergraduates participated for course credit.

Data from 12 participants were excluded from our analyses because they failed to follow the instructions ($n = 3$), accurately identified the purpose of the study ($n = 2$), responded to at least 10% of trials faster than 300 ms ($n = 1$; Greenwald et al., 2003), had an error rate greater than 30% ($n = 3$; Banse, Gawronski, Rebetez, Gutt, & Morton, 2010), or had an average reaction time on all trials that fell above three standard deviations above the mean ($n = 3$; Cvencek, Greenwald, & Meltzoff, 2011;

---

4. We selected these control conditions to address our specific research questions. However, another interesting comparison would be with smiling White and neutral Black faces. This would allow us to further determine whether smiling White faces and neutral Black faces show a similar magnitude of bias when they are categorized by either race or emotional expression. As this was not our central question, we did not include these comparisons. This is, however, a question worthy of future investigation.

Cvencek, Meltzoff, & Greenwald, 2011). This led to a final sample of 148 participants (107 female, M_{age} = 19 years, SD = 2.91), which included South Asian (n = 45 [31 female], M_{age} = 19 years, SD = 2.14), White (n = 41 [32 female], M_{age} = 18 years, SD = 1.64), East/Southeast Asian (n = 34 [23 female], M_{age} = 18 years, SD = 1.52), Middle-Eastern (n = 18 [13 female], M_{age} = 18 years, SD = 1.32), and non-Black multiracial/other race (n = 10 [8 female], M_{age} = 21 years, SD = 9.09) participants.

Measures

Implicit Attitudes. Implicit attitudes were measured using an Implicit Association Test (IAT; Greenwald et al., 2003), which is a computer-based reaction time task. In both the Race-Categorization and Emotion-Categorization conditions, target images included faces of smiling Black (n = 4) and neutral White (n = 4) men that were racially prototypical and matched for age and attractiveness. In the Race-Control condition, target images included matched faces of smiling Black (n = 4) and smiling White (n = 4) men, and in the Emotion-Control condition, target images included matched faces of smiling White (n = 4) and neutral White (n = 4) men. Thus target images in the two experimental conditions differed systematically by *both* race and emotional expression. By contrast, target images in the two control conditions differed systematically by either race *or* emotional expression. Across each of the conditions, pleasant and unpleasant attributes were depicted pictorially with valenced stimulus images taken from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008).

Participants completed seven blocks, as outlined by Greenwald et al. (2003). Blocks 1 and 2 were practice blocks consisting of 20 trials each. In Block 1, participants sorted the faces using two computer keys that corresponded to a header that appeared on the screen. In the Race-Categorization and Race-Control conditions, participants were asked to sort sequentially presented faces by race, and were shown a header with a Black cartoon face on one side and a White cartoon face on the other side. Each cartoon face on the header had the facial features removed so that no emotional expression information was available. In the Emotion-Categorization and Emotion-Control conditions, participants were asked to sort the sequentially presented faces by emotional expression and saw a header with a schematic smiling face on one side and a schematic neutral face on the other. In Block 2, participants sorted pleasant (sunflowers, kittens, puppies, beautiful landscape) and unpleasant (garbage, oil spill, demolished house, landfill) stimuli using a header with the words "pleasant" and "unpleasant" as labels.

Blocks 3 (20 practice trials) and 4 (40 test trials) were the first of two sets of critical blocks that combined the faces and valenced images. Participants in the Race-Categorization and Race-Control conditions saw a header containing the cartoon face of one race (e.g., White) and the word "pleasant" on one side, and a cartoon face of the other race (e.g., Black) and the word "unpleasant" on the other side (see Appendix A). Participants in the Emotion-Categorization and Emotion-Control conditions instead saw a header containing a schematic face with one expression (e.g., smiling) and the word "pleasant" on one side, and a schematic face with the
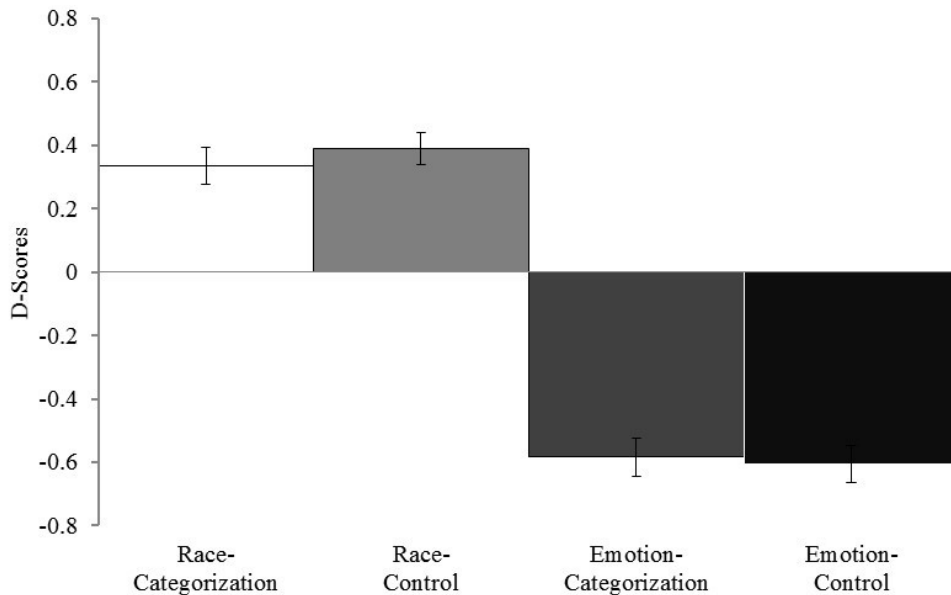
FIGURE 2. Study 2 IAT D-scores by condition. In the two Categorization conditions, higher D-scores represent greater implicit preference for neutral White versus smiling Black targets. In the Race-Control condition, higher D-scores represent greater implicit preference for White versus Black targets and in the Emotion-Control condition, higher D-scores represent greater implicit preference for neutral versus smiling targets. Error bars represent the standard error.

other expression (e.g., neutral) and the word "unpleasant" on the other side. Participants sorted all of the previously viewed images using this header. Block 5 was identical to Block 1, with the exception that the location of the two pictures on the header was reversed. Blocks 6 (20 practice trials) and 7 (40 test trials) were the final sets of critical blocks that combined the faces and valenced images. These were the same as Blocks 3 and 4, but with the new pairings. For example, participants in the Race-Categorization and Race-Control conditions who sorted White faces and pleasant images using one computer key and Black faces and unpleasant images using the other computer key in Blocks 3 and 4, now sorted Black faces and pleasant images using one key and White faces and unpleasant images using the other key in Blocks 6 and 7. The pairings for each of the critical blocks were counterbalanced between participants and the target images for each trial were presented in random order. Reaction times for each of the trials in the critical blocks were recorded.

Procedure

Participants were taken to individual testing rooms where they provided consent and then completed one of four randomly assigned IATs.

   After completing the study, participants were debriefed and thanked for their participation.

RESULTS AND DISCUSSION

D-scores were first calculated as outlined by Greenwald and colleagues (2003), with higher scores indicating an implicit preference for neutral White targets relative to smiling Black targets. A 2 (Categorization Task: Race or Emotion) × 2 (Condition: Experimental or Control) ANOVA revealed a significant main effect of Categorization Task, $F(1, 144) = 273.26$, $p < .001$, $\eta_p^2 = .66$. Participants categorizing by race ($D = .37$, $SD = .37$) had significantly different implicit attitudes than those categorizing by emotion ($D = -.60$, $SD = .33$). No main effect of Condition, $F(1, 144) = .07$, $p = .79$, $\eta_p^2 = .001$, or interaction, $F(1, 144) = .45$, $p = .51$, $\eta_p^2 = .003$, emerged.

As can be seen in Figure 2, participants in the Race-Categorization condition ($D = .34$, $SD = .38$), showed greater implicit positivity toward neutral White targets relative to smiling Black targets, $t(33) = 5.16$, $p < .001$, $d = 1.80$. The magnitude of this bias did not differ from those in the Race-Control condition ($D = .39$, $SD = .36$), who sorted faces with the same emotional expression, $t(77) = .65$, $p = .52$, $d = .15$. By contrast, those in the Emotion-Categorization condition ($D = -.58$, $SD = .31$), showed greater implicit positivity toward smiling Black targets relative to neutral White targets, $t(32) = -10.65$, $p < .001$, $d = 3.77$, and similarly, the magnitude of bias did not differ from those in the Emotion-Control condition ($D = -.61$, $SD = .34$), who saw only White faces, $t(67) = -.29$, $p = .77$, $d = .07$.

The first goal of this study was to determine what effect emotional expression, a dynamic social category, has on the attitudes spontaneously evoked by Black and White target faces. Our results provide evidence that the way in which targets are categorized can have a profound effect on people's implicit attitudes. Identical smiling Black and neutral White target faces elicited different attitudes depending on whether they were being categorized by race or by emotional expression. When participants categorized these faces by race, they showed an implicit preference favoring neutral White, as compared to smiling Black, faces. By contrast, when categorizing by emotional expression, participants showed an implicit preference favoring smiling Black, over neutral White, faces.

A second goal of this study was to determine whether the magnitude of bias would be affected by other target attributes that might have competing associations, or whether, once categorized, bias would be similar to what is found when faces differ only by one social category. Somewhat surprisingly given the results of Mitchell et al. (2003, Study 2), participants showed an equivalent magnitude of pro-White bias when they categorized smiling Black and neutral White faces by race in the Race-Categorization condition as when they categorized Black and White faces with identical emotional expressions. Unlike earlier work on stereotype activation and inhibition, we also found little evidence that multiple social categories had an inhibitory effect on attitudes—participants who completed an IAT containing targets who differed systematically along two perceptually salient social categories did not show *increased* bias favoring the dimension along which they categorized. Instead, once categorized, the magnitude of participants' implicit bias was comparable to when targets differed only in one way (i.e., by race

or by emotional expression). Put differently, once viewed through the lens of race, evidence of racial biases ensued regardless of other relevant social category information. Similarly, once categorized by emotional expression, racial biases did not attenuate the positive associations elicited by smiling Black targets. One limitation to this research was the relatively small sample size, and ensuing low statistical power, for this study. We addressed this concern in Study 3.

STUDY 3

The goal of Study 3 was to replicate and extend the findings of Studies 1 and 2 in several key ways. First, we aimed to increase our statistical power by replicating with a larger sample, the main finding of Study 2, that people's implicit attitudes toward smiling Black and neutral White targets, as measured by the IAT, depend on which social category is used as the basis for categorization. To make each condition more visually comparable, in Study 3 participants were presented with an identical header containing both race and emotional expression information, specifically, a smiling Black cartoon image on one side and a neutral White cartoon image on the other (see Appendix A). Prior to completing blocks containing smiling Black and neutral White target faces, participants in the Race-Categorization and Emotion-Categorization conditions were provided with written instructions asking them to categorize the pictures by either race or emotional expression, respectively.

Second, we examined which of these two social categories participants would spontaneously use as the basis for social categorization. We did this by randomly assigning participants to either the Race-Categorization condition, the Emotion-Categorization condition, or to a third Ambiguous-Categorization condition. In this third condition, participants used the same header provided in the other two conditions to sort smiling Black and neutral White faces. However, participants assigned to this condition were simply instructed to sort the pictures based on the header provided. Given that the pictures used in the task were of smiling Black and neutral White faces, this allowed these participants to successfully complete the task while categorizing by race, emotion, or both of these aspects of the target faces.

To determine how participants primarily categorized targets throughout the task, participants sorted three additional trials at the end of each IAT that appeared to participants to be part of the final block of trials. We selected three trials to ensure that we provided a sufficient number of trials to determine how participants were primarily categorizing (i.e., by race or by emotional expression), while simultaneously limiting the number of trials in such a way that they appeared seamlessly to be part of the task. The stimuli used in these three trials were randomly selected from four novel pictures, two of neutral Black faces and two of smiling White faces.

We again predicted that implicit attitudes would depend on how targets were categorized, with smiling Black faces eliciting greater positivity, and neutral White

faces eliciting less positivity, when participants categorized these faces by emotional expression as opposed to race.

METHOD

Participants

One hundred and seventy-two non-Black undergraduates participated for course credit. Data from three participants were excluded from our analyses because they had an average reaction time on all trials that was at least three standard deviations above the mean (Cvencek, Greenwald, & Meltzoff, 2011; Cvencek, Meltzoff, & Greenwald, 2011). The final sample of 169 was comprised of South Asian (n = 38 [27 female], M_{age} = 18 years, SD = 1.66), White (n = 36 [26 female], M_{age} = 19 years, SD = 3.04), Middle-Eastern (n = 26 [19 female], M_{age} = 20 years, SD = 3.26), East/Southeast Asian (n = 23 [16 female], M_{age} = 19 years, SD = 4.25), Hispanic (n = 4 [4 female], M_{age} = 23 years, SD = 4.27), and other (including non-Black multiracial/other race, n = 42 [31 female], M_{age} = 20 years, SD = 5.09) participants.

Measures

Implicit Racial Attitudes. Implicit attitudes were measured using similar IATs to those used in Study 1, with a few noteworthy differences. First, conditions now included only the Race-Categorization and Emotion-Categorization from Study 1, as well as a new Ambiguous-Categorization condition. Second, participants in each of our conditions saw an identical header (see Appendix A), which included a smiling Black cartoon face on one side and a neutral White cartoon face on the other. The instructions for the Race- and Emotion-Categorization conditions were the same as in Study 1, however the headers in these conditions now contained both race and emotion information. Participants in the Ambiguous-Categorization condition were simply asked to match the image in the center of the screen with the appropriate header. In this Ambiguous-Categorization condition, no reference to either race or emotion was made. Third, an additional critical block containing three trials was added to the end of each IAT (Block 8). For participants, these three trials appeared to follow seamlessly from the trials in Block 7, and contained an identical header. Each target image was selected from a set of four images that included *neutral* Black and *smiling* White faces. If participants had been categorizing by race, we anticipated that they would sort a neutral Black target face with the smiling Black header and a smiling White target face with the neutral White header. On the other hand, if they had been categorizing primarily by emotion, we anticipated that they would sort a neutral Black face with the neutral White header and a smiling White target face with the smiling Black header.
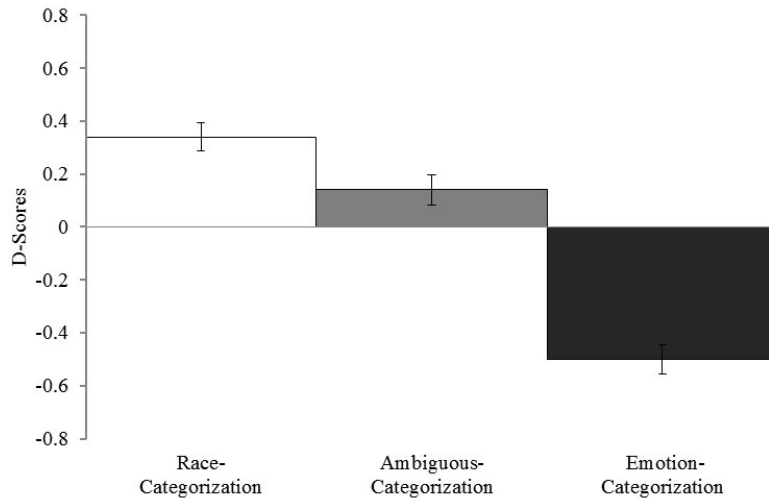
FIGURE 3. Study 3 IAT D-scores by condition. Higher D-scores represent greater implicit preference for neutral White versus smiling Black targets. Error bars represent the standard error.

Procedure

The procedure for Study 3 was identical to that of Study 2, with the exception that participants were now randomly assigned to one of three conditions (Condition: Race-Categorization, Emotion-Categorization, or Ambiguous-Categorization).

RESULTS AND DISCUSSION

As in Study 1, D-scores were first calculated as outlined by Greenwald and colleagues (2003), with higher scores indicating an implicit preference for neutral White targets relative to smiling Black targets. A one-way ANOVA using D-scores as the dependant variable revealed the expected significant effect of condition, $F(2, 166) = 64.09$, $p < .001$, $\eta_p^2 = .44$. Post-hoc Tukey HSD tests revealed that participants in the Race-Categorization condition ($D = .34$, $SD = .38$) showed significantly different attitudes than participants in both the Emotion-Categorization ($D = -.50$, $SD = .34$, $p < .001$, $d = 2.33$) and Ambiguous-Categorization ($D = .13$, $SD = .53$, $p = .04$, $d =.44$) conditions. Responses in the Emotion- and Ambiguous-Categorization conditions also differed significantly from each other ($p < .001$, $d = 1.46$, see Figure 3). As in Study 1, participants in the Race-Categorization condition ($D = .34$, $SD = .38$), showed greater implicit positivity toward neutral White targets relative to smiling Black targets, $t(57) = 6.83$, $p < .001$, $d = 1.81$. By contrast, those in the Emotion-Categorization condition ($D = -.50$, $SD = .34$) showed greater implicit positivity toward smiling Black targets relative to neutral White targets, $t(59) = -11.12$, $p < .001$, $d = 2.89$.
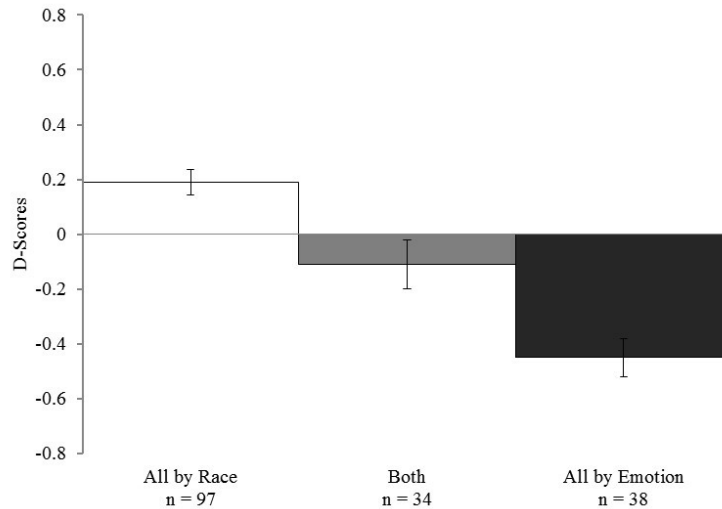
FIGURE 4. Study 3 IAT D-scores for participants who categorized the final three trials either exclusively by race, by both race and emotion, or exclusively by emotion. Higher D-scores represent a greater implicit preference for neutral White versus smiling Black targets. Error bars represent the standard error.

Next, we examined the spontaneous categorization trials across each of the conditions in order to determine how participants across conditions categorized the majority of these trials (see Table 1). Across all conditions, the majority (57%) of participants categorized all three of the spontaneous trials exclusively by race, while fewer categorized all three of the trials exclusively by emotion (23%) or by both race and emotion (20%). Individuals who categorized all three trials by race demonstrated significantly greater positivity toward neutral White, as opposed to smiling Black, target faces ($D = .19$, $SD = .47$), $t(96) = 3.99$, $p < .001$, $d = .81$. By contrast, those who categorized all three trials by emotional expression demonstrated significantly greater positivity toward smiling Black versus neutral White faces ($D = -.45$, $SD = .45$), $t(37) = -6.20$, $p < .001$, $d = 2.04$. These two groups differed significantly from each other, $t(133) = -7.21$, $p < .001$, $d = 1.25$. Participants who used a combination of race and emotion information to categorize the final three trials did not show a significant bias ($D = -.11$, $SD = .55$), $t(33) = -1.21$, $p = .24$, $d = .42$. This group differed significantly from those who categorized entirely by race, $t(129) = 3.11$, $p = .002$, $d = .55$, as well as from those who categorized entirely by emotion, $t(70) = -2.87$, $p = .005$, $d = .69$ (see Figure 4).

We also examined just those participants in the Ambiguous-Categorization condition. The majority of these participants categorized all three trials by race (69%), with fewer categorizing by emotion (14%) or by both race and emotion (18%). Overall, participants in this condition, who were not asked to categorize in a particular way, showed a preference for neutral White targets over smiling Black targets ($D = .14$, $SD = .52$), $t(50) = 1.98$, $p = .05$, $d = .42$. In addition, those who categorized all three trials by race ($D = .18$, $SD = .45$) showed a significantly different

TABLE 1. Number of Participants in Each Condition Who Categorized the Three Final Spontaneous Categorization Trials in Block 8 by Race, Emotion, or Both (Study 3)

	All by Race		Both		All by Emotion	
	#	%	#	%	#	%
Race-Categorization Condition (n = 58)	53	91	5	9	0	0
Spontaneous-Categorization Condition (n = 51)	35	69	9	18	7	14
Emotion-Categorization Condition (n = 60)	9	15	20	33	31	52

pattern of bias than those who categorized all three final trials by emotion (D = -.27, SD = .63), $t(40)$ = -2.25, p = .03, d = .82.

These findings are consistent with the results of Study 2 and provide additional evidence that the way in which a target is categorized has a direct impact on people's ensuing implicit attitudes. Participants who categorized these targets by race, either spontaneously or because of the instructions that they received, subsequently showed more positive attitudes toward neutral White targets and less positive attitudes toward smiling Black targets, than participants who categorized these targets based on their emotional expression.

In this study we also made use of three additional spontaneous categorization trials in the IAT to determine whether participants were primarily attending to race or emotional expression when categorizing our face stimuli. To our knowledge, this is the first time that the IAT has been used in this way to examine both spontaneous categorization and implicit attitudes. Overall, the majority of participants categorized these last three trials by race, suggesting that race may have been particularly salient when people were required to think categorically about others. As expected, participants' implicit biases generally showed a pattern that was consistent with these categorization trials. Participants who categorized these final trials by emotional expression showed an implicit preference for smiling Black targets, whereas those (the majority) who categorized these final trails by race showed an implicit preference for neutral White targets.

GENERAL DISCUSSION

Across three studies, we examined the impact of race and emotional expression on non-Black participants' implicit attitudes. In Study 1, we found that the race and the emotional expression of targets had independent effects on people's implicit attitudes. Previous research provides some evidence that social categories with competing associations, including social roles, occupations, or even sexual orientation, can attenuate biases on priming measures (Barden et al., 2004; Wittenbrink et al., 2011) or approach-avoidance tasks (Remedios et al., 2011). Using an established priming measure, we found that emotional expression, a dynamic attribute, similarly affected implicit attitudes. Our non-Black participants showed greater positivity following smiling, as opposed to neutral, face primes. Nonetheless, it

is important to note that racial biases on this priming measure still emerged. As a group, White face primes still evoked greater positivity than Black face primes.

In Studies 2 and 3, we examined non-Black participants' attitudes toward smiling Black and neutral White target faces on an implicit task that required participants to actively categorize these target faces. We found evidence that implicit attitudes are informed by contextually driven social categorization processes. When targets were categorized by race, a pro-White bias emerged. Neutral White faces were implicitly preferred to smiling Black faces. Importantly, the magnitude of this implicit bias did not differ from when participants were presented with White and Black targets with identical emotional expressions. This finding suggests that when categorizing by race, implicit attitudes will reflect racial biases, regardless of the emotion being expressed by the targets.

It is important to note, however, that when categorizing smiling Black and neutral White faces based on their emotional expression, participants showed an implicit preference for smiling Black over neutral White faces. The magnitude of positivity evoked by a smiling Black target did not differ from when the smiling target was White, suggesting that implicit racial biases can be obscured when targets are categorized differently, even when the criteria for social categorization is not a stable social category. This finding adds to the work of Mitchell and colleagues (2003) by demonstrating that novel targets and dynamic target attributes can influence implicit attitudes. This finding differs somewhat from the results of their second study, however, where the unique aspects of exemplars (specifically seeing liked Black and disliked White famous faces) attenuated implicit racial biases on the IAT. It seems possible that this difference in results was because participants in their study needed to select famous people from a list, making other aspects of their identity (being liked or disliked) salient just prior to completing the implicit measure. Future research will be needed to further identify those situations in which social categories with competing associations will and will not influence implicit attitudes.

In addition, in the current research, we found that when participants were not told how to categorize target faces when completing the IAT (Study 3), race was overwhelmingly used as a basis for categorization. Over two-thirds of our participants categorized by race, suggesting that when thinking categorically about others, race dominated over emotional expression, and this affected the attitudes that were spontaneously evoked. It is possible that when categorizing novel targets, participants will rely more heavily on stable social categories, such as race, as opposed to more dynamic aspects, such as emotional expression, however further research will be needed to better understand the conditions under which this is the case. In addition, future research establishing the reliability and validity of the ambiguous categorization condition as a measure of social categorization could provide useful insight. Although the final spontaneous categorization trials allowed us to infer whether participants were primarily categorizing by race or emotional expression by the end of the task, it is less clear whether participants categorized only one way throughout the task. It is possible that participants attended to both aspects of identity throughout (i.e., Happy Black targets) and/or used alternate

categorization strategies on different trials (i.e., categorizing by emotional expression on some trials and race on others), and future research should investigate these possibilities as well as individual differences in people's spontaneous categorization tendencies. It is also unclear from these results whether participants were recoding the response options during the task in Studies 2 and 3 (Meissner & Rothermund, 2013), and this would be worthy of future investigation.

The findings in these studies are consistent with recent theories of person construal (Freeman & Ambady, 2014), which focus on the interplay between facial cues, social categories, and contextual information in informing social categorization processes in a dynamic and interactive manner. The current research extends this work to show the effect of this dynamic interplay on attitudes. When the implicit measure primed smiling and neutral racial exemplars, both race and emotional expression informed implicit attitudes. Importantly, although the emotional expression of the targets had a much larger effect, racial biases still emerged. When the implicit measure required categorization, these implicit attitudes depended largely on the way in which the targets were categorized. This suggests that implicit attitudes can be informed by multiple perceptually salient social categories, however only under certain conditions. When the context strongly encourages categorization by one dimension, as was the case for those assigned to the race- or emotional-categorization conditions of the IAT, attitudes may reflect biases toward the dominant category. In daily living, this might be likely to occur when a specific social category (i.e., gender, race) is distinctive in a social context. For example, a Black women's race is likely to be most distinctive in a context with many White women, but her gender should be particularly distinctive when she is surrounded by Black men (Mitchell et al., 2003).

The findings also provide additional evidence that attitudes are online constructions that integrate pre-existing group-level associations with contextual cues. This conceptualization fits with a growing body of research to suggest that implicit racial attitudes are more malleable than was initially theorized (Blair, 2002; Lai et al., 2014). For example, there are now many studies demonstrating that exposure to counterstereotypical exemplars, prior to completing an implicit measure, can attenuate racial biases (Dasgupta & Greenwald, 2001; Gonzalez, Steele, & Baron, 2017; Lai et al., 2014; Lowery et al., 2001; Wittenbrink et al., 2001). It is unclear whether these interventions are effective because they challenge negative associations with the larger racial group or activate a particular subtype (e.g., Black professional) that has pre-existing positive associations.

Unlike this previous research, our findings should not be taken as evidence for the malleability of implicit racial biases. Instead, our findings provide additional evidence for the dynamic and contextual nature of implicit attitudes. In our studies the same targets elicited different attitudes, depending on how they were construed. This construal was often determined by contextual cues (i.e., task instructions and structure) directing social categorization. From a practical standpoint, these findings suggest that contexts which encourage racial categorization are likely to elicit racial biases, even when other relevant social information is available.

Taken together, the present studies suggest that both person perception and social categorization processes play an important role in determining the attitudes that are spontaneously evoked by others. When people were thinking categorically, race was likely to dominate over more dynamic emotional expressions to influence implicit attitudes. By contrast, when viewing people as individuals, as was the case in our priming task, emotional expression had a more prominent influence, yet racial biases still emerged. Through future research we hope to continue to enhance our basic understanding of the influence of multiple social categories on people's attitudes. It is clear that when other person attributes are available, these can attenuate some forms of implicit racial bias in the short term. While this provides theoretical insight into the influence of social categorization on racial biases, this is by no means a solution to racial prejudice. Instead, future research will need to continue to identify how implicit racial biases that contribute to the perpetuation of social inequalities can be challenged and reduced.

REFERENCES

Allport, G. W. (1954). *The nature of prejudice.* New York: Doubleday Anchor Books.

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., & Morton, J. B. (2010). The development of spontaneous gender stereotyping in childhood: Relations to stereotype knowledge and stereotype flexibility. *Developmental Science, 13*(2), 298-306. https://doi.org/10.1111/j.1467-7687.2009.00880.x

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*(1), 5-22. https://doi.org/10.1037/0022-3514.87.1.5

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*, 53-58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261. https://doi.org/10.1207/S15327957P-SPR0603_8

Brewer, M. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition, volume I,* (pp. 1-36). Hillsdale, NJ: Erlbaum.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16*(4), 330-350. https://doi.org/10.1177/1088868312440047

Cañadas, E., Lupiáñez, J., Kawakami, K., Niedenthal, P. M., & Rodríguez-Bailón, R. (2016). Perceiving emotions: Cueing social categorization processes and attentional control through facial expressions. *Cognition and Emotion, 30*(6), 1149-1163. https://doi.org/10.1080/02699931.2015.1052781

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*(6), 1314-1329. https://doi.org/10.1037//0022-3514.83.6.1314

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychological Science, 15*(12), 806-813. https://doi.org/10.1111/j.0956-7976.2004.00760.x

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The Preschool Implicit Association Test (PSIAT). *Journal of Experimental Child Psychology*, *109*(2), 187-200. https://doi.org/10.1016/j.jecp.2010.11.002

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*(3), 766-779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*(5), 800-814. https://doi.org/I0.1037//0022-3514.81.5.800

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*(3), 316-328. https://doi.org/10.1006/jesp.1999.1418

Donges, U. S., Kersting, A., & Suslow, T. (2012). Women's greater ability to perceive happy facial emotion automatically: Gender differences in affective priming. *PLoS One, 7*(7), e41745. https://doi.org/10.1371/journal.pone.0041745

Dovidio, J. F., Gaertner, S. E., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8*(2), 88-102. https://doi.org/10.1037//1099-9809.8.2.88

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-28. https://doi.org/10.1037//0022-3514.82.1.62

Dunahm, Y., Baron, A. S., & Banaji, M. R. (2007). Children and social groups: A developmental analysis of implicit consistency in Hispanic Americans. *Self and Identity, 6*(2-3), 238-255. https://doi.org/10.1080/15298860601115344

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*(6), 876-893. https://doi.org/0.1037/0022-3514.87.6.876

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297-327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1-74. https://doi.org/10.1016/S0065- 2601(08)60317-2

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological review, 118*(2), 247. https://doi.org/10.1037/a0022327

Freeman, J. B., & Ambady, N. (2014). The dynamic interactive model of person construal: Coordinating sensory and social processes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 235-248). New York: Guilford.

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS One, 6*(9), e25107. https://doi.org/10.1371/journal.pone.0025107

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283-310). New York: Cambridge University Press.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*(4), 509-517. https://doi.org/10.1037/0022-3514.60.4.509

Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing children's implicit racial bias through exposure to positive out-group exemplars. *Child Development, 88*(1), 123-130. https://doi.org/10.1111/cdev.12582

Govorun, O., & Payne, B. K. (2006). Ego-depletion and prejudice: Separating automatic and controlled components. *Social*

Cognition, 24(2), 111-136. https://doi.org/10.1521/soco.2006.24.2.111

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test (IAT) can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553-561.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test (IAT): I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test (IAT): III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17-41. https://doi.org/10.1037/a0015575

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science, 15*(5), 342-345. https://doi.org/10.1111/j.0956-7976.2004.00680.x

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13*(5), 505-521. https://doi.org/10.1080/026999399379168

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37*(9), 1539. https://doi.org/10.1037/0022-3514.37.9.1539

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284-308. https://doi.org/10.1037/0033-295X.103.2.284

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., et al. (2014). A comparative investigation of 17 interventions to reduce implicit racial preferences. Journal of Experimental Psychology: General, 143, 1765-1785.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Technical Report A-8. University of Florida, Gainesville, FL.

Levin, S., Sinclair, S., Veniegas, R. C., & Taylor, P. L. (2002). Perceived discrimination in the context of multiple group memberships. *Psychological Science, 13*(6), 557-560. https://doi.org/10.1111/1467-9280.00498

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*(1), 5-18. https://doi.org/10.1037/0022-3514.82.1.5

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*(5), 842-855. https://doi.org/10.1037//0022-3514.81.5.842

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*(3), 397-407. https://doi.org/10.1037/0022-3514.69.3.397

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution, 25*(4), 563-579.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology, 104*, 45-69. https://doi.org/10.1037/a0030734

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*(3), 455-469. https://doi.org/10.1037/0096-3445.132.3.455

Nelson, S. C. (2008). Feeling thermometer. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (p. 277). Thousand Oaks, CA: Sage.

Pauker, K., Williams, A., & Steele, J. R. (2016). Children's racial categorization in context. *Child Development Perspectives, 10*(1), 33-38. https://doi.org/10.1111/cdep.12155

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psy-*

chology, 81(2), 181-191. https://doi.org/10.1037//0022-3514.81.2.181

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277-293. https://doi.org/10.1037/0022-3514.89.3.277

Payne, B. K., & Lundberg, K. (2014). The Affect Misattribution Procedure (AMP): Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8*(12), 672-686. https://doi.org/10.1111/spc3.12148

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*(5), 729-738. https://doi.org/10.1162/089892900562552

Pittinsky, T. L., Shih, M., & Ambady, N. (1999). Identity adaptiveness: Affect across multiple identities. *Journal of Social Issues, 55*(3), 503-518. https://doi.org/10.1111/0022-4537.00130

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*(3), 811-832.

Remedios, J. D., Chasteen, A. L., Rule, N. O., & Plaks, J. E. (2011). Impressions at the intersection of ambiguous and obvious social categories: Does gay + Black = likable? *Journal of Experimental Social Psychology, 47*(6), 1312-1315. https://doi.org/10.1016/j.jesp.2011.05.015

Rohr, M., Degner, J., & Wentura, D. (2012). Masked emotional priming beyond global valence activations. *Cognition and Emotion, 26*, 224-244. https://doi.org/10.1080/02699931.2011.576852

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*(5), 885-904.

Williams, A., & Steele, J. R. (2017). Examining children's implicit racial attitudes using exemplar and category-based measures. *Child Development.* https://doi.org/10.1111/cdev.12991

Williams, A., Steele, J. R., & Lipman, C. (2016). Assessing children's implicit attitudes using the Affective Misattribution Procedure (AMP). *Journal of Cognition and Development, 17*, 505-525. https://doi.org/10.1080/15248372.2015.1061527

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*(5), 815-827. https://doi.org/10.1037//0022-3514.81.5.815
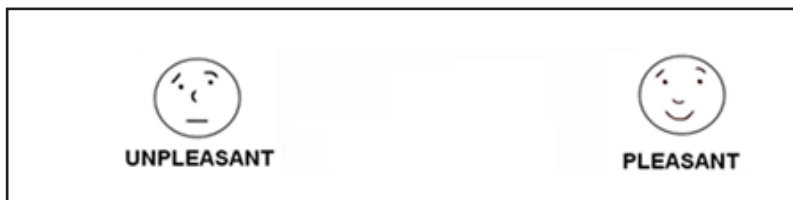
APPENDIX A

Sample headers presented in the Implicit Association Test (IAT).

Race-Categorization and Race-Control Conditions (Study 2)
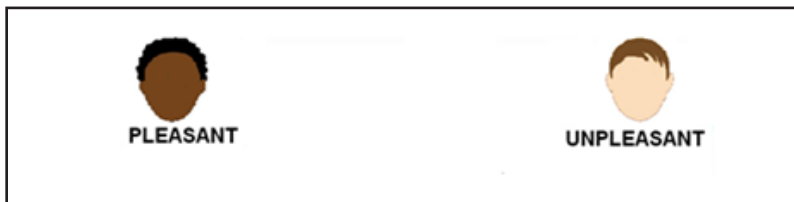
Emotion-Categorization and Emotion-Control Conditions (Study 2)

All Conditions (Study 3)

Note. The side and pairing for each header in the critical blocks were counterbalanced between participants. In Study 3, the header was identical for all participants, however, the instructions differed by condition.