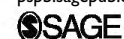


# Using Groups to Measure Intergroup Prejudice

Erin Cooley<sup>1</sup> and B. Keith Payne<sup>2</sup>

Personality and Social  
Psychology Bulletin  
2017, Vol. 43(1) 46–59  
© 2016 by the Society for Personality  
and Social Psychology, Inc  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0146167216675331  
pspb.sagepub.com



## Abstract

Implicit measures of racial attitudes often assess reactions to images of individuals to infer attitudes toward an entire social category. However, an increasing amount of research indicates that responses to individuals are highly dependent on context and idiosyncratic features of individual exemplars. Thus, using images of individuals to assess beliefs about a whole social category may not be ideal. Across three time points, we predicted that using images of groups would mitigate the influence of idiosyncratic features of individual targets and, thus, provide a better measurement tool to assess beliefs about a category to which all group members belong. Results revealed that an implicit measure that presented images of Black and White groups had greater construct validity, test–retest reliability, and predictive validity as compared with an implicit measure that presented the same exemplars individually. We conclude that groups provide a window into existing beliefs about social categories.

## Keywords

implicit cognition, group processes, affect misattribution procedure, prejudice

Received July 26, 2015; revision accepted September 7, 2016

## Using Groups to Measure Intergroup Prejudice

Allport (1954) described prejudice as hostile attitudes that are directed toward “a person who belongs to a *group*, simply because he belongs to that *group*” (p. 7). However, despite the centrality of groups to the development of prejudice, current measures of implicit prejudice (e.g., Implicit Association Tests [IATs] and affective priming measures) assess reactions to individual exemplars. On each trial of the task, a photo or name representing a single person from the group is presented. In the present article, we will propose that this discrepancy between what is being measured (i.e., reactions to individuals) and what is being inferred (i.e., attitudes toward social groups) may be problematic.

One drawback of using images of individuals to assess beliefs about a single social category such as “Black Americans” is that individuals simultaneously belong to myriad social categories (e.g., gender, race, socioeconomic status). Thus, responses to individuals can depend on whichever social category is currently salient (Mitchell, Nosek, & Banaji, 2003). Furthermore, individuals vary in their phenotypic characteristics such as skin tone, facial features, and attractiveness. Critically, variations in attributes such as these can greatly affect the types of associations and affective reactions that individual category members elicit

(Hagiwara, Kashy, & Cesario, 2012; Maddox, 2004). As a result, if implicit measures aim to assess attitudes toward a particular social category, then both the phenotypic variability of individuals as well as their multiple simultaneous social category memberships likely contribute to measurement error. We propose that one way to minimize the influence of idiosyncratic features of individual targets may be the use of groups rather than individuals to assess category beliefs. Given that groups are integral to the development of stereotyping and prejudice, we propose that groups (i.e., three people standing together) may activate beliefs about a social category to which all group members belong more effectively than individuals. Furthermore, groups may be less swayed by features of individual exemplars. As a result, groups may serve as a better tool to measure racial prejudice. This idea builds upon a long history of how to best measure attitudes about socially sensitive topics such as race using both explicit and implicit measures.

<sup>1</sup>Colgate University, Hamilton, NY, USA

<sup>2</sup>The University of North Carolina at Chapel Hill, USA

### Corresponding Author:

Erin Cooley, Colgate University, Olin Hall, 111 C, Hamilton, NY 13346, USA.

Email: ecooley@colgate.edu

## Measurement of Prejudice

Before the civil rights movement in America, racial prejudice was often easily visible and thus easily measured using explicit measures that asked people to directly report their beliefs about Black people. However, greater norms of tolerance following key civil rights legislation in the 1960s meant that self-reports of prejudice became complicated by increased motivations to control the expression of prejudice. For this reason, modern racial prejudice is often assessed using both explicit and implicit measures. While explicit measures assess consciously endorsed attitudes (Devine, 1989), implicit measures can capture attitudes that are independent of motivations. The automaticity of reactions captured by implicit measures often indicate that even people who intend to be egalitarian can hold negative associations with particular social groups such as Black Americans (Devine, 1989; Greenwald, McGhee, & Schwartz, 1998; Payne, Cheng, Govorun, & Stewart, 2005). Critically, the negative attitudes captured by implicit measures are able to predict racially biased behavior above and beyond explicit measures making them an important tool for understanding the many forms prejudice can take (Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). However, in addition to assessing attitudes that are independent of motivations to control prejudice, implicit measures also differ from explicit measures in a way that perhaps unintentionally hinders their ability to assess racial prejudice. In particular, while explicit measures often assess attitudes toward entire racial categories (e.g., Black people), implicit measures often assess attitudes toward individual exemplars (e.g., a Black individual). For example, the symbolic racism scale (Henry & Sears, 2002) captures explicit racial prejudice with items such as “Irish, Italian, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same.” Similarly, feeling thermometers assess attitudes toward entire social categories rather than individuals (e.g., “How warm do you feel toward Blacks?”). In contrast, implicit measures often use images of individuals to assess beliefs about the entire category. This is true in both IATs (Greenwald et al., 1998) and affective priming tasks (Fazio, Jackson, Dunton, & Williams, 1995; Payne et al., 2005). To demonstrate why the use of individual exemplars to assess social category beliefs may be problematic, we can think about the procedure for a common affective priming procedure that we will focus on in the present article, the Affect Misattribution Procedure (AMP; Payne et al., 2005)

The AMP takes advantage of the ambiguity associated with the origin of an attitude to examine the degree to which people misattribute the source of their affective reactions to a quickly flashing photo prime onto subsequent ambiguous stimuli. In particular, across many trials participants view quickly flashing photo primes (e.g., Black or White individuals) followed by ambiguous stimuli (Chinese symbols).

Participants are asked to ignore the photo and to make a judgment about the Chinese symbol such as whether the Chinese symbol is relatively unpleasant or relatively pleasant as compared with the average Chinese symbol. If participants systematically evaluate the Chinese symbols as more unpleasant after trials preceded by images of Black individuals versus White individuals, then an implicit bias is inferred. Indeed, despite the intention of ignoring the photo prime, research indicates that attitudes toward the primes are misattributed to the Chinese symbols that follow. This measure is a valid and reliable measure of implicit bias that effectively predicts behavior (Cameron et al., 2012). However, of particular relevance to the present article, this measure currently measures affective reactions to images of individual exemplars from a social category of interest to assess beliefs about the social category as a whole. The use of images of individuals in this measure may unintentionally increase measurement error due to idiosyncratic responses to specific exemplars. This proposition is supported by an increasing amount of research on how specific features of individual targets drive automatic responses in measures of automatic racial associations.

## Individuals Can Be Categorized in Multiple Ways

Individuals have multiple intersecting identities based on their many group memberships. Importantly, the way that individual exemplars are categorized in the context of an implicit measure can dictate how prejudice emerges (Govan & Williams, 2004; Mitchell et al., 2003). In fact, Mitchell and colleagues (2003) found that implicit bias as measured by an IAT varied depending on whether individual exemplars were categorized according to race or gender. When Black women were categorized by gender, they were evaluated more positively consistent with overall evaluations of women; in contrast, when Black women were categorized by race, they were evaluated more negatively consistent with overall evaluations of Black Americans. Like IATs, responses on affective priming tasks are also influenced by how individual exemplars are categorized (De Houwer, Hermans, & Spruyt, 2001; Olson & Fazio, 2003). For example, Olson and Fazio (2003) found that instructing participants to categorize faces according to race within a priming measure led to different responses than when participants were not given this instruction. Thus, the inherent multiple category memberships of individual exemplars lead them to activate highly divergent evaluations based on contextual factors that influence how they are categorized.

One reason that individuals can be categorized in multiple ways is that individuals vary in their phenotypic features. Research shows that phenotypic qualities of individual exemplars can drive responses on measures of prejudice (Bluemke & Fries, 2006; Livingston & Brewer, 2002; Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). For example, on both implicit and explicit measures, Black individuals

with darker skin tone (see Maddox, 2004 for review), or more afrocentric facial features (Blair, Judd, & Chapleau, 2004; Eberhardt, Goff, Purdie, & Davies, 2004), are more likely to be the targets of racial bias. For example, participants responded more quickly to negative words after viewing either a dark-skinned Black person (vs. light-skinned) or a Black person with more prototypically Black facial features (vs. less prototypical; Hagiwara et al., 2012). Furthermore, Ma and Correll (2011) found racial biases in shooter tasks actually reverse when Black targets do not have prototypically "Black" facial features. Together, these findings suggest that using images of individuals to assess stereotyping and prejudice toward Black Americans as a category may involve a lot of variability—varying greatly from one Black American individual to the next.

In the present article, we propose a theoretically driven way to minimize the measurement error that results from using individual exemplars in implicit measures of prejudice. In particular, we know from existing research that a prototypical Black American individual elicits prejudice associated with Black Americans more than a Black American individual perceived as less prototypical (Maddox, 2004). Building from this idea, we propose that a group of people that share a social category membership may be perceived as more typical of that category and thus elicit prejudice associated with that category more effectively. This reasoning is consistent with existing models of impression formation. In fact, both dual process models (Srull & Wyer, 1988) as well as continuum models (Fiske & Neuberg, 1990) of impression formation agree that perception of a target activates associated mental representations of categories to which that target belongs (e.g., age, race, gender, socioeconomic status). Furthermore, the degree to which the target is perceived as fitting the activated mental representations determines whether category information is applied to the target. Thus, we will first test whether groups are perceived as more typical or representative of their racial category than individuals. If so, we will then test whether perceiving groups of people from a particular social category may more effectively elicit people's implicit prejudice toward that social category. Consistent with this reasoning, other research indicates that groups accentuate the application of explicit stereotypes (Cooley & Payne, 2016). In particular, Black groups were rated more stereotypically (i.e., more aggressive and untrustworthy) than the same Black Americans presented as individuals. Similarly, groups of Asian Americans were judged more stereotypically (i.e., more hard working and better at math) than the same Asian Americans presented as individuals. Extending these findings to implicit measures, we suspect that groups will also automatically elicit prejudice associated with a category to which all group members belong more readily than individuals. If this is the case, then using images of groups (vs. individuals) in an implicit measure may improve the psychometric qualities of the measure.

## Study 1

In Study 1, we first evaluated whether photos of Black and White groups would be perceived as more representative of their racial category than photos of individuals from those groups. We predicted that, holding all else constant, merely putting people into same-race groups would lead those people to be perceived as more representative of their racial category. To create a particularly stringent test of our hypothesis, we began by choosing images of people from existing stimuli databases that our research has indicated are perceived as most typical of their race. Participants then rated these people on how representative and typical they appeared of their race when they were alone and when they were in groups. This allowed us to test whether even the most prototypical exemplars are perceived as more representative of their racial category when in groups.

## Method

**Participants.** Using G\*Power 3 software (Faul, Erdfelder, Lang, & Buchner, 2007), we determined that we needed a sample of at least 199 for this study to have adequate power ( $1 - \beta > .80$ ) to detect a small effect ( $f = .10$ ). To ensure adequate power, we recruited 250 participants on Amazon Mechanical Turk. One person submitted a completion code for the study, but did not actually complete the study. Thus, we were left with a final sample of 249 people (57% female; 88% White, 5% Black, 6% Asian, 1% Other). All participants were paid US\$0.50.

**Materials.** Individual and group photos were created by first selecting 18 Black individuals and 18 White individuals from an existing database (Minear & Park, 2004). The particular photos we used were selected based on previous research that found that these individuals were most representative of their racial categories. To create group photos, we randomly combined our 18 Black individuals and 18 White individuals into six White three-person groups and six Black three-person groups. Next we created two stimuli conditions so that no Black or White target appeared as an individual and in a group to the same participant. (However, across all participants, all stimuli appeared both as an individual and as a group.) These stimuli conditions were designed so that some participants saw half of the Black stimuli (nine faces) and half of the White stimuli (nine faces) as groups and the remaining nine Black stimuli and nine White stimuli as individuals. The other stimuli condition reversed which stimuli were viewed as individuals and which were viewed in groups.

**Procedure.** After agreeing to an electronic informed consent, participants viewed images of individual Black Americans, groups of Black Americans, individual White Americans, and groups of White Americans. Photos were presented in

counterbalanced blocks separated by race. Within each block, photos of individuals and groups were presented in a random order. For the White photos, participants rated each photo on two items: “How representative is this person [are these people] of White Americans in general?” and “How typical is this person [are these people] of White Americans in general?” Responses were made on 1 (*not at all*) to 100 (*extremely*) sliding scales. For the Black photos, participants made the same ratings but “White Americans” was substituted for “Black Americans.” Participants concluded with demographic information.

## Results and Discussion

Our main hypothesis was that groups would be rated as more representative of their race than individuals. To test this hypothesis, we first created composite “representativeness” scores separately for each type of photo by averaging ratings of representativeness and typicality. This index was highly reliable across all types of photos (Black individuals:  $\alpha = .97$ ; Black groups:  $\alpha = .95$ ; White individuals:  $\alpha = .97$ ; White groups:  $\alpha = .93$ ). Next we conducted a 2 (race: Black vs. White)  $\times$  2 (group: group vs. individual) repeated measures ANOVA. Consistent with our main hypothesis, groups were rated as more representative of their racial categories ( $M = 63.53$ ,  $SD = 19.74$ ) than individuals ( $M = 61.68$ ,  $SD = 18.65$ ),  $F(1, 248) = 14.83$ ,  $p < .001$ ,  $\eta_p^2 = .06$ . There was also an unexpected main effect of race such that Black people were rated as more representative of their race overall ( $M = 63.41$ ,  $SD = 19.06$ ) than White people ( $M = 61.80$ ,  $SD = 19.33$ ),  $F(1, 248) = 5.43$ ,  $p = .02$ ,  $\eta_p^2 = .02$ . Given that most participants were White, this may be an outgroup homogeneity effect. There was no interaction,  $F(1, 248) = 1.56$ ,  $p = .21$ ,  $\eta_p^2 = .01$ . Together, these results indicate that even when we used images of individuals that were chosen to be most representative of their race, placing these individuals into groups significantly increased the perception of how representative they were of their racial categories. Because theories of impression formation propose that targets who are perceived as representative of a category are most likely to have category information applied to them (Brewer, 1988; Fiske & Neuberg, 1990), we next tested whether groups may serve as a better tool to access people’s implicit racial category beliefs. We examined this hypothesis in the context of an AMP (Payne et al., 2005). We predicted that an AMP that used images of Black and White groups as primes (i.e., a Group AMP) would have better construct validity, test–retest reliability, and predictive validity than an AMP that uses Black and White individuals (i.e., an Individual AMP).

## Study 2: Time 1

At Time 1, we hypothesized that a Group AMP would have greater construct validity than an Individual AMP. To test this hypothesis, we examined whether implicit bias on a

Group AMP predicted variability in explicit measures of prejudice above and beyond variability predicted by an Individual AMP. We reasoned that if a Group AMP more effectively captures attitudes toward the social category (i.e., Black Americans), this should decrease measurement error and increase correspondence with explicit measures of the same construct (Cunningham, Preacher, & Banaji, 2001).

## Method

**Participants.** Participants were undergraduates enrolled in a marketing course at the University of North Carolina at Chapel Hill. For this course, participants had the option to complete three 1-hr research studies for course credit across the semester. Because all three time points were available to the same pool of about 200 students, this helped us have a high return rate across time points. We handled missing data by using the maximum number of data points available at each time point. Notably, participants were able to complete any time point regardless of whether they had completed prior time points.

At Time 1, 175 participants completed the study. Following convention, for all analyses that included either the Individual AMP or the Group AMP, we eliminated 29 people who reported being able to speak/read Chinese (given the use of Chinese symbols in the AMP) and eight participants who failed to follow task instructions by pressing the same button throughout the entire Individual AMP, the entire Group AMP, or both AMPs. This left us with a final sample of 138 people (66% male; 76% White, 7% Black, 8% Asian, 9% Other).

**Procedure.** After signing an informed consent, participants were told they would complete a task that would measure their ability to concentrate (i.e., the implicit measures) and another task that would ask them to report their attitudes (i.e., the explicit measures). We counterbalanced implicit and explicit measures. The implicit measures included two versions of an AMP: one version that used images of individuals as primes (Individual AMP), and one version that used images of groups as primes (Group AMP). A single trial of either AMP consisted of viewing a photo prime (i.e., a Black or White individual or group) for 200 ms followed by a Chinese symbol that appeared for 125 ms. Finally, participants saw a blank gray square until they made their response. Participants were asked to press “A” if they thought the symbol was relatively unpleasant and “L” if they thought the symbol was relatively pleasant. Participants were told that the measure was evaluating their ability to concentrate and that their task was to ignore the photo prime and to make a rating about the Chinese symbol by relying on their gut reaction to the Chinese symbol. For each measure, participants completed a total of two practice trials and 40 critical trials. For the Individual AMP, of the 40 critical trials, 20 were preceded by White individuals and 20 were preceded by Black individuals in a

random order. For the Group AMP, of the 40 critical trials, 20 were preceded by White groups and 20 were preceded by Black groups.

In addition to the implicit measures, participants completed two explicit measures of racial prejudice: the Symbolic Racism Scale (Henry & Sears, 2002) and a series of feeling thermometers. These feeling thermometers asked participants to rate how warm or cold they felt toward Black, White, Asian, and Hispanic people on 0 (*very cold/unfavorable*) to 100 (*very warm/favorable*) scales. Finally, participants reported their internal and external motivations to control prejudice (Plant & Devine, 1998), their confidence and certainty in their attitudes toward Black people and White people on 1 (*not at all*) to 7 (*extremely*) scales, demographic information, and whether they could understand the Chinese symbols in the implicit measures.

## Results

**Preliminary analyses.** First, we calculated implicit bias on the Individual AMP and the Group AMP. For both measures, we calculated the proportion of trials that participants responded with “pleasant” separately for trials preceded by White Americans and trials preceded by Black Americans. Overall, on the Individual AMP, participants responded more positively to White American individuals ( $M = 0.57$ ,  $SD = 0.17$ ) than to Black American individuals ( $M = 0.52$ ,  $SD = 0.19$ ),  $F(1, 137) = 8.69$ ,  $p = .004$ ,  $\eta_p^2 = .06$ . Similarly, on the Group AMP participants responded more positively to White American groups ( $M = 0.60$ ,  $SD = 0.19$ ) as compared with Black American groups ( $M = 0.53$ ,  $SD = 0.22$ ),  $F(1, 137) = 9.73$ ,  $p = .002$ ,  $\eta_p^2 = .07$ . To calculate implicit bias scores, we subtracted the proportion of pleasant responses on trials preceded by images of Black Americans from the proportion of pleasant responses on trials preceded by images of White Americans separately for the Individual AMP and the Group AMP. The Group and Individual AMP were also significantly correlated with each other,  $r = .56$ ,  $p < .001$ . However, consistent with our proposition that the Group AMP would have better psychometric properties than the Individual AMP, the Group AMP had significantly higher internal consistency ( $\alpha = .73$ ) than the Individual AMP ( $\alpha = .52$ ),  $\chi^2(1, N = 138) = 14.59$ ,  $p = .0001$ .

Next, we calculated explicit prejudice on both a feeling thermometer as well as the Symbolic Racism Scale. Overall, participants reported greater warmth toward White Americans ( $M = 8.78$ ,  $SD = 1.53$ ) than Black Americans ( $M = 7.60$ ,  $SD = 1.95$ ),  $F(1, 137) = 42.44$ ,  $p < .001$ ,  $\eta_p^2 = .24$ . Explicit prejudice on a feeling thermometer was calculated as the amount of warmth felt toward Black Americans subtracted from the amount of warmth felt toward White Americans. Higher values on this index indicated greater warmth toward White Americans than Black Americans ( $M = 1.19$ ,  $SD = 2.14$ ). Explicit prejudice on the Symbolic Racism Scale was calculated as the average response across the eight items ( $M = 2.22$ ,

**Table 1.** Correlations Among Group AMP, Individual AMP, and Explicit Prejudice, Time 1.

Measure	1	2	3
1. Group AMP	—		
2. Individual AMP	.56**	—	
3. Explicit prejudice	.41**	.38**	—

Note. AMP = affect misattribution procedure.

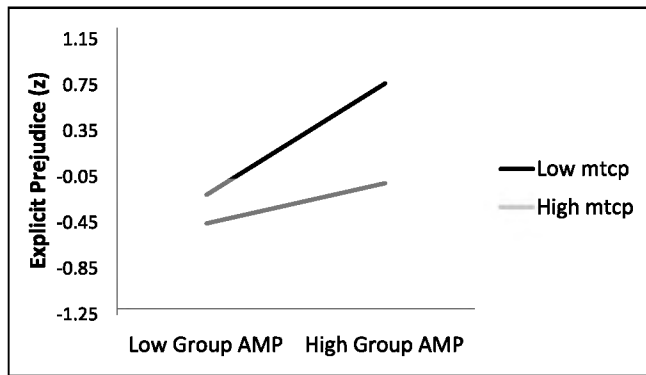
\*\* $p < .001$ .

$SD = 0.47$ ;  $\alpha = .78$ ). Finally, we averaged standardized scores on both the feeling thermometer and Symbolic Racism Scale to create a single standardized index of explicit prejudice ( $\alpha = .74$ ).<sup>1</sup>

**Main analyses.** Our main hypothesis was that a Group AMP would display better convergent validity with explicit measures of attitudes toward Black Americans than an Individual AMP. We tested this hypothesis with a series of regression analyses using standardized variables.

First, we predicted explicit prejudice from implicit bias on the Individual AMP. Higher implicit bias on the Individual AMP was associated with higher explicit bias,  $b = .34$ , 95% confidence interval (CI) = [.20, .48],  $t = 4.74$ ,  $p < .001$  (Table 1). Next we predicted explicit prejudice from implicit bias on the Group AMP. Again, higher implicit bias was associated with higher explicit bias,  $b = .37$ , 95% CI = [.23, .64],  $t = 5.25$ ,  $p < .001$ . Finally, we entered race bias on the Individual AMP and race bias on the Group AMP as simultaneous predictors of explicit prejudice. By entering both predictors simultaneously, we could evaluate the convergent validity of each measure above and beyond the influence of the other. Interestingly, at Time 1, the Individual AMP continued to predict unique variability in explicit prejudice,  $b = .19$ , 95% CI = [.03, .35],  $t = 2.32$ ,  $p = .02$ . However, most critical to our theoretical perspective, the Group AMP predicted variability in explicit prejudice above and beyond the Individual AMP,  $b = .26$ , 95% CI = [.10, .42],  $t = 3.15$ ,  $p = .002$ .

**Supplementary analysis.** An alternative explanation of our findings is that the Group AMP predicted explicit prejudice toward the racial category as a whole more effectively than the Individual AMP because participants were more likely to treat the Group AMP as an explicit measure. Under this reasoning, using images of groups as primes may have led participants to be more likely to ignore task instructions and explicitly rate the photo primes (see Bar-Anan & Nosek, 2012 for related criticisms of the AMP). To help rule out this possibility, we next conducted supplementary analyses with internal motivations to control prejudice. Because implicit measures capture attitudes that emerge before motivations can influence responding, we reasoned that bias captured on an implicit measure should be more likely to diverge from bias captured on an explicit measure among those high in



**Figure 1.** Moderation of the relationship between the group AMP and explicit prejudice by motivations to control prejudice, Time 1.

Note. AMP = affect misattribution procedure.

motivations to control prejudice. Thus, a way to evaluate whether the Group AMP was treated as an implicit measure is to examine whether the relationship between the Group AMP and explicit prejudice is moderated by motivations to control prejudice in a theoretically meaningful way. To test this, we conducted a linear regression predicting explicit prejudice by implicit bias on the Group AMP, average scores on the internal Motivations to Control Prejudice subscale (MCP; Plant & Devine, 1998;  $M = 15.75$ ,  $SD = 3.15$ ;  $\alpha = .81$ ), and the Group AMP  $\times$  MCP interaction. All variables were standardized for this analysis. Results revealed two main effects and the predicted interaction. A main effect for the Group AMP,  $b = .33$ , 95% CI = [.20, .46],  $t = 5.07$ ,  $p < .001$ , indicated that greater bias on the Group AMP predicted greater levels of explicit prejudice. The main effect for MCP indicated that greater motivations to control prejudice predicted less explicit prejudice,  $b = -.28$ , 95% CI = [-.41, -.15],  $t = 4.29$ ,  $p < .001$ . Finally, there was the significant, predicted interaction,  $b = -.16$ , 95% CI = [-.30, -.02],  $t = 2.17$ ,  $p = .03$  (Figure 1). Simple effects revealed that for those who were low in motivations to control prejudice ( $-1 SD$ ), higher implicit bias was associated with higher explicit prejudice,  $b = .49$ , 95% CI = [.30, .69],  $t = 5.01$ ,  $p < .001$ . Among those high in motivations to control prejudice ( $+1 SD$ ), greater implicit bias was associated with greater explicit prejudice, but to a lesser degree,  $b = .19$ , 95% CI = [.00, .37],  $t = 1.97$ ,  $p = .05$ . This is the pattern of results that we would expect to see if the Group AMP is operating as an implicit, not explicit, measure.

Although the analyses with the Group AMP were most critical in ruling out an alternative explanation of findings at Time 1 (i.e., that the Group AMP correlated more highly with explicit measures because it was treated as an explicit measure), we also conducted a parallel analysis to examine whether internal motivations to control prejudice moderated the relationship between implicit bias on the Individual AMP and explicit prejudice. Results of this analysis revealed two

main effects. Overall, higher implicit bias was associated with higher explicit prejudice,  $b = .32$ , 95% CI = [.19, .45],  $t = 4.88$ ,  $p < .001$ , and higher motivations to control prejudice were associated with less explicit prejudice,  $b = -.31$ , 95% CI = [-.44, -.18],  $t = 4.63$ ,  $p < .001$ . The predicted interaction did not reach significance,  $b = -.10$ , 95% CI = [-.25, .05],  $t = 1.29$ ,  $p = .20$ , although the pattern of effects were in the predicted direction.

## Discussion

Overall, Time 1 results were consistent with the greater convergent validity of the Group AMP (vs. the Individual AMP). Although both the Individual AMP and Group AMP were correlated with explicit prejudice and were highly correlated with each other, the Group AMP predicted variability in explicit prejudice above and beyond the Individual AMP. In addition, the internal consistency of the Group AMP was significantly greater than the internal consistency of the Individual AMP and was on par with other widely used measures (Cunningham et al., 2001; Robinson, Shaver, & Wrightsman, 1991). Together, these results are consistent with the idea that while the Individual AMP may partially capture racial prejudice, it may also capture variability in reactions to individual exemplars. In contrast, the Group AMP may serve as a more reliable indicator of the underlying construct of interest (i.e., racial prejudice). Furthermore, supplementary analyses were inconsistent with the alternative explanation that a Group AMP correlated more highly with explicit measures because images of groups led people to treat the Group AMP as an explicit measure. Instead, motivations to control prejudice moderated the relationship between the Group AMP and explicit prejudice in a way that was theoretically consistent with the AMP being an implicit measure. Although a parallel analysis with the Individual AMP did not yield a significant implicit bias by motivations to control prejudice interaction, the pattern of effects was in the predicted direction. Furthermore, if the lack of interaction of the Individual AMP with motivations to control prejudice predicting explicit prejudice were driven by people treating the Individual AMP as an explicit measure, we would have expected to find a stronger correlation of the Individual AMP (as compared with the Group AMP) with explicit prejudice. This is the opposite of what we found.

To further assess the psychometric properties of the Group AMP, we repeated the same measures from Time 1 at Time 2, 3 to 4 weeks later. Our purposes in Time 2 were threefold. First, we attempted to replicate the critical analyses from Time 1. Second, we compared the test-retest reliability of the Individual AMP and the Group AMP. We reasoned that if using images of groups minimizes the measurement error that results from using images of individuals, a Group AMP should be more reliable than an Individual AMP over time. Finally, we used confirmatory factor analyses (CFAs) and structural equation modeling to compare two possible

interpretations of our findings. One interpretation is our hypothesis: We expect that both the Group AMP and Individual AMP are assessing the same latent construct of implicit race bias, but that the Group AMP is tapping the construct better than an Individual AMP. Another possibility is that the Group AMP and Individual AMP are assessing different constructs all together (and that this drives the differences we are seeing in their relations with explicit measures of racial prejudice). To test these two interpretations empirically, we compared two nested structural equation models using Time 1 and Time 2 data: a one-factor model and a two-factor model.

## Study 2: Time 2

At Time 2, we tested two new hypotheses in addition to replicating our Time 1 findings. First, we predicted that the test–retest reliability between Time 1 and Time 2 would be greater for the Group AMP as compared with the Individual AMP. Second, we predicted that the Group AMP and Individual AMP assess the same latent construct, but that the Group AMP does so more effectively.

## Method

**Participants.** Participants were 172 undergraduates drawn from the same marketing class as Time 1 participants. As in Time 1, we eliminated 27 participants who reported being able to speak/read Chinese because of the use of Chinese symbols in the AMP and nine participants who hit the same button throughout at least one of the AMPs. Finally, due to a computer error, explicit data are missing for five participants. Thus, we were not able to use these participants' data for analyses that included explicit measures. This left a Time 2 sample of 131 participants.

**Procedure.** The procedure was identical to Time 1 with one exception. In particular, at Time 2, participants also explicitly evaluated the photo primes used in the implicit measures on a scale from 1 (*very unpleasant*) to 7 (*very pleasant*) after completing the implicit measures.

## Results

**Preliminary analyses.** As at Time 1, on the Individual AMP, participants indicated significantly more positive responses on trials preceded by White individuals ( $M = 0.55$ ,  $SD = 0.19$ ) than trials preceded by Black individuals ( $M = 0.49$ ,  $SD = 0.21$ ),  $F(1, 130) = 6.78$ ,  $p = .01$ ,  $\eta_p^2 = .05$ . Similarly, on the Group AMP, participants reported marginally more positive responses in trials preceded by White groups ( $M = 0.56$ ,  $SD = 0.21$ ) as compared with trials preceded by Black groups ( $M = 0.51$ ,  $SD = 0.22$ ),  $F(1, 130) = 3.75$ ,  $p = .06$ ,  $\eta_p^2 = .03$ . Implicit racial bias scores were calculated as in Time 1 and the Group and Individual AMP were significantly correlated with

**Table 2.** Correlations Among Group AMP, Individual AMP, and Explicit Prejudice, Time 2.

Measure	1	2	3
1. Group AMP	—		
2. Individual AMP	.65**	—	
3. Explicit prejudice	.47	.39**	—

Note. AMP = affect misattribution procedure.

\*\* $p < .001$ .

each other,  $r = .66$ ,  $p < .001$ . Importantly, replicating Time 1, and consistent with hypotheses, the Group AMP ( $\alpha = .80$ ) showed significantly higher internal reliability than the Individual AMP ( $\alpha = .68$ ),  $\chi^2(1, N = 131) = 11.27$ ,  $p < .001$ .

Finally, explicit prejudice was calculated in the same way as Time 1 for feeling thermometers ( $M = 1.21$ ,  $SD = 2.27$ ) and the Symbolic Racism Scale ( $M = 2.19$ ,  $SD = 0.45$ ;  $\alpha = .79$ ). In particular, we averaged standardized scores on each measure to create a single standardized measure of explicit prejudice ( $\alpha = .76$ ). All other variables were standardized before the following analyses.

**Main analyses replicated from Time 1.** To replicate Time 1 findings, we predicted explicit prejudice from implicit bias as assessed by the Individual AMP. Greater implicit bias was associated with greater explicit prejudice,  $b = .34$ , 95% CI = [.20, .48],  $t = 4.75$ ,  $p < .001$  (Table 2). Next we predicted explicit prejudice by the Group AMP. Greater implicit bias on the Group AMP was associated with greater explicit prejudice,  $b = .41$ , 95% CI = [.28, .55],  $t = 5.96$ ,  $p < .001$ . Finally, we conducted a regression analysis predicting explicit prejudice from implicit bias on the Individual AMP and implicit bias on the Group AMP as simultaneous predictors. Results strongly supported hypotheses and replicated Time 1 results. In particular, when both AMPs were included as predictors, the Group AMP predicted significant variability in explicit prejudice,  $b = .33$ , 95% CI = [.15, .51],  $t = 3.61$ ,  $p < .001$ , above and beyond the variability accounted for by the Individual AMP. In fact, at Time 2, the coefficient for the Individual AMP when controlling for the Group AMP was reduced to non-significance,  $b = .13$ , 95% CI = [−.05, .30],  $t = 1.39$ ,  $p = .17$ . Thus, extending Time 1 patterns, at Time 2, the Group AMP demonstrated strong convergent validity with explicit prejudice—above and beyond the Individual AMP.

**Test–retest reliability (Time 1 and Time 2 data).** Finally, we assessed test–retest reliability of the Individual AMP and the Group AMP. For the analyses involving test–retest reliability, we started with 151 undergraduates who completed both Time 1 and Time 2. From this sample, we filtered out 25 participants who reported being able to speak/read Chinese and 11 participants who hit the same button on at least one of the AMPs either at Time 1 or Time 2. This left us with 115 participants for our analysis of test–retest reliability.

**Table 3.** Test–Retest Reliability for the Individual AMP and Group AMP.

Measure	1	2	3	4
1. Time 1 Group AMP	—			
2. Time 2 Group AMP	<b>.66**</b>	—		
3. Time 1 Individual AMP	.61**	.52**	—	
4. Time 2 Individual AMP	.51**	.63**	<b>.50**</b>	—

Notes. AMP = affect misattribution procedure. Bolded values are significantly different from one another at  $p = .03$ .

\*\* $p < .001$ .

As can be seen in Table 3, results revealed that test–retest reliability was good overall—both for the Individual AMP ( $r = .49$ ) and the Group AMP ( $r = .63$ ). Critically, consistent with our hypothesis, the test–retest reliability for the Group AMP was significantly greater than the test–retest reliability for the Individual AMP, Pearson-Fillion,  $z = 2.23$ ,  $p = .03$ .

**Factor analysis (Time 1 and Time 2 data).** Next we compared the results of a CFA for a one-factor model and a two-factor model. We predicted that a one-factor model would fit the data well and that the two-factor model would not result in significant improvement in model fit. Furthermore, within the one-factor model, we predicted that the Group AMP would load more highly onto the latent construct of racial prejudice than would the Individual AMP.

Although we theoretically prefer a one-factor model, to adequately compare a one-factor model with a two-factor model, we first needed to structure our data so that a two-factor model had the possibility of fitting the data. In factor analyses, factors need to have at least three indicators (Bollen & Long, 1993); thus, for a two-factor model to have the potential to fit the data we needed at least three indicators to represent an Individual AMP and three indicators to represent the Group AMP. In our previous Time 1 and Time 2 analyses, we only had two indicators for each type of AMP: Time 1 Group AMP and Time 2 Group AMP and Time 1 Individual AMP and Time 2 Individual AMP. To increase our number of indicators, we calculated the proportion of “pleasant” responses separately for random halves of the Black prime trials and White prime trials for Individual and Group AMPs at Time 1 (Table 4) and Time 2 (Table 5). Then we created bias scores by subtracting the proportion of “pleasant” responses on Black trials from “pleasant” responses on White trials for each half. As a result, we had two Time 1 indicators each for the Individual AMP and Group AMP and two Time 2 indicators each for the Individual AMP and Group AMP.

First, we tested a one-factor CFA using maximum likelihood estimation using Mplus (Muthén & Muthén, 1998–2015). The scale of the latent factor was set by fixing one factor loading to 1. Consistent with our hypothesis, a one-factor model fit the data well,  $\chi^2(20, N = 115) = 29.66$ ,  $p =$

**Table 4.** Proportion of “Pleasant” Responses on Trials Preceded by Black Primes and White Primes for Random Halves of the Individual and Group AMPs, Time 1.

Primes	M	SD
<b>Black</b>		
Group AMP		
Half 1	0.52	0.23
Half 2	0.52	0.24
Individual AMP		
Half 1	0.50	0.22
Half 2	0.51	0.22
<b>White</b>		
Group AMP		
Half 1	0.60	0.20
Half 2	0.60	0.22
Individual AMP		
Half 1	0.57	0.20
Half 2	0.55	0.19

Note. AMP = affect misattribution procedure.

**Table 5.** Proportion of “Pleasant” Responses on Trials Preceded by Black Primes and White Primes for Random Halves of the Individual and Group AMPs, Time 2.

Primes	M	SD
<b>Black</b>		
Group AMP		
Half 1	0.50	0.24
Half 2	0.51	0.24
Individual AMP		
Half 1	0.49	0.25
Half 2	0.49	0.23
<b>White</b>		
Group AMP		
Half 1	0.57	0.24
Half 2	0.55	0.21
Individual AMP		
Half 1	0.57	0.22
Half 2	0.56	0.21

Note. AMP = affect misattribution procedure.

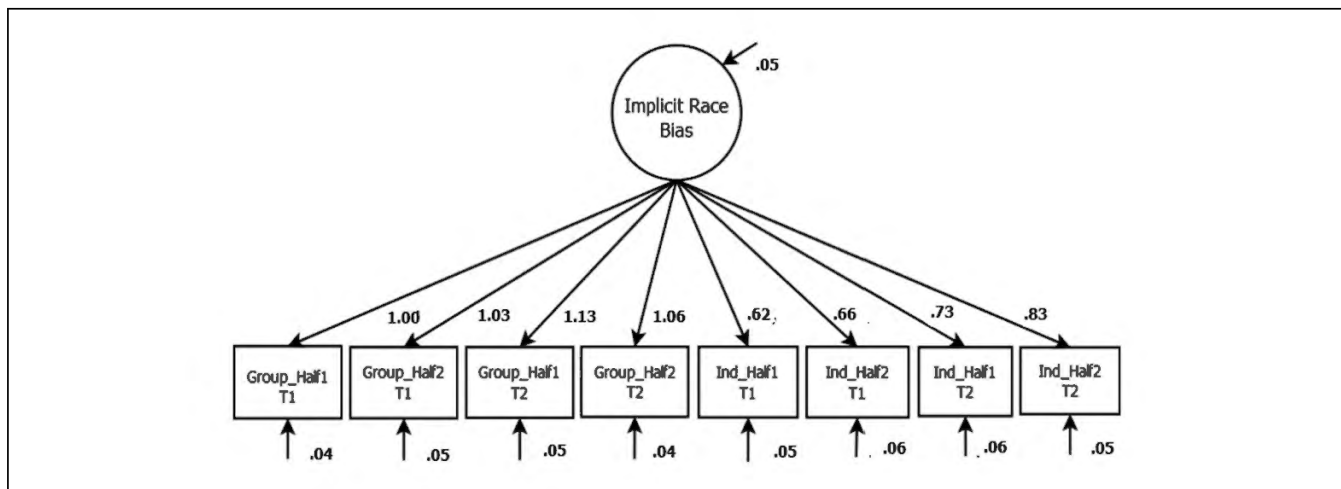
.08. In addition to a chi-square test, we also evaluated model fit using several other fit indices (Table 6). The Akaike information criterion (AIC) and Bayesian information criterion (BIC) evaluate comparative model fit with lower values indicating better model fit. The Tucker–Lewis index (TLI) evaluates model fit compared with the null model with a penalty for each parameter estimated. Values of greater than .95 indicates good model fit (Hu & Bentler, 1999). Finally, the root mean square error of approximation (RMSEA) examines differences between observed and expected model covariances. An RMSEA of .01 indicates excellent fit, .05 indicates good fit, and .08 indicates mediocre fit; a 90% CI near zero for RMSEA indicates strong model fit (MacCallum, Browne, &



**Table 6.** Goodness-of-Fit Indicators of One-Factor and Two-Factor Models for Implicit Race Bias ( $N = 115$ ).

Model	$\chi^2$	df	$\chi^2_{diff}$	AIC	BIC	TLI	90% CI RMSEA
One factor	29.66	20		118.08	183.96	0.96	[.00, .11]
Two factor	28.05	19	1.61	118.47	187.09	0.96	[.00, .11]

Note. A chi-square difference test of nested models indicated that a two-factor model did not provide a significant improvement in model fit over a one-factor model. AIC = Akaike information criterion; BIC = Bayesian information criterion; TLI = Tucker–Lewis index; CI = confidence interval; RMSEA = root mean square error of approximation.

**Figure 2.** Confirmatory factor analysis for a one-factor model, Time 1 and Time 2.

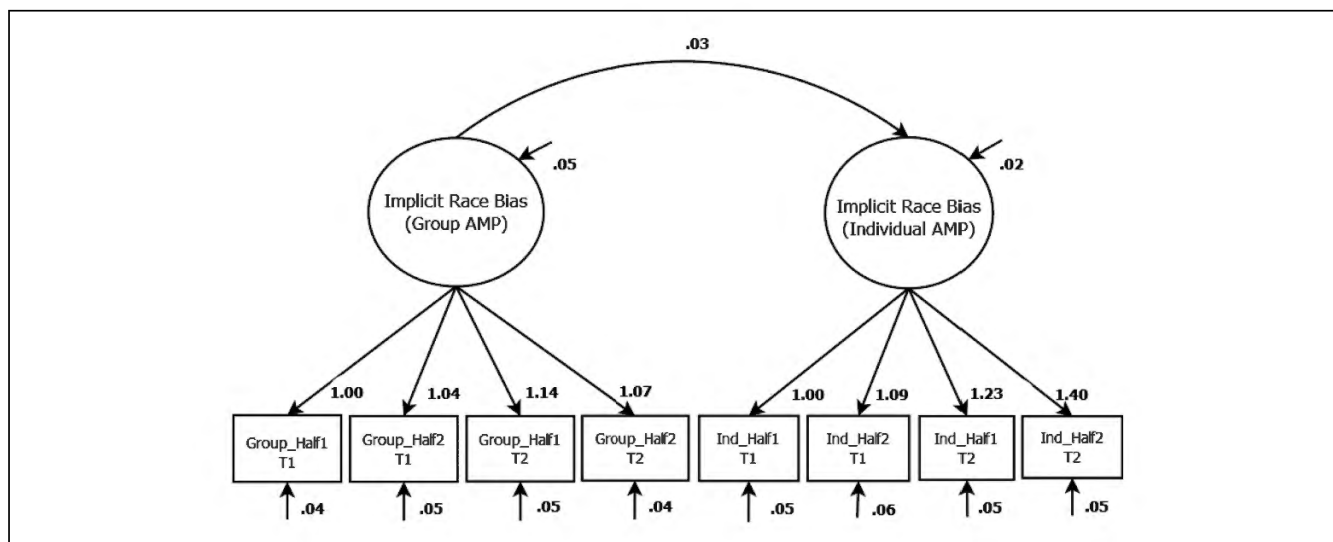
Note. The factor loading of the first indicator was set to 1 to set the scale of the factor. The model fits well,  $\chi^2(20, N = 115) = 29.66, p = .08$ . All factors loadings are significant,  $p < .001$ . Loadings for the Group AMP are descriptively larger than loadings for the Individual AMP. AMP = affect misattribution procedure.

Sugawara, 1996). As can be seen in Table 4, by all of these indices, the model fit well. Further consistent with hypotheses, indicators representing the Group AMP loaded more highly onto this single factor than indicators representing the Individual AMP (Figure 2). Thus, the Group AMP provided more valid indicators of the underlying latent construct of implicit racial bias than the Individual AMP.

Next, we tested a two-factor model (Figure 3). One way to think about a two-factor model in relation to a one-factor model is that, in a two-factor model, the correlation between two latent factors is allowed to be freely estimated, whereas in a one-factor model this pathway is restricted to be equal to 1. Because a two-factor model places fewer restrictions on the data than a one-factor model, all else being equal, we would expect a two-factor model to fit better. Thus, the critical question is whether the two-factor model fits significantly better than a one-factor model. Although the two-factor model fit the data,  $\chi^2(19, N = 115) = 28.05, p = .08$  (Table 6), it did not fit significantly better than a one-factor model,  $\chi^2_{diff}(1, N = 115) = 1.61, p = .20$ . In addition, the correlation between factors in the two-factor model was extremely high,  $r = .998, p < .001$ , indicative of a single underlying factor. Together, these analyses indicate that the predicted, and more parsimonious, one-factor model best describes our data.

## Discussion

Time 2 results were consistent with our hypotheses. First, replicating Time 1, the Group AMP continued to predict variability in an explicit measure of prejudice above and beyond the Individual AMP. In fact, when both the Group AMP and the Individual AMP were included as predictors of explicit prejudice, the coefficient for the Individual AMP was no longer a significant predictor. Moreover, the Group AMP also exhibited significantly greater internal reliability than the Individual AMP. Together, these findings replicate Time 1 and indicate that the Group AMP may capture the construct of interest—prejudice toward Black Americans—with less measurement error than an Individual AMP. Second, Time 2 data revealed that although both AMPs had strong test–retest reliability, the Group AMP provided significantly greater test–retest reliability than the Individual AMP. These results were particularly encouraging given that 3 to 4 weeks had elapsed between measurement periods. To our knowledge, this is some of the first data to examine test–retest reliability of the AMP (see also Gawronski, Morrison, Phillips, & Galdi, 2015). These findings are consistent with our proposal that the Group AMP may reduce measurement error and thus provide a more



**Figure 3.** Confirmatory factor analysis for a two-factor model, Time 1 and Time 2.

Note. The factor loading of the first indicator was set to 1 to set the scale of the factor. This model does not fit significantly better than a one-factor model. All factor loadings are significant,  $p < .001$ . AMP = affect misattribution procedure.

effective measure of the underlying construct—namely, prejudice toward a social category to which all group members belong. Finally, we compared a structural equation model that represented the data with one latent factor and a model that represented the data with two latent factors—one for the Group AMP and one for the Individual AMP. We predicted and found evidence for the superior model fit of the one-factor model. Furthermore, the indicators for the Group AMP loaded more highly on the latent factor than did the indicators for the Individual AMP. This pattern of findings is consistent with the idea that the Group AMP (as compared with the Individual AMP) is a more valid indicator of the latent construct of implicit racial bias. Finally, at Time 3, we evaluated whether the Group AMP had greater predictive ability than the Individual AMP.

## Study 2: Time 3

At Time 3, which took place 1 to 2 weeks later, we examined the ability of the Group AMP to predict racial biases in hiring decisions. We predicted that implicit bias as assessed by the Group AMP at Time 2 would be a better predictor of racially biased hiring than implicit bias as assessed by the Individual AMP at Time 2.

### Method

**Participants.** Participants for Time 3 were 194 undergraduates from the University of North Carolina at Chapel Hill drawn from the same pool of marketing students from Time 1 and Time 2. Our Time 3 sample was able to be larger than Time 1 and Time 2 because each time point was presented as an independent study. Thus, even participants who did not

complete both Time 1 and Time 2 had the option to complete Time 3.

For analyses predicting hiring decisions at Time 3 with Time 2 implicit measures, we started with 158 participants who completed both of these time points. Following convention, we eliminated 26 participants who reported being able to speak/read Chinese and nine who pressed the same button on one or both of the AMPs at Time 2. Finally, one participant who did not complete the hiring decision was not able to be included in our main analyses using this variable. This left us with 122 participants.

**Procedure.** At Time 3, participants learned that we were interested in how people make hiring decisions with only small amounts of information. In particular, we used a casuistry paradigm following Norton, Vandello, and Darley (2004). Participants were told to imagine they were in charge of a hiring decision for a job in a construction company that equally requires both experience in the engineering industry and a strong engineering background. Next, participants read about four (purportedly real) job candidates, two of whom clearly stood out from the rest: one had slightly more education and one had slightly more experience (Table 7). Critically, participants were randomly assigned to learn that one of these two top candidates had a name associated with Black Americans (i.e., Jamal Washington) and the other had a name associated with White Americans (i.e., Greg Schwartz). Thus, some participants learned that Jamal had slightly more experience and that Greg had slightly more education and some learned the reverse. After viewing the candidates in a random order, participants were next asked to evaluate the candidates on how likely they would be to hire the candidate, how skilled the candidate seemed, and how competent the

**Table 7.** Job Candidates for the Time 3 Hiring Decision.

Job candidates	Age	Experience	Education
Jamal Washington/Greg Schwartz	39	5 years in engineering industry	BS in engineering; NCMA certification
Greg Schwartz/Jamal Washington	37	9 years in engineering industry	BS in engineering
Brad Mueller	32	2 years in engineering industry	BA in management
Todd Schneider	37	1 year in construction	BA in English

Note. NCMA certification was described as a certification that educates architects and engineers about new technology and design practices being used in masonry construction. NCMA = National Concrete Masonry Association.

candidate seemed on 1 (*not at all*) to 100 (*extremely*) scales. Participants then ranked the candidates. Finally participants completed our main dependent variable that asked them to choose one candidate to hire for the position. Participants completed the study by rating their certainty and confidence in their decision on 1 (*not at all*) to 7 (*extremely*) scales and responding to an open response question about why they made the hiring decision that they did. Participants were debriefed for all three time points at the end of the semester by their course instructor.

## Results

**Preliminary analyses.** First, we tested whether participants showed a racial bias in hiring decisions overall by evaluating the likelihood of hiring Jamal (1) or one of the other White candidates (0). Overall, among the Time 3 sample, participants were equally likely to choose a candidate other than Jamal (46.9%) as they were to choose Jamal (53.1%),  $\chi^2(1, N = 192) = .75, p = .39$ . Interestingly, among our marketing sample, hiring preferences were strongly predicted by whom-ever had more experience,  $\chi^2(1, N = 192) = 75.01, p < .001$  (Table 8). Participants who learned that Jamal had experience were much *more* likely to hire Jamal (84%), whereas those who learned that Jamal had education (and Greg had experience) were much *less* likely to hire Jamal (21%).

**Main analyses.** Next, we assessed our main hypothesis that racial bias as assessed with the Group AMP would better predict racial biases in hiring decisions than would racial bias as assessed with the Individual AMP. To do this, we conducted a binary logistic regression predicting hiring decisions (0 = *not Jamal*; 1 = *Jamal*) by implicit race bias on the Group AMP as assessed at Time 2 and implicit race bias on the Individual AMP as assessed at Time 2. Results were consistent with hypotheses. In particular, the Group AMP significantly predicted hiring decisions such that greater bias on the Group AMP was associated with a reduced likelihood of hiring Jamal,  $B = -.51, SE = .26, p = .048$ , odds ratio ( $OR$ ) = .60. In contrast, the Individual AMP did not predict racial biases in hiring decisions,  $B = .28, SE = .25, p = .26, OR = 1.33$ . Because both implicit measures were included as simultaneous predictors of hiring decisions, these results indicate that the Group AMP predicted racial

**Table 8.** Decisions to Hire Jamal Based on Condition.

	Jamal education	Jamal experience	Total
Hire Jamal	20	82	102
Hire other candidate	74	16	90
Total	94	98	192

biases in hiring decisions above and beyond variability accounted for by the Individual AMP.

Although not central to the hypotheses tested in the current article, we also tested whether the Group AMP predicted hiring decisions above and beyond explicit racial prejudice and motivations to control prejudice. We tested this by conducting the same logistic regression as described above but by adding standardized explicit prejudice and internal motivations to control prejudice from Time 2 as an additional predictor. Even with these two variables included as predictors, the Group AMP continued to predict hiring decisions although the effect became marginal,  $B = -.51, SE = .27, p = .06, OR = .60$ . No other predictor in this model reached significance. Together, these results are consistent with the hypothesis that the Group AMP has strong predictive validity.

## Discussion

At Time 3, we hypothesized that implicit bias as assessed with the Group AMP at Time 2 would predict consequential racial biases in a hiring decision 1 to 2 weeks later better than an Individual AMP. Results supported our hypothesis. In particular, when both the Group AMP and the Individual AMP were used to predict hiring decisions, the Group AMP significantly predicted racial biases in hiring, whereas the Individual AMP did not. Those who had greater implicit bias toward Black Americans as measured by the Group AMP were less likely to hire the Black candidate. Furthermore, the Group AMP continued to predict hiring decisions when we added the Time 2 measures of explicit prejudice and internal motivations to control prejudice as a simultaneous predictor (although the effect became marginal). Thus, results are consistent with the conclusion that the Group AMP provides a valuable tool for predicting racially biased behavior. The significant predictive validity of the Group AMP is particularly

impressive given that our predictors were taken weeks earlier than the behavioral outcome.

## General Discussion

In the present article, we propose that groups are perceived as more representative of their racial category and thus activate prejudice associated with that race more effectively than individuals. As a result, we hypothesize that using images of groups rather than images of individuals within an implicit priming measure would increase the validity and reliability of the measure. This reasoning is consistent with models of impression formation that indicate that category information (or in this case, attitudes toward the category) are most likely to be applied to targets who are perceived as fitting the category (Brewer, 1988; Fiske & Neuberg, 1990). First, we assessed whether groups are perceived as more representative of their racial categories than individuals. Consistent with our hypothesis, we found that people rated images of Black and White groups to be more representative and typical of Black and White Americans in general than were Black and White individuals. These findings are particularly interesting, given that the Black and White individuals we used were selected because they were pretested (in other research from our lab) to be most representative of their racial categories. Next, we tested whether a Group AMP would display better psychometric properties than the Individual AMP across three time points using the same images.

First, at Time 1 and Time 2, we tested and compared the convergent validity of the Group AMP and Individual AMP with explicit measures of prejudice. We reasoned that if images of Black groups activate category beliefs about Black Americans more effectively than images of Black individuals, then bias scores on a Group AMP should predict unique variability in bias scores on explicit measures that assess attitudes toward Black Americans. Indeed, the Group AMP predicted variability in explicit prejudice above and beyond the Individual AMP at both time points. This finding also relates to an interesting way in which standard implicit measures and explicit measures of racial prejudice tend to differ that may unintentionally decrease their correspondence. In particular, unlike commonly used implicit measures that evaluate attitudes toward individual exemplars, explicit measures tend to assess beliefs about the social category as a whole (e.g., "How warm do you feel toward Black Americans?"). Thus, the Group AMP may increase implicit-explicit attitude correspondence by making the measures more structurally equivalent (Payne, Burkley, & Stokes, 2008).

At Time 2, we also compared the test-retest reliability of the Group AMP and the Individual AMP. To our knowledge, this is some of the first data to address the test-retest reliability of the AMP. We found that, overall, the AMP was reliable. This was impressive given that 3 to 4 weeks passed between measurements. The estimate of test-retest reliability for the Individual AMP ( $r = .50$ ) was quite strong. However, most critical to the present hypotheses, the test-retest reliability for

the Group AMP ( $r = .66$ ) was significantly stronger. As predicted, using images of groups in the implicit measure led to greater consistency in responding across time—a finding supportive of the idea that the Group AMP reduces measurement error.

Importantly, there are two possible reasons that the Group AMP and the Individual AMP may be diverging in their psychometric properties. One possibility is that the two AMPs are measuring the same construct (implicit race bias), but that the Group AMP is doing so better. This is our hypothesis. However, another possibility is that the Group AMP and Individual AMP are assessing distinct constructs. To compare these two explanations of our findings, we next used structural equation modeling to assess whether our Time 1 and Time 2 measures of implicit bias were better captured by a one-factor model or a two-factor model. Results indicated that a one-factor model fit the data well and that a two-factor model did not provide significant improvements in model fit. Finally, the factor loadings for the Group AMP in the one-factor model were much higher than those of the Individual AMP, indicating that the Group AMP is a more valid indicator of the latent factor.

At Time 3, we evaluated whether implicit prejudice as assessed with the Group AMP at Time 2 (taken 1-2 weeks prior) was able to predict racial discrimination in hiring above and beyond the Individual AMP. Consistent with hypotheses, the Group AMP significantly predicted racial biases in hiring decisions such that greater implicit race bias on the Group AMP led participants to be less likely to hire a Black candidate. The fact that the Group AMP was able to predict variability in hiring above and beyond the Individual AMP is impressive given the large amount of variability shared between the Individual AMP and Group AMP ( $r = .65$ ). Together, results across all three time points are strongly consistent with our hypotheses that groups may be an excellent tool for measuring intergroup prejudice.

Importantly, although we tested our hypotheses using images of Black and White groups, we expect that the current findings would generalize to the measurement of prejudice for any social category. For example, we suspect that a group of Asian American men would activate beliefs associated with Asian American men to a greater degree than an individual Asian American man; likewise, a group of Black American women should activate beliefs about Black American women to a greater degree than an individual Black American woman, and so on. We focused on Black American men in the present analyses merely because many race stereotypes of Black Americans are often those that are used to describe Black American men (Carpinella, Chen, Hamilton, & Johnson, 2015). Thus, a group of Black American men should activate these beliefs more readily.

## Conclusion

Although individuals certainly embody group memberships, they can be categorized in many different ways. These myriad

categorizations are often driven by many different visual cues that vary in their salience based on the context and motivations of the perceiver. As a result, using images of individuals to assess social category beliefs may be problematic when a measure aims to assess attitudes toward a single social category. In contrast, groups of people that share a salient social identity such as race should be easily and automatically categorized according to that shared group membership. Such categorization should facilitate the activation of associated stereotypes and prejudice and provide a window into people's social category beliefs. Given that prejudice exists because of group memberships, it is perhaps unsurprising that groups may also serve as an excellent tool for measuring prejudice.

### Authors' Note

The authors have reported all conditions, data exclusions, and how they determined their sample sizes. All measures used are reported in the article.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Supplemental Material

The online supplemental material is available at <http://pspb.sagepub.com/supplemental>.

### Note

1. The pattern of results does not change for any of our analyses if we treat these two scales as separate indicators of explicit prejudice

### References

- Allport, G. W. (1954). *The nature of prejudice*. Oxford, UK: Addison-Wesley.
- Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin*, 38, 1194-1208.
- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15, 674-679.
- Bluemke, M., & Fries, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, 42, 163-176.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models* (Vol. 154). Newbury Park, CA: SAGE.
- 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, 330-350.
- Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin*, 41(3), 405-419.
- Cooley, E., & Payne, B. K. (2016). *Putting the group back into intergroup bias*. Unpublished manuscript.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163-170.
- De Houwer, J., Hermans, D., & Spruyt, A. (2001). Affective priming of pronunciation responses: Effects of target degradation. *Journal of Experimental Social Psychology*, 37, 85-91.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- 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, 876-893.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013-1027.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum 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.
- Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2016). *Relative stability of implicit and explicit measures: A longitudinal analysis*. Unpublished manuscript.
- Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by redefining the category labels. *Journal of Experimental Social Psychology*, 40, 357-365.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17-41.
- Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *Journal of Experimental Social Psychology*, 48, 892-898.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23, 253-283.
- 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.
- Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in decision to shoot. *Journal of Experimental Social Psychology*, 10, 1016-1022.

- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130.
- Maddox, K. B. (2004). Perspectives on racial phenotypicity bias. *Personality and Social Psychology Review, 8*, 383-401.
- Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630-633.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455-469.
- Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87*, 817-831.
- Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636-639.
- Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16-31.
- 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*, 277-293.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. *Measures of Personality and Social Psychological Attitudes, 1*(3), 1-16.
- Srull, T. K., & Wyer, R. S. (1988). *A dual process model of impression formation*. Hillsdale, NJ: L. Erlbaum Associates.
- Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A. G., & Swanson, J. (2002). Subgroup prejudice based on skin color among Hispanics in the United States and Latin America. *Social Cognition, 20*, 198-225.