The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes

Elena V. Stepanova *, Michael J Strube

Department of Psychology, Campus Box 1125, One Brookings Drive, Washington University, Saint Louis, MO 63130, USA

A R T I C L E   I N F O

Article history:
Received 19 May 2011
Revised 10 December 2011
Available online xxxx

Keywords:
Racial categorization
Facial typicality
Implicit racial attitude
Skin color
Facial physiognomy

A B S T R A C T

Previous research has not sufficiently addressed factors that define and moderate racial categorization judgments. This study independently manipulated skin color and facial physiognomy to determine their relative weighting in racial categorization. Participants (N = 250) judged faces varying on 10 levels of facial physiognomy (from Afrocentric to Eurocentric) and 10 levels of skin color (from dark to light) under either no time constraints, a modest time constraint, and under a stringent time constraint. Skin color was a powerful predictor of racial typicality ratings at all levels of facial physiognomy, but participants relied upon facial physiognomy more when rating faces of light than dark skin color. Skin color was a more important cue than facial physiognomy under no time constraints, but as time constraints became more severe, skin color’s importance decreased, yet it remained a more important cue at extreme physiognomy levels. The relationship between skin color and racial typicality ratings was stronger for those with more negative implicit racial attitudes. These findings suggest the primary role of skin color in racial categorization and underscore the importance of implicit attitudes in explicit categorization judgments.

© 2012 Elsevier Inc. All rights reserved.

Introduction

In studies of prejudice and stereotyping, participants are frequently presented with African American and European American faces as stimuli intended to activate a racial concept. This practice is often carried out by researchers with seemingly little regard for the specific features of the facial stimuli that define them as African American or European American (for example). This oversight is problematic given the considerable variability in both skin tone/color and facial features within given racial and ethnic categories.

Research has found that the more that racial images represent typical Africentric phenotypic characteristics, the more negative are evaluations whether assessed by traditional explicit measures or by more automatic implicit measures (for a review, see Maddox, 2004). Despite the general consistency of these racial effects, the factors that define and moderate racial categorization have largely been ignored in categorization and impression formation research. It seems likely that some cues (e.g., skin color) are more important than others (e.g., facial features), and that various characteristics of the situation (e.g., time demand) and the perceiver (e.g., racial attitudes) affect weighting of these cues in racial categorization.

Within-group variability

The activation of “race” in studies of racial prejudice and stereotyping is often achieved with pictures (e.g., African American or European American faces). Researchers generally try to construct or select facial stimuli that clearly represent the racial categories of interest. The use of prototypical stimuli, however, belies the fact that faces vary along many dimensions in often subtle ways. For example, human skin color and facial physiognomy vary along a wide continuum, even within any one racial or ethnic group (Farkas et al., 2005; Jablonski, 2004; Parra, 2007). Such within-group variability is important to consider because the categorization of less typical group members may be especially sensitive to situational factors (e.g., cognitive resources) and individual differences (e.g., implicit racial prejudices).

Skin color and facial physiognomy

Skin color is usually assumed to be of primary salience in defining racial categorization judgments about non-White groups (for review, see Maddox, 2004). When White participants are asked to give verbal descriptions of a Black face, they tend to mention darkness of the face, kinkiness of the hair, and brown eyes more than other features.
(Deregowski, Ellis, & Shepherd, 1975). When White participants are explicitly asked what features are important in racial categorization, they rate skin color as the most important criterion in the categorization of Black targets (Brown, Dane, & Durham, 1998). Interestingly, pre-categorized African American faces are perceived to be darker-skinned than European American faces even when their actual skin color is identical (see Levin & Banaji, 2006; MacLin & Malpass, 2001; MacLin & Malpass, 2003).

In contrast, the role of facial physiognomy has rarely been addressed as an independent factor contributing to racial typicality judgments (but see Gitter & Satow, 1969). Previous research has shown that face pigmentation and shape independently contribute to face recognition (Russell & Sinha, 2007; Russell, Sinha, Biederman, & Nederhouser, 2006), suggesting that they might also contribute independently to categorization. In one early study, Gitter and Satow (1969) manipulated physiognomy and skin color separately, albeit using dolls presented as stimuli in a study of racial misidentification in children. They found that physiognomy and skin color of the stimuli were independent factors in children’s judgments of their own racial identification. More recently, Stepanova and Strube (2009) independently manipulated facial physiognomy, skin tone and stimulus presentation mode (grayscale vs. color) of computer generated faces. Consistent with Gitter and Satow’s (1969) findings, both skin tone and facial physiognomy contributed independently to racial typicality evaluations. However, this research used a limited number of levels of both skin color (3) and facial physiognomy (2). It may be that with finer manipulations of these two factors, they might interact to determine racial typicality ratings, suggesting that people rely upon a combination of visual cues in racial typicality judgments. For example, dark skin color might be used as a salient and sufficient marker for racial categorization; when it is present, faces may not require further processing, as they can be easily identified as non-White. When skin color is light, however, additional cues might also be examined (e.g., facial physiognomy) to determine if a face can be classified as White.

**Moderators of racial categorization**

**Explicit measures of prejudice**

Racial and ethnic categorization research has explored moderation by explicit measures of prejudice with some success. For example, Castano, Yzerbyt, Bourguignon, and Seron (2002) asked participants to categorize morphed faces as northern Italian or southern Italian when the target was composed of a varying percentage of a northern African and a northern European face. Consistent with the *in-group overexclusion hypothesis* (see Leyens & Yzerbyt, 1992), those participants who highly identified themselves with northern Italians (in-group) classified more target faces as out-group members (southern Italians) in comparison to those who did not highly identify with an in-group.

Other studies have investigated racial categorization of ambiguous faces among South Africans and White Americans (Blascovich, Wyer, Swart, & Kibler, 1997; Pettigrew, Allport & Barnett, 1958). Although Pettigrew and colleagues did not explicitly measure racial attitudes of White Afrikaners, they assumed this ethnic group to be prejudiced against others (“Colored,” Indians and Africans). Their findings indicated that White Afrikaners, when presented with racially mixed photographs, tended to include ambiguous faces in the extreme “African” group rather than in intermediate “Indian” or “Colored” groups. Blascovich et al. (1997) found that highly prejudiced individuals took longer to categorize racially ambiguous faces. These results suggest that to protect their identity, perceivers are very careful about whom to include in their in-group. The same faces can be judged quite differently depending on the racial attitudes of the perceiver, and this variation is likely largest when facial cues are mixed or ambiguous.

**Implicit measures of prejudice**

Explicit measures of racial attitudes do not always moderate racial categorization judgments (Stepanova, Strube, Yablonsky, Pehson, & Shuman, 2008), presumably because explicitly stated attitudes are under conscious control and their levels could be influenced by a number of motivational and situational factors. By contrast, implicit measures may be more sensitive indicators of underlying attitudes and offer an alternative approach to moderation of racial categorization. However, recent research has only begun to explore implicit moderators of racial categorization based on facial cues. For example, Hutchings and Haddock (2008) and Hugenberg and Bodenhausen (2003) found that participants with greater implicit racial prejudice were more likely to categorize angry (but not happy or neutral) ambiguous-race faces as Black. These findings suggest that racial judgments might be moderated by implicit attitudes, especially when facial cues are not extreme (e.g., cannot be classified as clearly indicating one racial group).

**Cognitive control and availability of cognitive resources**

When cognitive resources are limited (e.g., under time constraints or cognitive load) people are unable to exert as much control over their judgments (Bodenhausen, 1990; Govorun & Payne, 2006; Richeson et al., 2003) and responses then are more likely to be based on automatic associations. When participants are categorizing faces, lack of cognitive resources might fundamentally change how facial cues influence racial categorization. With fewer cognitive resources available, more weight is likely to be given to salient features such as skin color, which require less processing, than to subtler facial details that require more processing. Moreover, when cognitive resources are limited, individual differences in levels of implicit and explicit prejudice might also produce quite different outcomes for racial typicality evaluations, with implicit attitudes perhaps playing a more prominent role. Therefore, another important purpose of this research was to investigate how availability of cognitive resources affects racial typicality evaluations.

**The current research**

The current research examined skin color and facial physiognomy in greater detail than has been accomplished in previous research and examined individual differences and situational moderators that may alter the weighting of these facial features in determining judgments of racial typicality. We examined the following questions in this research: (a) What are the additive and interactive influences of skin color and facial physiognomy on racial typicality judgments when both skin color and facial physiognomy are varied independently in gradual increments? (b) Are those effects moderated by implicit and explicit racial attitudes?; and (c) Do time constraints alter the attention paid to (or weighting of) skin color and facial physiognomy?

**Experiment overview**

Participants were presented with 100 computer-generated faces varying in skin color and facial physiognomy. Each face was rated on a 6-point scale ranging from very African American to very Caucasian. The faces were rated three times, under varying levels of time constraint. Participants’ implicit and explicit racial attitudes were also assessed.

**Predictions**

**Hypothesis 1.** We expected an interaction between facial physiognomy and skin color. Darker faces were expected to be rated consistently as African American with little influence from facial physiognomy. Lighter colored faces, however, were expected to receive more varied racial typicality judgments and to depend to a greater extent on facial...
physiognomy. Dark skin color might be a salient racial marker that allows participants to confidently place dark faces in the extreme, very African American group, whereas light skin color is a less salient racial marker (i.e., it is shared by both African American and Caucasian groups) that requires reliance on other cues (i.e., facial physiognomy) to determine racial typicality.

**Hypothesis 2.** We expected that racial attitudes would moderate the effects of skin color and facial physiognomy on racial typicality ratings. We predicted that for participants with high racial prejudice, the pattern described in Hypothesis 1 would be even more prominent, with stronger reliance on darker skin color to categorize faces. A salient racial marker (dark skin color) that allows an accurate racial categorization about an out-group member should be particularly more important for those who have negative attitudes about that out-group. Because previous research has sometimes found moderation of racial categorization judgments by explicit attitudes and sometimes by implicit attitudes, we anticipated that either type of attitude would moderate categorization.

**Hypothesis 3.** We expected that, compared to no-time-constraint trials, time constraints (speeded categorization) would produce categorization decisions that would be more affected by skin color than by facial physiognomy (e.g., Response Deadline × Skin Color × Facial Physiognomy interaction). With fewer cognitive resources available under time constraints, more weight is likely to be given to salient features such as skin color that requires less processing, than to subtler facial physiognomy that requires more processing.

**Hypothesis 4.** We also expected to find a Response Deadline × Skin Color × Facial Physiognomy × Implicit Racial Attitudes interaction. Under time constraints, the effect described in Hypothesis 3 should be even more prominent in participants with higher implicit racial prejudice than in participants with lower implicit racial prejudice. When cognitive resources are limited, participants might not be able to rely upon their explicit pre-existing attitudes, but implicit attitudes should still be accessible.

Past research has not extensively examined racial categorization among non-White participants. Although not designed for this purpose, our sample did have sufficient numbers of non-White participants to provide some exploration of this potential moderator. Accordingly we report how our major hypotheses are qualified by participant ethnicity.

**Methods**

**Participants**

Undergraduates (N = 207) from a large private Midwestern university were recruited through standard subject pool procedures along with participants (N = 49) from the general population. Participants recruited from the general population received $10 for their participation; the remaining participants received course credit. Data from six participants were excluded from the analysis because they did not complete the experiment or were under 18, resulting in 250 participants (48 from the general population and 202 from the student population). The final sample included 137 (54.8%) European American, 39 (15.6%) African American, 4 (1.6%) Hispanic American, 47 (18.8%) Asian American, and 23 (9.2%) “Other.” Participants’ mean age was 21.92 years (SD = 7.28) with a range of 18–62 years. Sixty percent of the final sample was female.

**Facial stimuli**

Facial stimuli were created with Poser 6™ software. The faces were designed to be equivalent for affective expressions but to vary systematically in skin color and facial physiognomy (see Fig. 1 for sample stimuli). Skin color varied from very light to very dark (10 levels) and facial physiognomy varied from very Afrocentric to very Eurocentric (10 levels). The facial physiognomy manipulation encompassed several phenotypic characteristics (e.g., width of the nose, fullness of the lips, bone structure, etc.). A set of those characteristics was manipulated simultaneously in the Poser 6™ software using a control that globally modified the faces to make them “less/more African” (for European faces) or “less/more European” (for African faces). Two sets of stimuli were used to insure generalizability, with each set consisting of 100 faces.3

**Racial typicality ratings task**

Participants were asked to rate 100 facial stimuli on the dimension of racial typicality (a continuum of Afrocentric–Eurocentric typicality) using a 6-point scale: 1 (Very African American), 2 (Moderately African American), 3 (Somewhat African American), 4 (Somewhat Caucasian), 5 (Moderately Caucasian) and 6 (Very Caucasian). The choice of the one-dimensional typicality scale is substantiated by previous research.4 Participants rated faces presented in a random order. The particular

---

3 A separate sample of participants (N = 51 for set 1 and N = 58 for set 2) completed two rank-ordering tasks in order to determine if the faces could be reliably distinguished along the manipulated dimensions. They were instructed to rank-order faces on the basis of (a) skin tone and (b) facial physiognomy. On each trial, they were presented with a row of faces (for skin judgments) or a column of faces (for facial physiognomy judgments) (see Fig. 1). The column or row a participant saw on each trial was randomly determined.

4 To determine if faces could be distinguished along the two manipulated dimensions, we performed analyses for skin color ranks when data were collapsed across facial physiognomy levels and for facial physiognomy ranks when data were collapsed across skin color levels. There was a statistically significant difference in rankings of physiognomy for Set 1, χ²(9) = 445.1, p < .001 and for Set 2, χ²(9) = 353.28, p < .001. There was a statistically significant difference in rankings of skin tone for Set 1, χ²(9) = 352.52, p < .001 and for Set 2, χ²(9) = 263.63, p < .001. Each level of skin tone was ranked differently from adjacent ones (for both sets), and most levels of facial physiognomy were ranked differently from adjacent ones (except for levels 1 and 2, 7 and 8, and 9 and 10 for Set 1 and levels 2 and 3 and 9 and 10 for Set 2), as indicated by significant follow-up Wilcoxon signed ranked tests with Bonferroni adjustments. All higher levels of physiognomy received higher mean ranks (1.75, 2.24, 3.14, 4.00, 5.16, 6.20, 7.02, 7.76, 8.63, and 9.12 for Set 1 and 1.40, 2.34, 2.83, 4.12, 4.97, 6.21, 7.12, 7.93, 8.84, 9.24 for Set 2). All higher skin levels always received higher mean ranks (2.08, 2.82, 3.61, 4.39, 5.08, 5.96, 6.63, 7.51, 8.02, 8.90 for Set 1 and 1.53, 2.48, 3.33, 4.29, 5.02, 6.03, 6.76, 7.78, 8.26, 9.52 for Set 2). These data indicate that participants were able to distinguish the incremental differences in both skin tone and facial physiognomy.

In order to insure that dark skin color did not physically obscure variations in facial physiognomy, we analyzed the physical physiognomy ranks data after collapsing

---

set of faces (from 2 sets) presented to each participant in the racial
typicality task was randomly determined.

The racial typicality ratings task consisted of three separate blocks. In
each block, participants rated the same set of 100 faces. In the first
block, participants performed the task without any response dead-
line. Participants were told to rate faces according to how representa-
tive they were of two racial groups (i.e., Caucasian and African
American) after careful examination. They were instructed to use
any standards they liked when making this judgment, to take as
much time as needed to make their judgments, and not to rush
through to get done quickly.

Most cognitive categorization research has employed only dichot-
omous judgments and short response deadlines when studying
categorization under time pressure (600–1600 ms, Lamberts, 1995;
Lamberts, 2000). Because we used a multiple-category response
scale, it was not initially clear what response deadline we should em-
ploy to limit cognitive resources. To determine response deadlines,
we examined how fast participants performed the racial typicality
rating task without time pressure in the first block. In the second
block, each participant’s median response time from the first block
was used as their deadline for the second block. In the third block,
each participant’s 25th percentile reaction time from the first block
was used as their deadline. Therefore, during Block 2 we required
participants to respond more quickly than they did on 50% of their
no-deadline trials, and during Block 3 participants had to respond
more quickly than on 75% of their no-deadline trials (as suggested
by K. Lamberts, personal communication, February 25, 2009). Instruc-
tions were also modified in the second and third blocks. Participants
were told to make a decision about each face as fast as possible be-
cause they would be given a limited time to rate each face; if they
took longer than allowed, they would receive a warning (“Too
Slow! Go Faster!”).

In all three blocks, after a participant made a decision on each trail,
there was a blank screen and brief interval of 1.5 s between displays
of faces, so that when two adjacent faces were similar, a participant
would know that a new face was displayed. Also, in all three blocks,
if participants responded in less than 250 ms, a warning was issued:
“You are responding too quickly to be giving any thought to your
judgments. Please take enough time to provide a careful judgment.
Press OK to continue.” After the participant pressed the “Ok” button
to clear the warning message, a blank screen appeared for 1.5 s and
then the same face was shown again.

Procedure

The study was conducted on computers and took approximately
1 h. Half of the participants performed the racial typicality ratings of
the faces first, and then the IAT (Greenwald, McGhee & Schwartz,
1998; Greenwald, Nosek & Banaji, 2003). The other half of the partic-
ipants performed the IAT first, and then the racial typicality ratings.
The IAT score was derived according to the procedures described by
Greenwald et al. (2003) and represents a standardized response time
difference. Higher scores indicate more favorable implicit attitudes to-
ward Whites compared to Blacks. Facial stimuli employed in the IAT
were a subset of faces used during the racial typicality ratings task
and included 5 faces with high Afrocentric physiognomy and dark
skin color and 5 faces with high Eurocentric physiognomy and light
skin color. After completing the IAT and racial typicality ratings, partic-
ipants were asked to complete the explicit racial attitude measures:
Symbolic Racism Scale (Henry & Sears, 2002), Social Dominance

Please cite this article as: Stepanova, E.V., & Strube, M.J, The role of skin color and facial physiognomy in racial categorization: Moderation by
Orientation Scale (Pratto, Sidanius, Stallworth, & Malle, 1994), Modern Racism Scale (McConahay, 1986), and Feeling Thermometers. For the Feeling Thermometers, participants were asked to indicate how favorably they viewed different social and ethnic groups on a scale of 1 (very cold or unfavorable) to 7 (very warm or favorable). These questionnaires were randomly ordered for each participant. The explicit measures were collected last so that they would not sensitize participants or produce inadvertent priming effects. At the end of the study, participants were asked to provide demographic information.

Results

First, we examined the typicality ratings in a 10 (Skin Color) × 10 (Facial Physiognomy) × 3 (Response Deadline) × Explicit Racial Attitudes × Implicit Racial Attitudes repeated measures multiple regression, with the last two factors treated as between-subjects continuous variables. Although some interactions involving Explicit Racial Attitudes were significant, none were of substantive interest and no predicted effects or interactions emerged. To simplify the presentation that follows, we excluded explicit measures and report results from a 10 (Skin Color) × 10 (Facial Physiognomy) × 3 (Response Deadline) × Implicit Racial Attitudes repeated measures multiple regression.

Second, we explored the additional moderating influence of participant ethnicity by including ethnicity as a between-subjects continuous variable. Although some interactions involving Explicit Racial Attitudes were significant, none were of substantive interest and no predicted effects or interactions emerged. To simplify the presentation that follows, we excluded explicit measures and report results from a 10 (Skin Color) × 10 (Facial Physiognomy) × 3 (Response Deadline) × Implicit Racial Attitudes repeated measures multiple regression.

The assumption of sphericity was violated in our analyses (an unsurprising result given the nature of the data), so we used the Greenhouse–Geisser corrected degrees of freedom when determining statistical significance. All follow-up comparisons were conducted with Bonferroni corrections.

Tests of stated hypotheses

The analysis revealed significant main effects for Skin Color, Greenhouse–Geisser $F(1.46, 361.56) = 230.46, \text{MSE} = 12223.87, p < .001,$ $\eta^2 = .48,$ and for Facial Physiognomy, Greenhouse–Geisser $F(1.34, 331.31) = 248.26, \text{MSE} = 12963.19, p < .001, \eta^2_g = .50.$ Unsurprisingly, darker faces were rated as more African American than lighter faces, and, faces with Afrocentric physiognomy were rated as more African American than faces with Eurocentric facial physiognomy (see Fig. 2). Pair-wise comparisons indicated that mean ratings for each level of skin color were significantly different from mean ratings of every other level of skin color; likewise, mean ratings for each level of facial physiognomy were significantly different from mean ratings of every other level of facial physiognomy (all $p < .05,$ Bonferroni adjustment). These effects verify the effectiveness of the manipulations in producing clear differences along the skin tone and facial physiognomy dimensions.

As predicted by Hypothesis 1, a Skin Color × Facial Physiognomy interaction emerged, Greenhouse–Geisser $F(48.61, 12053.93) = 7.73, \text{MSE} = 7.09, p < .001, \eta^2_g = .03$ (see Fig. 2). Follow-up comparisons revealed that skin color was an important cue for racial typicality judgments within each level of physiognomy (all simple effects of skin color, $p < .05$) and that physiognomy was also an important cue for racial typicality judgments within each level of skin color (all simple effects of facial physiognomy, $p < .05$). More revealing are the proportions of variance that each cue accounts for at levels of the other cue. Fig. 3 illustrates the proportions of the variance explained ($\eta^2_g$) for simple effects of skin color within each level of physiognomy, and, the proportions of the variance explained for simple effects of physiognomy within each level of skin color. Readily apparent in this figure is that proportions of variance show more variability for facial physiognomy than for skin color and that the smallest effect sizes for physiognomy occur for darker skin colors. In other words, although both skin color and facial physiognomy are powerful cues, skin color appears to be more uniformly important.

Hypothesis 2 predicted that racial attitudes would moderate the effects of skin color and facial physiognomy. There was a significant Skin Color × Implicit Racial Attitudes interaction, Greenhouse–Geisser $F(1.46, 361.56) = 3.98, \text{MSE} = 210.85, p = .03, \eta^2_g = .02,$ which indicated that participants with greater implicit racial prejudice rated faces as less Caucasian (and more African American) than participants with lower implicit racial attitudes for darker levels of skin color; that pattern reversed for lighter levels of skin color (see Fig. 4). Stated differently, skin color was more strongly related to typicality ratings for participants with higher implicit racial prejudice than for participants...
Physiognomy interaction, Greenhouse condition, there was also a significant interaction. Greenhouse–Geisser $F(52.02, 12899.95) = 3.37, p < .001, \eta^2_g = .013$, although it was less significant and accounted for less variance than in the No Response Deadline condition. In the 25th Percentile Response Deadline condition, there was also a significant Skin Color × Facial Physiognomy interaction, $F(52.63, 13050.95) = 2.94, p < .001, \eta^2_F = .012$ of a magnitude similar to the Median Response Deadline condition. The separate Skin Color × Facial Physiognomy interactions for each Response Deadline condition are illustrated in Fig. 5.

To further explore this interaction, we examined the pair-wise comparisons of each physiognomy level within each level of skin color and each skin color level within each level of physiognomy in each response deadline condition (see Fig. 6 for proportion of significant pair-wise comparisons among skin color within physiognomy levels and among physiognomy within skin color levels for each of the response deadline conditions). Of note are the following patterns: (a) skin color was a more important cue than facial physiognomy under no time constraints, as evidenced by the higher proportion of significant pair-wise comparisons for skin color within physiognomy levels as compared to the proportion of significant pair-wise comparisons of physiognomy within skin color levels, (b) as time constraints became more severe, skin color became a less important cue, as evidenced by the lower proportion of significant pair-wise comparisons among skin color levels in the Median Response Deadline and 25th Percentile Response Deadline than in the No Response Deadline condition, and (c) yet it remained a more important cue than physiognomy at all but one intermediate level of physiognomy in the Median Response Deadline condition and a more important cue than physiognomy when levels of the other cue were extreme in the 25th Percentile Response Deadline condition.

We also used the same approach described previously and examined the proportions of variance accounted for by skin color within levels of facial physiognomy, and the proportions of variance accounted for by facial physiognomy within levels of skin color. Fig. 7 illustrates these proportions separately by time constraint condition. The pattern is rather complex, but one feature is notable: The effect of time constraint is most obvious when skin color is dark or facial features are African American. When either of these dimensions is at its extreme, the other dimension accounts for less variance when there is a time constraint on the decision. This is less likely to occur when faces are light or clearly European American. Overall, the total
amounts of variance explained and a proportion of significant pairwise comparisons suggest that skin color was a more important cue than facial physiognomy under time constraints.

Hypothesis 4 predicted that, under time constraints, the effect described in Hypothesis 3 would be even more prominent in participants with higher implicit racial prejudice than in participants with lower implicit racial prejudice. However, the Response Deadline × Skin Color × Facial Physiognomy × Implicit Racial Attitudes interaction did not achieve significance, Greenhouse–Geisser $F(82.73, 20516.64) = 1.17, p = .14$.

Exploratory analyses: ethnicity as moderator

Because we were able to recruit a number of non-White participants, we examined participants’ ethnicity as a between-participants factor in an exploratory analysis. There was a limited number of Hispanic participants ($N = 4$) so we excluded that group from the analyses. Data from the following ethnic groups were included in the analysis: European Americans ($N = 127$), African Americans ($N = 39$), Asian Americans ($N = 47$), and Other ($N = 23$).

Racial typicality ratings were analyzed in a $10 \times 10 \times 3 \times 4$ (Facial Physiognomy) × (Response Deadline Condition) × (Implicit Racial Attitudes) × (Participants’ Ethnicity) repeated measures multiple regression that mirrored the previously reported tests for the stated hypotheses. Participants’ Ethnicity was treated as a categorical between-participants factor. To examine the interaction of Implicit Racial Attitudes and Participants’ Ethnicity, their product was entered in the model. Because of the duplication of effects from previously described analyses, only new effects involving Participant Ethnicity are described. Because these analyses are exploratory, we considered only effects significant at $p < .01$ and interactions only up to third-order inclusively.

A significant Skin Color × Participants’ Ethnicity interaction emerged, Greenhouse–Geisser $F(4.48, 360.37) = 4.87, p < .001, \eta^2_p = .06$. Follow-up tests examined ethnicity group differences within each skin color level.

Significant differences were found at skin color levels 1, 2, 3, 4, 5, and 9 (all $F$s > 4.87, ps < .018). Pairwise comparisons showed that in the darker skin color levels (1–5), African American participants rated faces as more Caucasian (less African American) than did European American and Asian American participants. However, that pattern reversed for lighter skin colors (9 and 10, although pairwise comparisons did not reach significance for Skin Color level 10) (see Fig. 8). Stated differently, the relationship between skin color and racial typicality ratings was stronger for European American and Asian American participants than for African American participants.

The analysis also revealed a significant Response Deadline × Physiognomy × Participants’ Ethnicity interaction, Greenhouse–Geisser $F(35.46, 2848.5) = 1.69, p = .006, \eta^2_p = .02$. We examined the Physiognomy × Participants’ Ethnicity interaction within each of the response deadline conditions. As Fig. 9 illustrates, the relationship of facial physiognomy to racial typicality ratings was more pronounced for African American participants than for European American and Asian American participants, and this effect became even stronger with more stringent response deadlines. Follow-up tests within response deadline conditions revealed significant Physiognomy × Participants’ Ethnicity interactions in each response deadline condition, but they were stronger in the two response deadline conditions than in the no-response deadline condition; Greenhouse–Geisser $F(5.54, 444.79) = 4.60, p < .001, \eta^2_p = .05$ for the No Response Deadline condition; Greenhouse–Geisser $F(5.26, 422.90) = 6.87, p < .001, \eta^2_p = .08$ for the Median Response Deadline condition; Greenhouse–Geisser $F(6.21, 498.95) = 6.45, p < .001, \eta^2_p = .07$ for the 25th Percentile Response Deadline condition.

Discussion

This research addressed the relative weighting of variability in two markers—skin color and facial physiognomy—in determining racial typicality judgments, as well as how individual (racial attitudes) and situational (time constraints) factors might moderate the effects...
Fig. 6. Proportion of significant pair-wise comparisons for skin color within levels of facial physiognomy and for facial physiognomy within skin color levels by time constraint condition. Facial physiognomy levels varied from Afrocentric (1) to Eurocentric (10) and skin color levels varied from dark (1) to light (10).

Fig. 7. Proportion of variance explained ($\eta^2$) for simple effects of skin color within levels of facial physiognomy and simple effects of facial physiognomy within skin color, by time constraint condition. Facial physiognomy levels varied from Afrocentric (1) to Eurocentric (10) and skin color levels varied from dark (1) to light (10).
of those markers. Exploratory analyses further examined the moderating role of participant ethnicity.

Consistent with Hypothesis 1, skin color and facial features interacted to influence racial typicality ratings (Fig. 2). The current findings suggest that facial features might be a more important judgment cue when skin color is intermediate and lighter than when it is very dark. Specifically, classification of faces into the African American category can be done largely on dark skin color alone. As faces become lighter, however, a simple skin color heuristic is no longer as useful and the additional information supplied by facial features is relied upon more heavily. Classification of faces into racial groups thus may follow a two-step process, with an initial evaluation of skin color that terminates in a classification into the African American category if the face is dark but that is followed by an evaluation of facial features if skin color is lighter. Additional evidence supporting this proposed two-step model of racial categorization is provided by recent social neuroscience research (Brebner, Krigolson, Handy, Quadflieg, & Turk, 2011). Brebner et al. (2011) employed event-related brain potentials (ERPs) to investigate temporal effects of racial categorization when skin color and facial physiognomy of faces were manipulated independently (each had two levels). Their analyses revealed that early ERP components associated with perceptual processing of faces (N170) and the individuation of facial exemplars (N250) were sensitive to skin color, whereas in the later ERP components (N200) a modulation on the basis of facial physiognomy was observed.

An important implication is that equally atypical faces are not treated the same way. That is, a very dark Eurocentric face is, objectively, as unusual as a very light Afrocentric face—at least in the mismatch of skin color and physiognomy. Yet, the former faces are not distinguished as much from very dark Afrocentric faces—all are viewed as African American faces—whereas light Afrocentric faces are more clearly distinguished from light Eurocentric faces. We can only speculate about the causes of the proposed two-step process. Perhaps the relative salience of skin color and facial physiognomy in visual processing determined the effects obtained. Further research is needed to examine this and other potential underlying causes (e.g., the extent to which variability in skin color is more naturalistic than variability in facial features in computer generated images, perceptual familiarity with certain markers, and reliance on certain cues to effectively include or exclude faces from an in-group).

Although the Skin Color × Facial Physiognomy interaction suggests that skin tone may play a primary role in racial classification, other evidence from this study suggests that reliance on skin tone and facial physiognomy varies by ethnicity. For example, White participants and Asian American participants relied more on skin tone than did African American participants (Fig. 8). On the other hand, the use of facial physiognomy was more pronounced for African American participants than for European American and Asian American participants, especially when decisions had to be made quickly (Fig. 9). These findings suggest
that skin color and facial physiognomy play somewhat different roles or are weighted differently in racial categorization for different ethnic groups. Skin color is a more salient out-group marker or diagnostic cue for European Americans and Asian Americans than for African Americans, whereas facial physiognomy is a more meaningful in-group marker or diagnostic cue for African Americans. There are several ethnic groups that include individuals with dark skin tone such as Hispanics, Asian Indians or American Indians. Therefore, skin color is not a very meaningful in-group marker for African Americans and they rely upon other cues such as facial features when making important in-group classification decisions. Perhaps, dark skin color and Eurocentric physiognomy are the category-diagnostic facial characteristics for categorization of other-race faces for European-Americans/Asian-Americans and African Americans correspondingly.

The interactive influence of skin color and facial features on racial categorization judgments also depended on availability of cognitive resources (see Figs. 5–7). Skin color was a more important cue than facial physiognomy under no time constraints, but as time constraints became more severe, the importance of skin color—but not facial physiognomy—declined. Yet, skin color was a more important cue for all but one intermediate level of physiognomy under moderate time constraints and for most extreme levels of physiognomy under severe time constraints. Also, as time constraints became more severe, if one of cues was extreme, a participant relied less on the other cue, and this was especially true for Afrocentric physiognomy and dark skin color. Yet skin color was a more important cue under time constraints, as it accounted for a bigger proportion of the variance when facial physiognomy was Afrocentric (as opposed to the proportion of the variance explained by facial physiognomy when skin color was dark). Skin color plays a primary role in categorizations judgments, and the relative weighting or importance of skin color and facial physiognomy is not fixed but varies with the processing demands imposed by the situation.

Collectively, these results are consistent with previous research indicating that facial cues associated with Black features are more salient than White features in social categorization (Smith & Zarate, 1992). This research also suggests that categorization depends upon the degree of Black features in a face (Freeman, Pauker, Apfelbaum, & Ambady, 2010). Importantly, however, skin color and facial physiognomy are not independently manipulated in previous work. The results of the current research thus offer an important qualification. Facial cues certainly are important in racial judgments, but their importance depends on the nature of other cues (skin color), racial identification of the perceivers, and on the degree to which the situation provides the opportunity to process those cues completely. Therefore, not all features are always salient, and some disproportionately affect racial categorization. These findings underscore the role of within-stimuli variability in racial categorizations.

Implicit racial attitudes as moderators of racial categorization

Hypothesis 2 predicted that racial attitudes would further moderate the impact of skin color and facial physiognomy on racial typicality judgments. Partial support was found for this prediction. Skin color was more related to racial categorization for individuals with greater implicit racial prejudice than for those with lower implicit racial prejudice (see Fig. 4). These results suggest that individuals with high implicit racial prejudice rely upon the most salient out-group marker in racial categorizations, skin color. Previous research (for review, see Maddox & Dukes, 2008) has established that dark skin color leads to negative attitudes; this study shows that implicit attitudes influence weighting of skin color in racial categorization. On the other hand, the influence of implicit attitudes on racial categorization was not qualified by the time constraints, suggesting that implicit attitudes influence such judgment regardless of limitations on cognitive resources. Prejudice and stereotyping research has postulated for a long time that social categorization is a pre-cursor to prejudice (Allport, 1954). But, it appears that implicit racial attitudes influence social categorization as well.

Overall, these findings are consistent with previous research indicating that the same faces can be judged quite differently depending upon racial attitudes of a perceiver (Blascovich et al., 1997; Castano et al., 2002; Pettigrew et al., 1958), especially when there is within-group variability. Our findings add an important caveat: different facial markers carry different weights for individuals with various levels of implicit—and not explicit—racial prejudice. Perhaps the “racial” nature of the task signaled to participants to control their responses, even when no response deadline was imposed; and when time constraints were imposed, participants could not rely upon their explicit attitudes, as they were not able to exert much control over their judgments that became even more automatic.9

Methodological and applied significance

This study has important methodological implications because research on racial prejudice and stereotyping depends critically on adequate choice of the facial stimuli used. As demonstrated by the variability in typicality judgments driven by skin color and facial physiognomy, the characteristics of facial stimuli do matter, and effect sizes in studies employing these stimuli might be influenced by the visual properties of facial racial stimuli. Specifically, a choice of extremely dark faces to represent African American primes would potentially produce stronger racial effects, especially so in those whose implicit racial attitudes are high, because they rely more than other groups on skin color in racial categorization.

The practical implications are even more important. Race-relevant decisions occur in many important contexts with considerable potential for bias. Determining the factors that drive those decisions has substantial applied importance, and the design of interventions will depend critically on an understanding of the underlying mechanisms. For example, recent research on cross-racial eyewitness identification has addressed how social categorization influences encoding and memory (e.g., Bernstein, Young, & Hugenberg, 2007; Hehman, Mania, & Gaertner, 2010; Hugenberg, Young, Bernstein, & Sacco, 2010; Pauker et al., 2009; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008; Young, Bernstein, & Hugenberg, 2010) and how the effects of social categorization on encoding are moderated by the racial context (Cassidy, Quinn, & Humphreys, 2011). While we did not explicitly address the effects of skin color and facial physiognomy on encoding and memory, our findings suggest that due to differences in categorization outcomes, these effects might emerge.

Furthermore, under some circumstances, individuals might be exposed to faces for a very brief period of time, which affects the process of categorization as well. Payne (2001) has discussed the case of Amadou Diallo, an African immigrant who was shot by New York City police officers who had mistaken his wallet for a gun; that research investigated stereotypic associations people have with African Americans and weapons and showed how easily misclassifications (e.g., of the wallet as a weapon) can occur and be moderated by ethnicity of targets. However, similar mistakes might be made for categorization of people into racial categories when a fast decision is required. We suggest that, under such conditions, the categorization of a target face might be altered, and factors that drive this categorization are weighed

---

8 We believe that our response deadline manipulation tapped into availability of cognitive resources rather than practice effects. In preliminary studies (e.g., Stepanova et al., 2008) when participants were asked to complete the racial typicality tasks twice without time constraints, no practice or order effects were observed.

9 We are aware that the IAT measures both automatic and controlled aspects of cognition on behavioral responses (see Sherman et al., 2008), and that implicit attitudes do not simply measure automatic associations. The distinction made between implicit and explicit attitudes in this work simply refers to types of measures employed.
differentially. For example, under time constraints, faces will be more likely to be categorized on the basis of skin color only, regardless of facial physiognomy. A person briefly exposed to a dark-skinned individual might weigh skin color more than facial features and report seeing “African American”—this might be even more prominent for individuals with high implicit racial prejudice and prompt especially high behavioral discrimination.

Limitations

This study explored racial categorization through explicit categorization judgments but does not allow inferences about the time course of social categorization. There is emerging research addressing the temporal effects of categorization in which participants’ hand movements en route to racial category alternatives are recorded by tracking the spatial coordinates of the mouse (see Freeman, Ambady, Rule, & Jonson, 2008; Freeman et al., 2010). Social neuroscience research also shows promise by employing ERPs to investigate temporal effects of racial categorization (Ito, Thompson, & Cacioppo, 2004; Ito & Urdan, 2005; Ito, Urdan, Willadsen-Jensen, & Corell, 2006; Kubota & Ito, 2007), including studies with racially ambiguous faces (Willadsen-Jensen & Ito, 2006) and a manipulation of skin color and facial physiognomy (Brebner et al., 2011). Such methodologies will allow investigating the temporal sequence of categorization and the role played by moderators at different stages.

Facial perception and recognition research has stressed that faces are processed holistically in a Gestalt-like manner (e.g., Farah, Wilson, Drain, & Tanaka, 1998; Goffaux & Rossion, 2006; Hole, 1994; Homa, Haver, & Schwartz, 1976; Sergent, 1984; Tanaka & Farah, 1993; Young, Hellawell, & Hay, 1987), and our emphasis on separation of skin color and facial features in racial categorization might seem somewhat artificial. However, experimental evidence suggests that other-race faces are perceived less holistically than own-race faces (e.g., Michel, Caldara, & Rossion, 2006; Michel, Rossion, Han, Chung, & Caldara, 2006). Likewise, new research investigating separate effects of skin color and facial features on modulation of neural responses (Balas & Nelson, 2010; Brebner et al., 2011), face recognition (Bar-Haim, Saidaè, & Yovel, 2009) and racial categorization (Stepanova & Strube, 2009) shows that separation of these two factors is justified.

The facial physiognomy manipulation in the current research did not center on any specific facial feature, rather it included a collection of facial features manipulated as a combination. We demonstrated that facial physiognomy has an important impact on categorization, but it will be important for future research to determine the particular facial features that are especially important. That task was beyond the scope of the present work. There are other procedural limitations as well. The stimuli that we employed are artificially constructed computer generated images. They undoubtedly deviate from real faces in some respects. However, initial questions about how the nature of facial stimuli affects racial judgments are best answered when the features of those stimuli can be carefully controlled and manipulated—the distinct advantage of the approach taken in this research. Nonetheless, it will be important for future research to verify our key findings using real faces. Similarly, this research only examined facial typicality judgments along an African-American-European American continuum. Perhaps racial typicality judgments for other groups (e.g., Asian-European continuum) do not follow the same pattern involving two types of cues—facial physiognomy and skin color. This area is a potential new direction for future research that can examine racial typicality judgments employing other racial groups and judgment tasks.

Concluding remarks

Despite these limitations, the questions that we attempted to answer here regarding the factors that are important for racial categorization are a key step to a greater understanding of how skin color and facial features can affect race perception. Given the substantial variability in the facial features and skin color that exists within racial groups outside the laboratory, this work is also an important step toward linking laboratory work to the settings in which research and theory should apply.

Role of the funding source

This work was partially funded by the 2008–2009 Psi Chi Graduate Research Grant awarded to the first author. Funding supported data collection.

Acknowledgments

This research was based on the first author’s doctoral dissertation from Washington University in Saint Louis, conducted under the supervision of the second author. We thank Randy Larsen, William Lowry, Michael Mueller, Simine Vazire, and Desireè White, who served on the dissertation committee for their helpful comments, and all staff members and volunteers from HealthStreet, the Center for Community-Based Research, at the Institute of Clinical and Translational Sciences of Washington University School of Medicine, for their assistance in participant recruitment. We also thank Bruce Bartholow for his critical reading of the manuscript.

References


