Social Cognition, Vol. 36, No. 6, 2018, pp. 603–625

RACE SIGNALING FEATURES: IDENTIFYING MARKERS OF RACIAL PROTOTYPICALITY AMONG ASIANS, BLACKS, LATINOS, AND WHITES

Debbie S. Ma
California State University, Northridge

Kolina Koltai
University of Texas at Austin

Ryan M. McManus and Adrian Bernhardt
California State University, Northridge

Joshua Correll
University of Colorado at Boulder

Bernd Wittenbrink
University of Chicago Booth School of Business

Racial prototypicality informs categorization and profoundly influences stereotyping and evaluative judgments. Although previous research has examined the physical facial features that correspond with racial prototypicality in Blacks and Whites, little research has investigated the features that predict prototypicality among Asians and Latinos. Using a large database of facial stimuli that included Asian, Black, Latino, and White faces, we modeled physical measures of the faces to identify features that influence subjective judgments of prototypicality for each of these groups with three different exploratory strategies. Our research provides a critical extension of previous research by including Asians and Latinos and seeks to test the replicability of the existing research on this topic for Blacks and Whites.

Keywords: racial prototypicality, racial categorization, face perception, face database, facial physiognomy

---

Support for this work was provided by National Science Foundation Continuing Grant 1226143 to the first author. We thank Brianna Herrera, Heather Rees, and Michelle Revels for their meticulousness in the exciting work of measuring faces.

Correspondence should be addressed to Debbie S. Ma, California State University, Northridge, Department of Psychology, 18111 Nordhoff Street, Northridge, CA 91330-8255; E-mail: Debbie.Ma@csun.edu.

Humans can quickly and deftly read complex social information from faces. After even very brief exposure to a face (50–100 milliseconds), humans can extract extraordinarily nuanced information about a person, including emotion, personality traits, intention, identity, and—most critically for the current article—social category membership (Ito, 2011). Social categorization describes the assignation of individuals to social groups, such as race, gender, and age, and underlies consequential psychological phenomena, including stereotyping and prejudice. By most accounts, social categorization is a necessary condition for stereotyping and prejudice, and some argue that social categorization is sufficient for eliciting stereotypes and prejudice (Allport, 1954; Lepore & Brown, 1997). Given the ubiquity and significance of social categorization, a great deal of attention has been dedicated to understanding the process of social categorization (Taylor, 1981; Zarate & Smith, 1990). Especially relevant for the current research is the role that physical features play in racial categorization. Research in this area finds (perhaps unsurprisingly) that exemplars who possess more race-signaling features are categorized as group members more quickly and imbued with group-level traits and evaluations to a greater extent (Blair, Judd, & Fallman, 2004; Blair, Judd, Sadler, & Jenkins, 2002; Kahn & Davies, 2011; Ma & Correll, 2011).

Unfortunately, researchers have primarily investigated Black and White racial categorization and social judgment and comparatively little research has examined social categorization of other racial or ethnic groups. We view the goal of making research more inclusive (racially and otherwise) as paramount to the health of the field as social psychology aims at building a more international community (Henrich, Heine, & Norenzayan, 2010). Worldwide, Asians comprise over 60% of the population and Latinos make up roughly 8% of the total population (Population Reference Bureau, 2014). The aim of the current article is thus to identify the features that signal category membership among Asians and Latinos,[1] while also offering a replication of previous research investigating Blacks and Whites.

FEATURES ASSOCIATED WITH RACIAL PROTOTYPICALITY

Researchers have employed a variety of methods to determine which features correspond with prototypicality. Blair and colleagues (2002), for example, asked participants to "make a single, global assessment of the degree to which each face had features that are typical of African Americans, using a scale from 1 (not at all) to 9 (very much)." Using this measure to assess prototypicality, Blair and Judd

---

[1] Throughout this article, we describe "Latino" as a racial category, but acknowledge that defining Latino as a racial category conflicts with conventional views of race (Snipp, 2010) and conflates race and ethnicity. For example, the United States Census includes Asian, Black, Native American and Alaskan Native, and White as racial categories and considers Latino an ethnicity: Hispanic or Non-Hispanic. However, many perceivers in the United States conceive of Latino as a racial category and commonly equate race and ethnicity (Cornell & Hartmann, 2007) and a recent study conducted by the Pew Research Center (2015) reported that two-thirds of Hispanic adults living in the United States considered their Hispanic ethnicity a part of their racial background. For these reasons, we adopt a folk understanding of race for this article, but we fully recognize that "Latino" refers to an ethnicity rather than racial category.

(2010) were able to identify features associated with Afrocentricity (i.e., Black prototypicality). In particular, they carried out extensive physical measurements of pictures of Black and White faces taken from published experiments (Blair, 2006; Blair, Chapleau, & Judd, 2005; Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Blair et al., 2002). Additionally, they asked research assistants to rate the images in terms of hair texture and volume. After obtaining participants' subjective Afrocentricity ratings of these targets, they regressed average ratings of Afrocentricity on the physical measures of the face. Using exploratory regression, they identified five features that predict subjective ratings of Black prototypicality. These include luminance, lip fullness, nose shape, hair texture, and hair quantity, whereby lower luminance, fuller lips, broader nose, kinkier hair, and more voluminous hair corresponded with Afrocentricity. In total, these attributes explained nearly 90% of the variability in subjective ratings of Black prototypicality. Although some researchers view prototypicality in terms of this constellation of features (see also Stepanova & Strube, 2012a), others have paid special attention to skin tone (Ellis, Deregowski, & Shepherd, 1975; Dunham, Stepanova, Dotsch, & Todorov, 2014; Maddox & Chase, 2004), whereby darker skin tone is associated with higher Black prototypicality and lighter skin tone relates to higher White prototypicality.

Although research has historically emphasized studying Black and White targets, recent research has begun exploring features and categorization in other groups. For example, Dunham, Dotsch, Clark, and Stepanova (2016), presented participants with computer-generated Asian and White faces to determine the extent to which individuals relied on facial physiognomy and skin tone in racial categorization of these targets. These faces were generated to look more or less prototypically East Asian by varying features like forehead size, face width, and nose shape. It was found that both adults and children rely on facial physiognomy and skin tone in their categorizations; however, it is worth noting that their analysis treated facial physiognomy as a cluster of features, rather than as individual features. A second study that has examined the features that correspond with Asian (along with Black and White) prototypicality was conducted by Strom, Zebrowitz, Zhang, Bronstad, and Lee (2012). Black, White, and Korean participants rated the prototypicality of Black, White, and Korean faces and the relative contribution of skin tone and features were assessed controlling for target gender, target attractiveness, target babyfacedness, and proportion of individuals who stated the target appeared to be smiling. Features included: vertical eye height, jaw width, eye separation, nose width, nose length, eyebrow separation, horizontal eye width, mouth width, lip thickness, eyebrow height, and chin to pupil height. The results of the study were fairly complicated and involved interactions between targets' and perceivers' social category membership. However, most relevant to the current review, ratings of Korean prototypicality were more affected by facial features than skin tone. Although the relative importance of a given feature depended on whether the researchers were comparing Korean to White or Korean to Black targets, jaw width, eye separation, eye brow separation, and mouth width appeared to uniquely correspond with Korean prototypicality.

THE ROLE OF FEATURES IN JUDGMENT

Most stereotyping and prejudice research focuses on category-level differences (e.g., how negatively do people feel toward Blacks *relative to* Whites; Allport, 1954; Fiske & Taylor, 1991); however, as we alluded to earlier, researchers have established that within-category variation can influence stereotyping and prejudice over and above category membership alone. Over the past decade, researchers have demonstrated the role that racial prototypicality has in informing a variety of social judgments—again, this work has almost exclusively focused on Blacks in reference to Whites (Hagiwara, Kashy, & Cesario, 2012; Maddox & Gray, 2002; Stepanova & Strube, 2012b). In an especially profound illustration of the role of racial prototypicality on judgment, Blair, Judd, and Chapleau et al. (2004) found that individuals, both Black and White, who were rated as having more Afrocentric features received harsher prison sentences even after controlling for prior criminal records and crime severity (see also Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Their analysis revealed a 7–8 month difference in sentence length comparing those one standard deviation below and one standard deviation above the mean in Black prototypicality. Notably, this pattern emerged despite the lack of a mean-level difference between Blacks and Whites in sentence length, demonstrating the unique role that racial prototypicality can have on impacting judgments in some contexts. Demonstrating yet another profound consequence of features on individuals, Hagiwara, Penner, Gonzalez, and Albrecht (2013) demonstrated that Black individuals who possessed more Afrocentric features reported receiving less fair treatment, and that this perception mediated the relationship between features and poorer physical and mental health. By contrast, Kahn, Goff, Lee, and Motamed (2016) have reported that individuals who look more prototypically White are treated less severely than their counterparts by law enforcement.

Laboratory investigations have established that more prototypic targets elicit increased stereotyping on explicit judgment tasks. Blair and colleagues (2002) employed an impression formation paradigm in which participants were presented with a description of an individual that varied in terms of stereotypically Black behavior and valence. Participants were then given photographs of Blacks and Whites who varied in Afrocentricity and were asked to rate the probability that each photograph was the individual being described. Researchers found that more Afrocentric targets were rated as more likely to be the person in the stereotypically Black descriptions (see also Blair, Judd, & Fallman, 2004). Recent research from social psychologists has also highlighted the consequence of skin tone on evaluations of Blacks, showing that individuals expressed greater negativity toward darker-skinned Blacks than lighter-skinned Blacks (Maddox & Gray, 2002). Similar skin tone results have been demonstrated with regard to American Hispanics' and Chileans' attitudes toward lighter- and darker-complected Latinos (Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). Further, convergent evidence using functional magnetic resonance imaging (fMRI) has demonstrated that exposure to dark-skinned White males (unprototypic Whites) elicited greater amygdala activation relative to light-skinned White males (Ronquillo et al., 2007).

The role of racial prototypicality on judgment is robust, and all attempts to undermine its effects have been ineffective to date. Blair, Judd and Fallman (2004) showed that participants can successfully avoid the influence of race, when instructed (i.e., they show no difference between their ratings of Black and White targets, on average). However, participants cannot avoid using a target's prototypicality in their judgments even when explicitly instructed to do so. Additionally, although cognitive load manipulations can increase participants' use of race, participants' reliance on prototypicality remains unchanged under cognitive load (Blair, Judd, & Fallman, 2004). Consistent with these findings, research from our lab has similarly shown that, with training, participants can avoid using categories in their decisions, but the influence of features persists (Ma & Correll, 2011).

THE CURRENT RESEARCH

Afrocentric features impact a host of important judgments (Blair et al., 2002; Livingston & Brewer, 2002). However, far less is known about prototypicality among Asians (generally, but see Dunham et al., 2016; Strom et al., 2012), Latinos, and, for that matter Whites (for two exceptions see Kahn, Unzueta, Davies, Alston, & Lee, 2015; Uhlmann et al., 2002). Previous research has predominantly focused on Black prototypicality and it has perhaps been implicitly assumed that the same features that predict Black prototypicality have the inverse relationship to White prototypicality (e.g., if Black prototypicality corresponds with darker skin, White prototypicality corresponds with lighter skin). Moreover, some of the research that has been conducted on Asians utilizes artificial, computer-generated faces that were created to look more or less Asian by researchers and the analysis that was reported does not allow us to examine the contributions of individual features in prototypicality (Dunham et al., 2016).

Our research advances the field's current understanding in four ways. First, we attempt to replicate previous research examining Blacks using a different set of stimuli than was used by Blair and Judd (2010) and Strom et al. (2012). Second, we aim to identify the physical attributes that signal good category fit among Asians, Blacks, Latinos, and Whites using images of real—rather than computer-generated—faces (Dunham et al., 2016). Third, we deviate from previous research by only including objective physical measurements in our models. Previous research establishing Black and White prototypicality included subjective ratings of targets' hair volume and texture, which raises some concerns because cues regarding race can distort perception and/or judgment (Correll, Cloutier, & Mellinger, 2016; Correll, Wittenbrink, Crawford, & Sadler, 2015). A rater's judgment of hair texture could be unintentionally influenced by another feature, appearing kinkier if that person has other more Black prototypic features. To avoid this possibility, we utilize only objective measurements. Fourth, and most critically, we examine prototypicality of each racial group without directly referencing other groups (see Blair & Judd, 2010; Strom et al., 2012). In the current work, we obtained separate ratings of racial prototypicality for each racial group and ran separate analyses for each race. We employed the same modeling techniques used by Blair and Judd (2010). In particular, we regressed the subjective ratings of prototypicality on the objective measures using best subsets regression, multiple regression, and stepwise regression. Separate models were carried out to predict Asian, Black, Latino, and White prototypicality. We then looked across the resulting models to identify common factors and examined the goodness of fit of each model. This research contributes to the field's current understanding by revealing the features corresponding to racial prototypicality.

METHOD

To model racial prototypicality we needed a stimulus set that contained a reasonably representative and large number of face stimuli. For this we turned to the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015; www.chicagofaces.org). The CFD is a free database of high-resolution, standardized digital photographs of almost 600 individuals with neutral emotion expressions. The original release of the CFD was comprised of 158 images of Black (48 female, 37 male) and White (37 female, 36 male) individuals. Since that time, the CFD has been greatly expanded and now includes: 57 Asian females, 52 Asian males, 104 Black females, 93 Black males, 56 Latinas, 52 Latinos, 90 White females, and 93 White males, making it the largest database of standardized facial stimuli available. To supplement the stimulus set, the current investigation includes an additional set of 22 Latinas and 19 Latinos, who were photographed in the same conditions as the other CFD targets. The sample size was determined by the number of targets that were available in the CFD, but we believe our study was sufficiently powered. We anticipated a medium-large effect size ($f^2 = .25$), based on the results reported by Blair and Judd (2010). Powered at 0.80, with 17 predictors, and with a probability level of .05, we estimated that we needed a sample of 94 to reasonably power the study. Each of our racial categories includes at least this many targets.

The CFD also contains a host of physical measures of the targets. These include cheekbone height, cheekbone prominence, chin length, eye closeness, eye luminance, eye shape, eye height, eyebrow height, eyelid height, facial width-to-height ratio (fWHR), hair luminance, heartshapeness, lip fullness, midface length, nose shape, skin luminance, and upper-head length[2] (see Table 1). These specific features were identified as significant in the face perception research and research on racial categorization taken from both psychology and anthropology (e.g., Blair & Judd, 2010; Cartmill, 1998; Dibennardo & James, 1983; Oliver, Jackson, Moses, & Dangerfield, 2004; Zebrowitz, 1997). A full description of how these measurements were

---

[2] Several other measures were initially included in our preliminary analyses. For example, we included face roundness, but this metric was highly correlated with fWHR ratio, $r(637) = .84$, $p < .001$, and was automatically excluded in the multiple regression analyses on the basis of collinearity. We also included RBG values for the eye and hair in exploratory analyses, but observed that there were no effects of these variables. The inclusion of these variables did not affect the results we report herein. As a result, we excluded these factors from the analysis.

TABLE 1. Facial Metrics Used to Predict Racial Prototypicality

Feature	Measure/Computation of Measure
Cheekbone Height	(Average midcheek to chin for right and left) divided by face length
Cheekbone Prominence	(Face width at cheek minus face width at mouth) divided by face length
Chin Length	Bottom of lip to chin divided by face length
Eye Distance	Distance between center of pupils
Eye Luminance	Median luminance for eye inclusive of iris and pupil
Eye Shape	Eye height divided by eye width
Eye Height	Eye height divided by face length
Eyebrow Height	Eyebrow height divided by face length
Eyelid Height	Eyelid height divided by face length
fWHR	Facial width to height ratio
Hair Luminance	Median luminance for hair (taken of eyebrow for bald targets)
Heartshapeness	Face width at cheeks divided by face width at mouth
Lip Fullness	Lip thickness divided by face length
Midface Length	(Average pupil to lip for right and left) divided by face length
Nose shape	Nose width divided by nose length
Skin Luminance	Median luminance for target's face only
Upper Head Length	Forehead divided by face length

obtained is available in Ma and colleagues (2015, see Figure 1). Additionally, we needed to obtain racial prototypicality ratings for each of these targets. Participants were 355 individuals recruited from Amazon's MTurk who received nominal monetary compensation in exchange for their participation. No demographic information was collected about the sample, but large-scale analyses of MTurk respondents indicates these samples are equally comprised of males and females, have a mean age of between 30–50 years old, and are largely White (Levay, Freese, & Druckman, 2016). For this data collection, participants were limited to those who resided in the United States. After providing consent, participants were randomly assigned to one of eight possible race-by-gender sets of faces to rate, for example, Black males or Asian females. Participants provided a maximum of 60 ratings to avoid rater fatigue. For categories with more than 60 photos (Black and White faces of both genders), we obtained twice as many ratings to ensure that we had a stable estimate. Each target received a minimum of 30 ratings. We also obtained a separate sample of 33 participants who provided ratings of the 41 additional Latino faces. All participants were instructed to rate targets relative to the targets' gender and race group; for example: "In this survey, you will be shown pictures of [Asian males]. These people differ in terms of how much their physical features resemble the features of [Asian] people. For example, their skin color, hair, eyes, nose, cheeks, lips, and other physical features, may be more [Asian] (i.e., typical of [Asians]) or less [Asian] (i.e., less typical of [Asians]). For this study, we will show you pictures of people one at a time and your job will be to rate how [Asian] looking each person's physical features are on a scale from 'Less Typically [Asian] Looking' to 'Very Typically

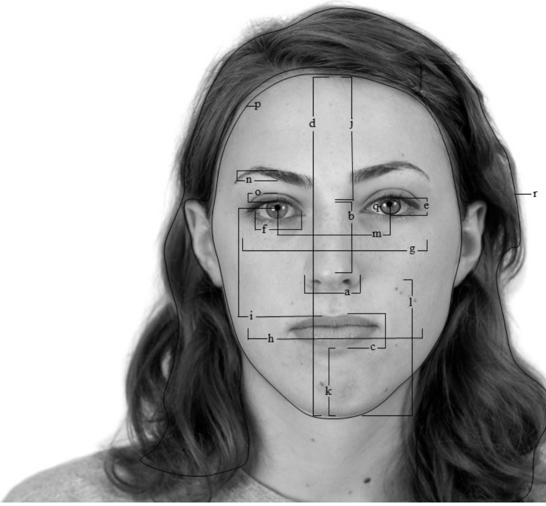

FIGURE 1. a: Nose Width, b: Nose Length, c: Lip Thickness, d: Face Length, e: Eye Height, f: Eye Width, g: Face Width at Cheeks, h: Face Width at Mouth, i: Pupil Center to Lips, j: Forehead, k: Bottom Lip to Chin, l: Midcheek to Chin, m: Eye Distance, n: Eyebrow Height, o. Eyelid Height, p: Skin Luminance, q: Eye Luminance, r: Hair Luminance. Printed with permission from the Chicago Face Database.

[Asian] Looking.'" Participants were presented neutrally expressed targets one at a time and were asked to rate each target on a 1–5 scale (1 = Less Typically [Asian] Looking, 5 = Very Typically [Asian] Looking). From these prototypicality ratings, we computed an average for each target and ran separate regressions for each of the four race groups using this average prototypicality rating as the criterion.[3] We collapsed across gender within each racial group in the current research. Predictors for each model included a standard set of objective measures: cheekbone height, cheekbone prominence, chin length, eye closeness, eye luminance, eye shape, eye height, eyebrow height, eyelid height, facial width-to-height ratio (fWHR), hair luminance, heartshapeness, lip fullness, midface length, nose shape, skin luminance, and upper-head length as the predictors.

Within each racial group we modeled the data using three different regression techniques: best subsets regression, multiple regression, and stepwise regression. Best subsets regression is an automatic linear modeling procedure that exhaustively tests all possible models resulting from every combination of predictors. Within the best subsets options in SPSS, we allowed for automatic variable transformations and data trimming.[4] Multiple regression in the current context is a simul-

---

[3] In a separate set of analyses, we also regressed racial categorization (e.g., for Asians we regressed the proportion of respondents who categorized the target as Asian) on the predictors and obtained highly similar results.

[4] We also conducted best subsets analyses disallowing transformations of variables and data trimming. When compared to the analyses we presented, we observed that the models were very similar. Unfortunately, because the data that contributed to these two models was different (i.e., these models either included or excluded the outliers), there is currently no way to statistically compare whether including or excluding outliers yielded a significantly different model fit. However, we do note that the predictors that were identified were highly consistent with and without transformations and data trimming, and further note the similarity between the best subset, simultaneous, and stepwise regression results.

taneous regression procedure in which all predictors are entered into the model and each predictor's unique association can be determined. Stepwise regression, like best subsets regression, is an automatic modeling procedure; however, the selection process occurs iteratively such that at each step predictors are added or removed based on the predictors' significance levels. The threshold to enter a factor into the model was set at $p \leq .05$ and to remove a factor was $p \geq .10$.

In follow-up analyses, we also tested for possible differences as a function of target gender. That is, we were interested in determining whether target gender moderated the predictive value of individual features in explaining prototypicality. To do this, we ran analyses in which we let target gender (male = -1, female = 1) interact with each predictor and entered these interaction terms along with coded gender into the models we describe below. We reported instances in which a given feature's effect on prototypicality depended on target gender.

RESULTS

Prototypicality ratings were highly reliable (αs $\geq$ .905), indicating exceptional agreement among raters. Average prototypicality ratings for the groups were as follows: Asian (M = 3.02; SD = 0.81), Black (M = 3.48; SD = 0.77), Latino (M = 2.82; SD = 0.66), and White (M = 3.44; SD = 0.84). A Shapiro-Wilk test for normality revealed non-normal distributions for Black prototypicality, $W(197) = 0.90$, $p < .001$; Latino prototypicality, $W(150) = 0.98$, $p = .03$; and White prototypicality, $W(183) = 0.92$, $p < .001$. The effect for Asian prototypicality was not statistically significant, $W(109) = 0.98$, $p = .09$.

ASIANS

Best Subsets Regression. The resulting model for Asians yielded a significant model fit, $F(10, 94) = 20.93$, $p < .001$, with an Akaike information criterion (AIC) of -143.49. Automatic data trimming identified 5 outliers and 9 significant predictors emerged. In order of importance (a value computed automatically by best subsets regression), these included: eye height, $b = -106.78$, $p < .001$; eyelid height, $b = -21.09$, $p < .001$; eye shape, $b = 10.34$, $p < .001$; upper-head length, $b = -7.88$, $p = .001$; nose shape, $b = -1.52$, $p = .002$; fWHR, $b = 4.68$, $p = .004$; skin luminance, $b = 0.01$, $p = .01$; chin length, $b = -7.36$, $p = .01$; lip fullness, $b = 9.44$, $p = .01$. Asian prototypicality corresponded with narrower, yet rounder eyes, thinner eyelids, smaller foreheads, narrower noses, larger fWHR, lighter skin tones, shorter chins, and fuller lips (see Table 2). The Breusch-Pagan test found no evidence of heteroscedasticity, $F(9, 94) = 1.31$, $p = .24$. Follow-up tests showed no evidence of interactions by target gender.

Multiple Regression. Next, we regressed prototypicality on the objective measures in a simultaneous regression. The resulting model was significant, $F(17, 91) = 12.55$, $p < .001$, $R^2_{adj} = .70$. Four factors were significantly associated with racial prototypicality: eye shape, $b = 10.19$, $p = .001$; eye height, $b = -104.39$, $p < .001$; eyelid height, $b = -22.28$, $p < .001$; and nose shape, $b = -1.50$, $p = .007$. None of the

TABLE 2. Results from Best Subsets Regression, Multiple Regression, and Stepwise Regression Predicting Asian Prototypicality Ratings of Asian Targets Using 17 Facial Features

	Asian - Best Subsets					Asian - Multiple Regression					Asian - Stepwise Regression				
Facial Feature	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***
Cheekbone Height						2.573	2.400	-2.193:7.340	1.072	0.286					
Cheekbone Prominence						3.068	16.123	-28.960:35.095	0.190	0.850	6.502	1.860	2.813: 10.192	3.496	0.001
Chin Length	-7.356	2.856	-13.027: -1.684	-2.575	0.012	-7.670	4.670	-16.947: 1.607	-1.642	0.104					
Eye Distance						0.003	0.003	-0.004: 0.009	0.809	0.421					
Eye Luminance						0.006	0.007	-0.008: 0.019	0.804	0.424					
Eye Shape	10.343	2.726	4.930:15.756	3.794	< .001	10.187	3.033	4.163: 16.211	3.359	0.001					
Eye Height	-106.777	18.600	-143.709: -69.846	-5.741	< .001	-104.385	20.439	-144.985: -63.786	-5.107	< .001	-37.664	8.088	-53.706: -21.622	-4.657	< .001
Eyebrow Height						0.872	7.153	-13.337: 15.082	0.122	0.903					
Eyelid Height	-21.094	5.530	-32.073: -10.114	-3.815	< .001	-22.281	5.560	-33.325: -11.238	-4.008	< .001	-27.713	5.396	-38.415: -17.011	-5.136	< .001
fWHR	4.679	1.584	1.529:7.830	2.949	0.004	5.119	2.725	-0.295: 10.533	1.878	0.064					
Hair Luminance						-0.001	0.004	-0.009: 0.008	-0.225	0.823					
Heartshapeness						0.524	5.955	-11.305: 12.353	0.088	0.930					
Lip Fullness	9.435	3.746	1.998:16.873	2.519	0.013	8.144	6.132	-4.036: 20.324	1.328	0.187	12.533	3.464	5.662: 19.403	3.618	< .001
Midface Length						0.035	5.362	-10.616: 10.687	0.007	0.995					
Nose Shape	-1.517	0.483	-2.475:-0.558	-3.143	0.002	-1.503	0.543	-2.581: -0.425	-2.768	0.007	-1.925	0.440	-2.797: -1.052	-4.376	< .001
Skin Luminance	0.011	0.004	0.003:0.019	2.623	0.010	0.008	0.004	-0.001: 0.017	1.703	0.092	0.014	0.004	0.006: 0.022	3.628	< .001
Upper Head Length	-7.876	2.333	-12.508:-3.245	-3.376	0.001	-4.138	3.851	-11.788: 3.513	-1.074	0.286					

other predictors met conventional levels of statistical significance, though there were some marginal effects (see Table 2 for full results). These predictors emerged in the best subsets analysis; however, chin length, fWHR, lip fullness, skin tone, and upper-head length were not significant in this model. The Breusch-Pagan test found no evidence of heteroscedasticity, $F(17, 91) = 0.87$, $p = .61$. Follow-up tests showed no evidence of interactions by target gender.

Stepwise Regression. Stepwise regression yielded a significant model fit, $F(6, 102) = 28.18$, $p < .001$, $R^2_{adj} = .62$. The factors that the model identified were generally consistent with the previous two models. Eyelid height, $b = -27.71$, $p < .001$; skin luminance, $b = 0.01$, $p < .001$; cheekbone prominence, $b = 6.50$, $p = .001$; lip fullness, $b = 12.53$, $p < .001$; eye height, $b = -37.66$, $p < .001$; and nose shape, $b = -1.93$, $p < .001$ all corresponded with ratings of prototypicality (see Table 2). The Breusch-Pagan test found no evidence of heteroscedasticity, $F(6, 102) = 0.86$, $p = .53$. Follow-up tests showed no evidence of interactions by target gender.

BLACKS

Best Subsets Regression. Overall, the model was significant, $F(9, 186) = 34.35$, $p < .001$, with an AIC of -274.70. Ten outliers were identified and omitted from the analysis and 7 significant predictors emerged. In order of importance, these were: skin luminance, $b = -0.02$, $p < .001$; nose shape, $b = 1.20$, $p = .001$; eye brow height, $b = -14.68$, $p = .001$; lip fullness, $b = 7.97$, $p = .002$; cheekbone height, $b = 2.86$, $p = .002$; upper-head length, $b = 5.45$, $p = .004$; and eye shape, $b = -2.63$, $p = .006$. More prototypic Black faces were associated with darker skin tones, broader noses, thinner eyebrows, fuller lips, higher cheekbones, longer foreheads, and narrower eyes (see Table 3). The Breusch-Pagan test revealed evidence of heteroscedasticity, $F(7, 189) = 2.30$, $p = .03$. A plot of the unstandardized predicted value against the squared residuals indicated greater variability at lower than higher values of prototypicality.

A follow-up analysis including target gender and interaction terms for target gender uncovered a gender × fWHR interaction, $b = 4.18$, $p < .001$. We found that lower fWHR was associated with greater prototypicality for males, $b = -7.62$, $p < .001$, but not for females, $b = -0.45$, $p = .54$. We also found a gender × nose shape interaction, $b = 0.90$, $p < .001$, such that wider noses were associated with greater prototypicality in males, $b = 2.32$, $p < .001$, and females, $b = 1.411$, $p = .002$, but the effect was larger for males. There was also evidence for a gender × cheekbone prominence, $b = -3.72$, $p < .001$. Whereas less prominent cheekbones were associated with greater prototypicality for females, $b = -3.85$, $p = .01$, the opposite was true for males, $b = 4.31$, $p = .01$. Lastly, we found a significant gender × hair luminance interaction, $b = -0.01$, $p = .003$. Lighter hair was associated with prototypicality for females, $b = -0.01$, $p = .02$, but darker hair was associated with greater prototypicality for males, $b = 0.01$, $p = .001$.

Multiple Regression. Simultaneous, multiple regression yielded a significant overall model, $F(17, 179) = 18.25$, $p < .001$, $R^2_{adj} = .63$. The following 5 predictors significantly related to racial prototypicality: cheekbone height, $b = 3.91$, $p = .005$;

TABLE 3. Results from Best Subsets Regression, Multiple Regression, and Stepwise Regression Predicting Black Prototypicality Ratings of Black Targets Using 17 Facial Features

	Black - Best Subsets					Black - Multiple Regression					Black - Stepwise Regression				
Facial Feature	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***
Cheekbone Height	2.858	0.929	1.025: 4.690	3.077	0.002	3.913	1.382	1.186: 6.641	2.831	0.005	2.906	0.920	1.091: 4.722	3.158	0.002
Cheekbone Prominence						7.641	9.502	-11.109: 26.391	0.804	0.422					
Chin Length						-5.029	3.278	-11.496: 1.439	-1.534	0.127	-4.702	2.189	-9.020: -0.384	-2.148	0.033
Eye Distance						0.001	0.002	-0.003: 0.004	0.354	0.724					
Eye Luminance						0.001	0.003	-0.006: 0.007	0.205	0.838					
Eye Shape	-2.629	0.938	-4.480: -0.778	-2.802	0.006	-3.006	2.219	-7.385: 1.373	-1.355	0.177	-2.538	0.920	-4.354: -0.723	-2.758	0.006
Eye Height						3.084	13.510	-23.575: 29.743	0.228	0.820					
Eyebrow Height	-14.678	4.538	-23.630: -5.726	-3.235	0.001	-14.152	4.736	-23.496: -4.807	-2.988	0.003	-15.771	4.505	-24.657: -6.886	-3.501	0.001
Eyelid Height						7.755	4.580	-1.283: 16.793	1.693	0.092					
fWHR						-2.278	1.717	-5.666: 1.111	-1.327	0.186					
Hair Luminance						-0.005	0.004	-0.012: 0.002	-1.289	0.199					
Heartshapeness						-2.983	3.897	-10.672: 4.706	-0.766	0.445					
Lip Fullness	7.973	2.539	2.964: 12.983	3.140	0.002	6.970	3.616	-0.166: 14.105	1.928	0.055	7.753	2.539	2.743: 12.762	3.053	0.003
Midface Length						-0.154	3.458	-6.978: 6.671	-0.044	0.965					
Nose Shape	1.196	0.339	0.527: 1.866	3.524	0.001	1.422	0.403	0.627: 2.217	3.531	0.001	1.226	0.340	0.555: 1.897	3.605	< .001
Skin Luminance	-0.020	0.002	-0.025: -0.016	-8.905	< .001	-0.020	0.003	-0.025: -0.015	-7.822	< .001	-0.020	0.002	-0.024: -0.015	-8.717	< .001
Upper Head Length	5.451	1.861	1.779: 9.122	2.929	0.004	5.994	2.313	1.430: 10.558	2.591	0.010	5.991	1.836	2.369: 9.613	3.263	0.001

eyebrow height, $b = -14.15$, $p = .003$; nose shape, $b = 1.42$, $p = .001$; skin luminance, $b = -0.02$, $p < .001$; and upper-head length, $b = 5.99$, $p = .01$. These 5 predictors were significant in the best subset models (see Table 3 for a full set of results). The Breusch-Pagan test revealed evidence of heteroscedasticity, $F(17, 179) = 2.05$, $p = .01$. Inspection of a scatterplot showed a similar pattern of variability in errors that was evident in the best subsets modeling.

We also observed moderating effects of features by target gender. There was a gender × cheekbone prominence interaction, $b = -8.92$, $p = .04$. Cheekbone prominence was not a predictor for either Black males, $b = 17.20$, $p = .16$, or females, $b = -6.30$, $p = .57$, but the direction of the effect of cheekbone was significantly different and had opposing effects. There was a significant gender × fWHR, $b = 4.68$, $p < .001$. Greater fWHR was associated with lower ratings of prototypicality for males, $b = -10.07$, $p < .001$, but there was no evidence of an effect for females, $b = 0.13$, $p = .95$. We also observed a gender × hair luminance interaction, $b = -0.01$, $p = .003$, whereby darker hair predicted prototypicality among females, $b = -0.01$, $p = .03$, but lighter hair predicted higher prototypicality among males, $b = 0.01$, $p = .05$. Finally, there was a gender × nose shape, $b = -0.69$, $p = .04$. Wider noses was associated with greater prototypicality for females, $b = 1.36$, $p = .01$, and males, $b = 2.66$, $p < .001$, but the effect was larger for males.

Stepwise Regression. Stepwise regression yielded a significant model fit, $F(8, 188) = 38.07$, $p < .001$, $R^2_{adj} = .62$. Eight predictors were associated with prototypicality, including: cheekbone height, $b = 2.91$, $p = .002$; chin length, $b = -4.70$, $p = .03$; eye shape, $b = -2.54$, $p = .006$; eyebrow height, $b = -15.77$, $p = .001$; lip fullness, $b = 7.75$, $p = .003$; nose shape, $b = 1.23$, $p < .001$; skin luminance, $b = -0.02$, $p < .001$; and upper-head length, $b = 5.99$, $p = .001$. These 8 predictors subsumed those that emerged from the best subsets and multiple regression analyses (see Table 3). Again, the Breusch-Pagan test revealed heteroscedasticity, $F(6, 190) = 3.20$, $p = .005$. Inspection of the scatterplot was consistent with the previous two models.

A follow-up analysis including interaction terms for target gender uncovered three different interaction effects. First, we again observed a significant gender × hair luminance interaction, $b = -0.01$, $p < .001$. Consistent with the multiple regression analysis, we found that darker hair predicted prototypicality among females, $b = -0.01$, $p = .02$, but lighter hair predicted higher prototypicality among males, $b = 0.01$, $p = .001$. Second, we found a significant gender × cheekbone prominence interaction, $b = -3.89$, $p < .001$. The pattern of the interaction was consistent with the other models; cheekbone prominence negatively predicted prototypicality for females, $b = -3.85$, $p = .01$, but had a positive association for males, $b = 4.31$, $p = .01$. Third, we found a gender × lip fullness interaction, $b = -3.23$, $p = .04$. Fuller lips predicted higher ratings of prototypicality, but this effect was stronger among males, $b = 13.64$, $p < .001$, than females, $b = 6.89$, $p = .008$.

LATINOS

Best Subsets Regression. Overall, we observed significant model fit, $F(5, 143) = 7.85$, $p < .001$, with an AIC of -162.26. Ten outliers were dropped and 3 significant

TABLE 4. Results from Best Subsets Regression, Multiple Regression, and Stepwise Regression Predicting Latino Prototypicality Ratings of Latino Targets Using 17 Facial Features

	Latino - Best Subsets					Latino - Multiple Regression					Latino - Stepwise Regression				
Facial Feature	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***
Cheekbone Height						2.190	1.949	-1.665: 6.046	1.124	0.263	2.962	.974	1.037: 4.886	3.042	0.003
Cheekbone Prominence						-9.933	12.355	-34.372: 14.505	-0.804	0.423					
Chin Length						-4.488	4.878	-14.136: 5.161	-0.920	0.359					
Eye Distance						0.000	0.004	-0.008: 0.007	-0.033	0.974					
Eye Luminance	-0.013	0.003	-0.019: -0.006	-3.656	< .001	-0.012	0.004	-0.020: -0.004	-2.946	0.004	-0.015	0.003	-0.021:-0.009	-4.669	< .001
Eye Shape	3.862	1.156	1.578:6.147	3.342	0.001	3.211	2.798	-2.323: 8.746	1.148	0.253	2.502	1.039	.448: 4.556	2.407	0.017
Eye Height						-2.797	19.334	-41.042: 35: 448	-0.145	0.885					
Eyebrow Height						-4.753	8.805	-22.171: 12.665	-0.540	0.590					
Eyelid Height						10.025	7.568	-4.946: 24.995	1.325	0.188					
fWHR						2.236	2.507	-2.723: 7.194	0.892	0.374					
Hair Luminance						0.001	0.004	-0.007: 0.010	0.348	0.728					
Heartshapeness						3.597	5.009	-6.312: 13.505	0.718	0.474					
Lip Fullness						-2.464	5.352	-13.051: 8.123	-0.460	0.646					
Midface Length						1.569	4.769	-7.865: 11.004	0.329	0.743					
Nose Shape						1.206	0.637	0.053: 2.466	1.894	0.060					
Skin Luminance	-0.008	0.004	-0.016:-0.000	-1.994	0.048	-0.010	0.005	-0.019: 0.000	-2.055	0.042					
Upper Head Length						-1.378	2.507	-6.337: 3.581	-0.550	0.584					

predictors emerged: eye luminance, *b* = -0.01, $p < .001$; eye shape, *b* = 3.86, $p = .001$; and skin luminance, *b* = -0.01, $p = .05$. Latino faces were associated with dark skin tone and eye color and rounder eyes (see Table 4). There was evidence of heteroscedasticity using the Breusch-Pagan test, $F(3, 146) = 4.64$, $p = .004$. Examination of a scatterplot of error variance suggested greater variability among the lower than higher prototypic targets.

A follow-up analysis in which we included target gender and gender by feature interactions uncovered a gender × hair luminance interaction, *b* = 0.01, $p = .02$. For females, lighter hair predicted prototypicality, *b* = 0.01, $p = .02$, whereas there was no evidence of a relationship between hair luminance and prototypicality for males, *b* = -0.01, $p = .14$.

Multiple Regression. There was significant model fit resulting from the multiple regression, $F(17, 132) = 2.48$, $p = .002$, $R^2_{adj} = .24$. Only 2 factors, eye luminance, *b* = -0.01, $p = .004$, and skin luminance, *b* = -0.01, $p = .04$, reached conventional levels of significance, although several other factors were marginally associated with prototypicality (see Table 4 for a full set of results). A Breusch-Pagan test showed no evidence of heteroscedasticity, $F(17, 132) = 1.31$, $p = .20$.

Follow-up analyses including target gender and target gender interactions identified two significant effects. There was a gender × eye luminance, *b* = -0.01, $p = .008$. For Latinas, darker eye color predicted prototypicality significantly, *b* = -0.03, $p < .001$, whereas there was no evidence of an association between eye color and prototypicality for males, *b* = -0.00, $p = .84$. We also found a significant gender × hair luminance interaction, *b* = 0.01, $p = .01$. Lighter hair color predicted prototypicality for Latinas, *b* = 0.01, $p = .04$; however, there was no association between hair color and prototypicality for male targets, *b* = -0.01, $p = .10$.

Stepwise Regression. Finally, stepwise regression yielded a significant model fit, $F(3, 146) = 9.83$, $p < .001$, $R^2_{adj} = .17$. Three factors were significant, eye luminance, *b* = -0.02, $p < .001$; cheekbone height, *b* = 2.96, $p = .003$; and eye shape, *b* = 2.50, $p = .02$. Across all models, darker eye color predicted Latino prototypicality. Consistent with the best subsets analysis, we also saw evidence of eye shape predicting prototypicality, although higher cheekbones also predicted prototypicality. Unlike the other models, there was a null effect of skin tone. There was marginal evidence of heteroscedasticity using the Breusch-Pagan test, $F(3, 146) = 2.48$, $p = .06$. Follow-up tests showed no evidence of interactions by target gender.

WHITES

Best Subsets Regression. Best subsets regression yielded a significant model fit, $F(14, 166) = 17.43$, $p < .001$, with an AIC of -194.00. Eight outliers were removed from the analysis and 9 factors emerged as significant predictors. In order of importance these were: eye luminance, *b* = 0.02, $p < .001$; lip fullness, *b* = -20.33, $p < .001$; hair luminance, *b* = 0.01, $p < .001$; skin luminance, *b* = 0.02, $p < .001$; eyelid width, *b* = -23.50, $p < .001$; nose shape, *b* = -1.96, $p < .001$; cheekbone height, *b* = 4.20, $p = .003$; chin length, *b* = -10.83, $p = .005$; and midface length, *b* = -9.57, $p =$

.02. White faces were associated with lighter eye, hair, and skin luminance, thinner lips, thinner eyelids, narrower noses, higher cheekbones, shorter chins and mid-face lengths. There was no evidence for heteroscedasticity, $F(9, 173) = 0.89$, $p = .54$.

Follow-up analyses including target gender and interactions revealed a gender × luminance interaction, $b = 0.01$, $p = .007$. For females, $b = 0.03$, $p = .001$ lighter skin tone was associated with greater prototypicality, but no evidence of such an association was observed among males, $b = 0.01$, $p = .15$. We also found a gender × nose shape interaction, $b = -0.99$, $p = .009$. Whereas narrower noses were associated with greater prototypicality among females, $b = -2.87$, $p < .001$, there was no evidence of this relationship among males, $b = -0.69$, $p = .28$.

Multiple Regression. Multiple regression returned a significant model fit, $F(17, 165) = 14.62$, $p < .001$, $R^2_{adj} = .60$. Eleven of the 17 factors were significantly associated with White prototypicality: cheekbone height, $b = 3.71$, $p = .02$; chin length, $b = -11.10$, $p = .006$; eye closeness, $b = 0.01$, $p = .04$; eye luminance, $b = 0.02$, $p < .001$; eye size, $b = -39.52$, $p = .05$; eyelid height, $b = -20.46$, $p < .001$; hair luminance, $b = 0.01$, $p < .001$; heartshapeness, $b = -8.09$, $p = .05$; lip fullness, $b = -19.44$, $p < .001$; nose shape, $b = -1.68$, $p = .001$; and skin luminance, $b = 0.02$, $p < .001$. White prototypicality is associated with higher cheekbones, shorter chins, wider interocular distance, lighter eye, hair, and skin color, thinner eyelids, less heartshapeness, thinner lips, and narrower noses (see Table 5 for full results). These results and the direction of the effects were consistent with the results from the best subsets model; however, eye height was not significant in this model. A Breusch-Pagan test revealed no evidence for heteroscedasticity, $F(17, 165) = 0.95$, $p = .51$. Follow-up tests showed no evidence of interactions by target gender.

Stepwise Regression. Stepwise regression yielded a significant model fit, $F(8, 174) = 26.91$, $p < .001$, $R^2_{adj} = .55$. Eight factors comprised the final model: eye luminance, $b = 0.02$, $p < .001$; hair luminance, $b = 0.01$, $p < .001$; eyelid thickness, -23.55, $p < .001$; nose shape, $p = -1.70$, $p < .001$; lip fullness, $b = -9.55$, $p < .001$; skin luminance, $b = 0.01$, $p = .001$; heartshapeness, $b = -1.78$, $p = .002$; and eyebrow thickness, $b = -12.68$, $p = .04$ (see Table 5). The Breusch-Pagan test revealed no evidence for heteroscedasticity, $F(8, 174) = 1.60$, $p = .13$. Follow-up tests showed no evidence of interactions by target gender.

GENERAL DISCUSSION

Previous research highlights the importance of racial prototypicality in a host of judgment tasks (Anderson & Cromwell, 1977; Blair, Judd, & Fallman, 2004; Livingston & Brewer, 2002). The aim of the current study was to identify the markers of racial prototypicality and respond to the need for more racial inclusivity in research (Henrich et al., 2010; Sadler, Correll, Park, & Judd, 2012). Through exploratory analysis, we explained a significant proportion of the variability in subject ratings of racial prototypicality Asian, Black, Latino, and Whites using physical measures of the face and hair. Our data suggest that perceptions of Asian prototypicality consistently correspond with narrower eyes, thinner eyelids, and nar-

TABLE 5. Results from Best Subsets Regression, Multiple Regression, and Stepwise Regression Predicting White Prototypicality Ratings of White Targets Using 17 Facial Features

	White - Best Subsets					White - Multiple Regression					White - Stepwise Regression				
Facial Feature	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***	**b**	**b (se)**	**b 95% CI**	***t***	***p***
Cheekbone Height	4.201	1.392	1..452:6.950	3.018	0.003	3.713	1.530	0.691: 6.735	2.426	0.016					
Cheekbone Prominence						17.062	9.834	-2.356: 36.480	1.735	0.085					
Chin Length	-10.833	3.809	-18.354:-3.312	-2.844	0.005	-11.103	3.947	-18.985: -3.310	-2.813	0.006					
Eye Distance						0.006	0.003	0.000: 0.001	2.029	0.044					
Eye Luminance	0.017	0.003	0.012:0.023	6.401	< .001	0.018	0.003	0.013: 0.023	6.651	< .001	0.019	0.003	0.013: 0.024	7.059	< .001
Eye Shape						4.886	3.313	-1.656: 11.427	1.475	0.142					
Eye Height						-39.522	19.570	-78.161: -0.882	-2.020	0.045					
Eyebrow Height						-11.432	6.018	-23.315: .451	-1.900	0.059	-12.675	6.068	-24.652: -0.698	-2.089	0.038
Eyelid Height	-23.502	5.621	-34.600: -12.403	-4.181	< .001	-20.463	5.755	-31.827: -9.099	-3.555	< .001	-23.547	5.493	-34.388: -12.706	-4.287	< .001
fWHR						-1.984	2.091	-6.113: 2.144	-0.949	0.344					
Hair Luminance	0.010	0.002	0.006:0.015	4.527	< .001	0.008	0.002	0.004: 0.012	4.165	< .001	0.009	0.002	.005: .013	4.746	< .001
Heartshapeness						-8.088	4.055	-16.095: -0.081	-1.994	0.048	-1.775	0.560	-2.880: -.669	-3.168	0.002
Lip Fullness	-20.329	4.097	-28.418: -12.240	-4.962	< .001	-19.435	4.179	-27.686: -11.183	-4.650	< .001	-9.550	2.617	-14.715: -4.385	-3.650	< .001
Midface Length	-9.565	4.105	-17.670: -1.460	-2.330	0.021	-7.660	4.331	-16.211: 0.891	-1.769	0.079					
Nose Shape	-1.958	0.548	-3.040: -0.877	-3.575	< .001	-1.682	0.516	-2.701: -0.663	-3.258	0.001	-1.698	.442	-2.570: -.827	-3.846	< .001
Skin Luminance	0.020	0.005	0.011:0.029	4.260	< .001	0.023	0.005	0.013: 0.033	4.556	< .001	0.013	0.004	.005: .021	3.235	0.001
Upper Head Length						-5.351	3.068	-11.409: 0.707	-1.744	0.083					

rower noses. Across all three modeling techniques, these three metrics explained a significant amount of variability in Asian prototypicality, $F(3, 105) = 33.38$, $p < .001$, $R^2 = .49$. Higher cheekbones, thinner eyebrows, broader noses, darker skin tone, and a longer upper head significantly predicted Black prototypicality across all models and accounted for a significant amount of variability, $F(5, 191) = 46.65$, $p < .001$, $R^2 = .55$. Latino prototypicality was more difficult to capture and darker eye color alone predicted across the modeling strategies. This factor accounted for a significant, but fairly small amount of variance in Latino prototypicality, $F(1, 148) = 13.87$, $p < .001$, $R^2 = .09$. Finally, White prototypicality was dependably associated with lighter eye color, lighter hair color, lighter skin, thinner lips, and narrower noses. Together, these five features explained significant variability in White prototypicality, $F(5, 144) = 4.81$, $p < .001$, $R^2 = .14$.

One of the findings that stood out in the current analysis is the dissociation between features relevant to Black and White prototypicality. As we alluded to in the introduction, previous research has at times (perhaps unintentionally) assumed that Black and White targets are mutually exclusive by studying contrasting these groups with each other. For example, participants might be asked to rate both Black and White faces in terms of Afrocentricity (Blair et al., 2002; Blair, Judd, & Fallman, 2004). By only using Black and White targets in research, participants may infer that what defines one group has the inverse effect for the other group. It is possible that these experimental contexts create these perceptions. This may have inadvertently led previous research to define one group's prototypicality in opposition to the other group (e.g., a broad nose is Afrocentric rendering a narrow nose Eurocentric). Here we see that this is not entirely the case. Although the skin tone effect is clearly negative for Blacks and positive for Whites, cheekbone height has no observed value in predicting prototypicality in Whites, though it does for Blacks. To us this provides evidence cautioning against the view that perceptions of Blacks and Whites ought to be dichotomized.

We also want to highlight the fact that, despite research suggesting that Hispanic individuals living in the U.S. regard their ethnicity as part of their racial background, our models provided relatively little insight into the features that correspond with Latino prototypicality. This finding may provide support for the idea that treating Latinos/Hispanics as a racial category in this kind of research may be problematic. Alternatively, it could mean that the features that we have included in our models (most of which come from research that again has focused on different groups) are simply insufficient to capture Latino prototypicality.

Although we can account for significant variability in racial prototypicality, we raise several points of caution readers should consider when interpreting our results. There may very well be other facial metrics that can explain racial prototypicality that we have not considered and that were not included in the current analysis. Although we scoured the psychological and anthropological literatures relating to race perception to select the objectively measurable features we ultimately included in our modeling (Cartmill, 1998; Oliver, Jackson, Moses, & Dangerfield, 2004; Zebrowitz, 1997), there are a host of other features that might be associated with prototypicality that we have not considered. As we noted earlier, we initially recorded a variety of other measures, such as RGB (red, green, blue) values of skin, hair, and eyes, but there was no evidence that these were systematically associated with racial prototypicality for any of the four groups. In order to simplify the models and reduce susceptibility to Type I errors, we elected to exclude them in these analyses. It was also the case that other facial metrics like face roundness highly correlated with fWHR, which lead to collinear predictors. That said, there may be other features that are highly predictive of racial prototypicality that we have overlooked and this vulnerability is a shortcoming of our approach. Relatedly, features may *interactively* impact psychological perceptions (e.g., the combination of narrower noses and higher cheekbones may make a person look especially White) and we did not account for such a possibility here by allowing factors to interact. Without a basis in the literature for motivating this sort of prediction, doing so seems arbitrary and could further compound Type I errors.

Additionally, and along similar lines, whereas previous research used factors that were obtained from judges to predict prototypicality, we deliberately chose not to include more subjective features that may correlate with racial prototypicality in our analysis. This was done for two reasons. First, hair volume and hair texture were the two predictors that have previously been considered in assessing Black and White racial prototypicality; however, these two features are not fixed and may change wildly as a function of stylistic trends. Second, we wanted to avoid the possibility that judges might be influenced by top-down information when providing ratings of one aspect of the target. For example, participants might come to see a target's hair as kinkier if that target has other more prototypically Black features (Blair & Judd, 2010). It is also possible that one's judgment about one feature may be influenced by another feature and we aimed to isolate individual features here. We mitigate these concerns by only including objectively measurable features.

Third is the issue of representativeness. Our approach assumes that the targets we used are representative of Asian, Black, Latino, and White people in the world; however, that is likely not the case. Our sample was largely American-born and raised individuals and as such, their features may be uniquely American. Moreover, subsets of individuals may be overrepresented or underrepresented given the locations in which the targets were gathered and immigration patterns in the U.S. For example, the Latino faces were collected from a convenience sample in Los Angeles and many Latino immigrants in the regions come from Mexico and Guatemala. Had the sample been collected in a different region of the United States, for example, Miami, we would have a very different sample. Representativeness may also affect the external validity of our Asian prototypicality results. The majority of the Asians in the sample are East Asians, rather than Southeast Asians, which could similarly affect the results we observed. Given our data-driven approach and our reliance on the available sample for data, it is possible—indeed likely—that the constituents of the sample could affect which features emerge as predictive. Validating this research with different stimulus sets may help us verify the associations we observe here. For instance, one could imagine testing the generalizability of these models on a different set of facial stimuli or alternatively, creating computerized faces using these models and testing these artificial faces for racial prototypicality. Relatedly, an area for future research (and a limitation of the current work) is further investigation regarding the representativeness of the perceivers who provided prototypicality ratings. As we described in the introduction, others have demonstrated that the racial or ethnic background of the perceiver can influence the extent to which different features figure into perceptions of prototypicality (Strom et al., 2012; see also Chen, de Paula Couto, Sacco, & Dunham, 2017). Unfortunately, we do not have perceiver data and as such, we cannot examine possible differences in the features different perceivers might utilize when judging prototypicality. We therefore caution that the conclusions that we might draw from the current study might only reflect the perceptions of American individuals.

In comparing our findings with previous research, we explain much less of the variability in prototypicality than Blair and Judd (2010) who report being able to account for 87–88% of the variance in Afrocentricity ratings with just five features: skin tone, hair texture, hair volume, lip fullness, and nose shape. Besides excluding subjective measures like hair volume and texture, we also differ from previous research in how we selected and classified targets by race. Whereas Blair and Judd used targets who were reliably categorized as Black or White (over 99% agreement was found among raters) and Dunham et al. (2016) created faces meant to vary in terms of how Asiatic or White they appeared, we grouped targets based on how the targets themselves racially identified, which in some cases did not comport with perceivers' categorizations and likely introduced more featural variation among targets and possibly the ratings. To us, it seems valuable to determine which features predict prototypicality within a more inclusive rather than narrower set of targets, but this likely contributed to a less clear result. Future research examining perceived prototypicality based on perceiver consensus versus self-identified race would be an interesting next step toward understanding how features influence this important judgment.

REFERENCES

Allport, G. W. (1954). *The nature of prejudice.* Oxford, England: Addison-Wesley.

Anderson, C., & Cromwell, R. L. (1977). "Black is beautiful" and the color preferences of Afro-American youth. *Journal of Negro Education*, *46*(1), 76-88.

Blair, I. V. (2006). The efficient use of race and Afrocentric features in inverted faces. *Social Cognition, 24*(5), 563-579.

Blair, I. V, Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology, 35*(1), 59-68.

Blair, I. V., & Judd, C. M. (2010). Afrocentric facial features and stereotyping. In R. B. Adams, Jr., N. Ambady, K. Nakayama, & S. Shimojo (Eds.), *The science of social vision* (pp. 306-320). New York: Oxford University Press.

Blair, I. V, Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679.

Blair, I. V, Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology, 87*(6), 763-778.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*(1), 5-25.

Cartmill, M. (1998). The status of the race concept in physical anthropology. *American Anthropologist*, *100*(3), 651-660.

Chen, J. M., de Paula Couto, M. C. P., Sacco, A., & Dunham, Y. (2017). To be or not to be (Black or multiracial or White): Cultural variation in racial boundaries. *Social Psychological and Personality Science*. [online first] https://doi.org/10.1177/1948550617725149

Cornell, S., & Hartmann, D. (2007). *Ethnicity and race: Making identities in a changing world* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Correll, J., Cloutier, J., & Mellinger, C. (2016). Discriminating perception. *Psychological Inquiry, 27*(4), 281-285.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of personality and social psychology, 108*(2), 219-233.

Dibennardo, R., & James V. T. (1983). Multiple discriminant function analysis of sex and race in the postcranial skeleton. *American Journal of Physical Anthropology, 61*(3), 305-314.

Dunham, Y., Dotsch, R., Clark, A. R., & Stepanova, E. V. (2016). The development of White-Asian categorization: Contributions from skin color and other physiognomic cues. *PLoS ONE, 11*(6), e0158211. https://doi.org/10.1371/journal.pone.0158211

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2014). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science, 18*(3), 469-483.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*(5), 383-386.

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of White and Black faces by White and Black subjects. *International Journal of Psychology, 10*(2), 119-123.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Hagiwara, N., Kashy, D., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *Journal of Experimental Social Psychology, 48*(4), 892-898.

Hagiwara, N., Penner, L. A., Gonzalez, R., & Albrecht, T. L. (2013). Within-group health disparities among Blacks: The effects of Afrocentric features and unfair treatment. *Cultural Diversity and Ethnic Minority Psychology, 19*(4), 477-480.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(29), 29-29.

Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes and Intergroup Relations, 14*(5), 569-580.

Kahn, K. B., Goff, P. A., Lee. J. K., & Motamed, D. (2016). Protecting Whiteness: White phenotypic racial stereotypicality reduces police use of force. *Social Psychological and Personality Science, 7*(5), 403-411.

Kahn, K. B., Unzueta, M. M., Davies, P. G., Alston, A. T., & Lee, J. K. (2015). Will you value me and do I value you? The effect of phenotypic racial stereotypicality on organizational evaluations. *Journal of Experimental Social Psychology, 59*, 130-138.

Ito, T. A. (2011). Perceiving social category information from faces: Using ERPs to study person perception. In A. Todorov, S. T. Fiske, & D. A. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 85-100). New York: Oxford University Press.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*(2), 275-287.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open, 6*(1), 1-17

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*(1), 5-18.

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47*(2), 391-396.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135.

Maddox, K. B., & Chase, S. G. (2004). Manipulating subcategory salience: Exploring the link between skin tone and social perception of Blacks. *European Journal of Social Psychology, 34*(5), 533-546.

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin, 28*(2), 250-259.

Oliver, M. B., Jackson, R. L., II, Moses, N. N., & Dangerfield, C. L. (2004). The face of crime: Viewers' memory of race-related facial features of individuals pictured in the news. *Journal of Communication, 54*(1), 88-104.

Pew Research Center. (2015). *Multiracial in America—The many dimensions of Hispanic racial identity*. Retrieved May 12, 2017, from http://www.pewsocialtrends.org/2015/06/11/ chapter-7-the-many-dimensions-of-hispanic-racial-identity/

Population Reference Bureau. (2014). *2014 world population data sheet.* Retrieved from https://www.prb.org/2014-world-population-data-sheet/

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z.-L., Nandy, A., & Maddox, K. B. (2007). The effects of skin tone on race-related amygdala activity: An fMRI investigation. *Social Cognitive and Affective Neuroscience, 2*(1), 39-44.

Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not Black and White: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues, 68*(2), 286-313.

Snipp, C. M. (2010). Defining race and ethnicity: The Constitution, the Supreme Court, and the Census. In H. R. Markus & P. M. L. Moya (Eds.), *Doing race*: *21 Essays for the 21st century* (pp. 105-122). New York: Norton.

Stepanova E. V., & Strube, M. J. (2012a). The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes. *Journal of Experimental Social Psychology, 48*(4), 867-878.

Stepanova, E. V., & Strube, M. J. (2012b). What's in a face? The role of skin tone, facial physiognomy, and color presentation mode of facial primes in affective priming effects. *The Journal of Social Psychology, 152*(2), 212-227.

Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS ONE, 7*(7), e41193. https://doi.org/10.1371/journal.pone.0041193

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83-114). London, England: Taylor and Francis.

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*(3), 198-226.

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8*(2), 161-185.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.