Empirical Research Paper

Personality and Social Psychology Bulletin
1–10

DOI: 10.1177/0146167221989836
journals.sagepub.com/home/pspb
SAGE

Are Morphs a Valid Substitute for Real Multiracial Faces in Race Categorization Research?

Debbie S. Ma[1], Justin Kantner[1], Jonathan Benitez[1], and Stephanie Dunn[1]

Abstract

The rise of the multiracial population has been met with a growing body of research examining multiracial face perception. A common method for creating multiracial face stimuli in past research has been mathematically averaging two monoracial "parent" faces of different races to create computer-generated multiracial morphs, but conclusions from research using morphs will only be accurate to the extent that morphs yield perceptual decisions similar to those that would be made with real multiracial faces. The current studies compared race classifications of real and morphed multiracial face stimuli. We found that oval-masked morphed faces were classified as multiracial significantly more often than oval-masked real multiracial faces (Studies 1 and 2), but at comparable levels to unmasked real multiracial faces (Study 2). Study 3 examined factors that could explain differences in how morphs and real multiracial faces are categorized and pointed to the potential role that unusualness/distinctiveness might play.

Keywords

biracial, multiracial, face perception, racial classification, morphs

Received August 4, 2020; revision accepted December 29, 2020

The U.S. Census Bureau began permitting respondents to choose multiple racial categories in 2000. Between 2000 and the time of the next census in 2010, the population of Black-White biracials more than doubled and Asian-White biracials increased by 87% (Parker et al., 2015). While astonishing on their own, these figures may actually *underestimate* the number of multiracial individuals in the population. According to Pew Research, whereas the census estimates 2.1% of American adults reported being multiracial, the actual population may be more than three times that (Parker et al., 2015). The rise of multiracialism is not limited to the United States. Countries around the world are seeing explosive growth in multiracial marriages and multiracial babies being born (Aspinall, 2018; Morning, 2012). As the multiracial population increases, social psychologists have attempted to understand how multiracial individuals are perceived and judged by others. As this nascent literature continues to grow, we believe it is imperative to evaluate the validity of a common operationalization of multiracialism—facial morphs.

A Brief Overview of Multiracial Perception Research

In some of the earliest published studies on multiracial face perception, Peery and Bodenhausen (2008) presented participants with "racially ambiguous" faces along with information about their racial heritage within the context of an impression formation task. Racially ambiguous faces in this case were digital morphs of Black and White faces. Morphing, in the context of creating multiracial faces, involves the blending of two "parent faces" or "source faces," both of which have been previously judged to be monoracial. Morphing relies on Euclidean geometry to create images that mathematically combine two or more inputs. The relative contribution of the inputs can be varied based on users' specifications (e.g., a morphed face may have an 80% contribution from the White parent face and a 20% contribution from the Black parent face); however, multiracial morphs are most typically averages of monoracial parent faces and are thus referred to as biracial. Across two studies, Peery and Bodenhausen (2008) showed that when participants learned that a multiracial target had Black and White ancestry, they were categorized as Black more frequently. This pattern is indicative of hypodescent—the tendency to assign multiracial children to the culturally subordinate racial group. Ho et al.

[1]California State University, Northridge, USA

Corresponding Author:
Debbie S. Ma, Department of Psychology, California State University, 18111 Nordhoff Street, Northridge, CA 91330, USA.
Email: debbie.ma@csun.edu

(2011) conducted conceptually similar studies in which they presented participants with morphed Black-White faces. The relative contributions of these faces varied, such that the morphed faces could be 50% Black-50% White, 40% Black-60% White, and so on. Categorization of these morphed faces illustrated that participants required a relatively minimal non-White contribution to a morph before indicating that the morph was non-White. Ho and colleagues, like Peery and Bodenhausen (2008), concluded that the categorization of racially ambiguous individuals is largely governed by hypodescent.

Chen and Hamilton (2012) asked participants to make speeded classifications of images of monoracial and morphed multiracial faces. Participants made fewer multiracial than monoracial classifications (suggesting a reluctance to categorize others as biracial) and took significantly longer to classify a target as multiracial (suggesting a greater difficulty in making multiracial than monoracial categorizations). Follow-up work by Chen et al. (2014) asked participants to categorize Black, White, and multiracial faces. Here, both real and morphed Black-White faces were included in the study. The central finding of the study was that individuals who had higher internal motivation to control prejudice were more accurate at categorizing multiracial faces as such, compared with those who were lower on this individual difference. Notably, the researchers tested whether the eight real and eight morphed multiracial faces differed in terms of accurate classifications and reported no difference.

Others have examined the representation of multiracial faces through less direct means, focusing instead on how individuals spontaneously encode multiracial people by measuring the well-documented cross-race effect (CRE)—the difficulty people have in individuating and remembering faces from racial outgroups compared with ingroups (Meissner & Brigham, 2001). The rationale is that if multiracial people are construed as outgroup members, participants should show less recognition of those faces relative to racial ingroup faces when later tested. Pauker and Ambady (2009) presented White, Asian, and Asian-White multiracial participants with pictures of real Asian, White, and Asian-White multiracial faces, which they were instructed to learn and memorize. Real multiracial faces were either labeled as Asian or White at the time of learning. Subsequently, participants were shown White, Asian, and Asian-White multiracial faces, some of whom they had previously seen and some of whom were new. Participants indicated whether each test face was previously seen. For monoracial participants, multiracial face memory was better when those faces were labeled as part of the ingroup (e.g., Asian participants remembered multiracials better when they were labeled Asian than White), although multiracial participants showed no evidence that the label affected recognition. These data suggest that participants' perceptions of multiracial faces are subject to influence, at least among monoracial perceivers. Gaither et al. (2013) found that White children aged 4 to 9 who demonstrated race essentialism, an individual difference belief that an immutable, genetic/biological essence defines all members of a racial group (Williams & Eberhardt, 2008), remembered White faces better than Black-White racially ambiguous and Black faces. However, children who had lower levels of race essentialism remembered Black-White racially ambiguous faces as well as White faces. These data suggest that race essentialism may correspond with greater category distinctiveness.

Issues of Validity Surrounding Morphs

This review represents a thin slice of an ever-growing literature devoted to understanding multiracial face perception. These particular papers showcase the types of tasks that have been used to study multiracial face perception, as well as some of the operationalizations of multiracialism. In the broader literature, morphing represents a commonly used means for operationalizing multiracialism (Chen, Pauker, et al., 2018; Ho et al., 2011; Krosch et al., 2013; Pauker et al., 2013). The popularity of morphing undoubtedly stems from necessity, as only very recently have there been resources developed including real multiracial face stimuli (Chen et al., 2020; Ma, Kantner, & Wittenbrink, 2020). Moreover, morphs afford convenience and experimental control. Creating multiracial stimuli from monoracial faces requires little time and cost. Morphing also has appeal from an experimental standpoint, because researchers can create infinite numbers of stimuli and any type of multiracial face (e.g., Black-White, Asian-White, Asian-Black, etc.), vary the contribution of each parent face, and select parent faces based on specific criteria (see Gaither et al., 2019, for related points regarding face stimuli created with FaceGen software).

Despite the logical and practical rationale for using morphs to operationalize biracialism, morphed multiracial faces come with several artifacts that may raise concerns regarding their external validity—the extent to which they perceptually represent real biracial faces. First, morphing produces faces that have a softer and hazier appearance than a high-resolution image of a real face. In the literature, this issue has been described and explored in terms of pixelation (Chen & Hamilton, 2012). In addition, morphing results in faces that tend to be more attractive, because averaging procedures remove idiosyncrasies in the face and reduce imperfections (Halberstadt & Rhodes, 2003). Either of these documented physical differences between morphed and real biracial stimuli may impact how they are classified by observers. In addition, because morphing software yields hazy, unrealistic hair, morphed faces are oval-masked to remove hair. We believe that relying on stimuli that generally cannot include hair could be problematic for studying race perception, because hair is a useful and disambiguating cue in making racial classifications (Maclin & Malpass, 2003).

Current Studies

Researchers have often used morphed multiracial faces in the past; however, as we noted, there are limitations in using morphed stimuli that may undermine their use in race perception and categorization experiments. We argue that when morphs are used to study race perception, their purpose is essentially to serve as a stand-in for real biracial faces that are encountered and classified in the world. This use of morphs is valid if morphs yield race categorization patterns similar to those that would be made with real biracial stimuli. If morphs do not, their use in indexing real-world classification decisions would be limited. In the current experiments, we directly compared classifications of morphs and real multiracial faces. The motivation for this work was similar to that of Gaither et al. (2019), who compared classifications of real multiracial and computer-generated (FaceGen) multiracial faces and found that evidence for hypodescent was more pronounced with real faces.

In the current research, we compare classifications of morphed biracial and real biracial faces on a larger scale than has been possible in the past, using a newly developed database of multiracial faces that greatly expands the number of available stimuli (Ma, Kantner, & Wittenbrink, 2020). Although we are aware of other studies that used both morphs and real multiracial faces (Chen & Hamilton, 2012; Chen et al., 2014; Chen, Pauker, et al., 2018), these studies were not specifically designed to test for differences between the two face types, and thus included relatively few exemplars of either type. Nicolas et al. (2019) directly compared the emergent race phenomenon (the tendency for multiracial faces to be categorized as neither multiracial nor as either of their parent races) in morphed versus real biracial faces, using 10 exemplars of each type and finding no difference in emergent race classifications between the two biracial stimulus types. Here, we test a larger sample of faces than has been used previously (88 faces of each type), allowing for increased generalizability. Study 1 compares racial categorizations of morphed multiracial faces and oval-masked, real multiracial faces. Study 2 compares racial categorizations of morphed multiracial faces, oval-masked, real multiracial faces, and unmasked, real multiracial faces. Finally, Study 3 attempts to identify, using mediation analysis, what might explain differences in categorization rates between morphed multiracial faces and oval-masked, real multiracial faces. Data and analysis scripts for these studies are available for download at the Open Science Framework [https://osf.io/sa9cd/]

Study 1

Study 1 compares racial categorizations of morphed multiracial faces and oval-masked, real multiracial faces. As we noted above, it is conventional to remove hair from morphed images, because morphing procedures produce very hazy, translucent, and artificial looking hair. Thus, to make these two stimulus types comparable, we opted to oval-mask all of the faces. If morphing can produce stimuli that perceptually represent real multiracial faces, we expect that categorization of morphed multiracial faces and oval-masked, real multiracial faces will yield equivalent rates of categorizations as multiracial.

Method

Participants

Existing research suggests that mixed effects models involving random effects for both participants and stimuli require 6,400 observations for a power of .80 (Adelman et al., 2014; Brysbaert & Stevens, 2018). Given that we had 88 stimuli per condition, this led us to recruit a minimum of 73 participants per condition. We recruited 218 individuals (123 male, 94 female, 1 other; M_{age} = 36.6, SD = 11.3) from Amazon's Mechanical Turk (mTurk). Participants self-identified as follows: 173 White, 17 Black, 11 Asian, 7 Latino, 9 Biracial/Multiracial, and 1 Native American. We included a captcha to prevent bots from completing the study. Participants were compensated $3.

Materials

Faces included images of 88 morphs and 88 real multiracial individuals (Ma, Kantner, & Wittenbrink, 2020). Images of real multiracial individuals were obtained from self-reported multiracials who agreed to be photographed as part of an expansion of the Chicago Face Database (CFD; Ma et al., 2015), a standardized database of facial stimuli. The real multiracial targets were asked to report the racial and ethnic backgrounds of their maternal and paternal grandparents, their parents, and themselves in a free-response format. These data can be obtained at www.chicagofaces.org. Critically, all targets listed more than one racial category when reporting their ancestry. Morphed faces were created from Black, Asian, and White individuals taken from the CFD. The faces were rated high in racial prototypicality and matched for face shape to create more realistic morphs (Steyvers, 1999). Based on these criteria, Asian faces were all East Asian in appearance. Notably, East Asians are prototypically Asian in the United States (Ma et al., 2018). Faces with distinguishing features (e.g., moles, scars) or hair covering a substantial portion of the face were excluded. Morphs were created by combining images of Black and White or Asian and White faces using FantaMorph software (FantaMorph 5 Deluxe edition; Abrosoft, 2018, www.fantamorph.com). Prominent facial features were mapped with points that the software aligned to create a new face that was a 50%-50% combination of the two faces. This procedure was used to create 88 morphs (22 female Asian/White, 22 female Black/White, 22 male Asian/White, and 22 male Black/White) all of which are on our Open Science Framework registry (OSF; https://osf.io/sa9cd/). Again, in keeping with convention (e.g., Chen, Pauker, et al., 2018; Chen & Hamilton, 2012; Gaither et al., 2019), we oval-masked both types of stimuli.

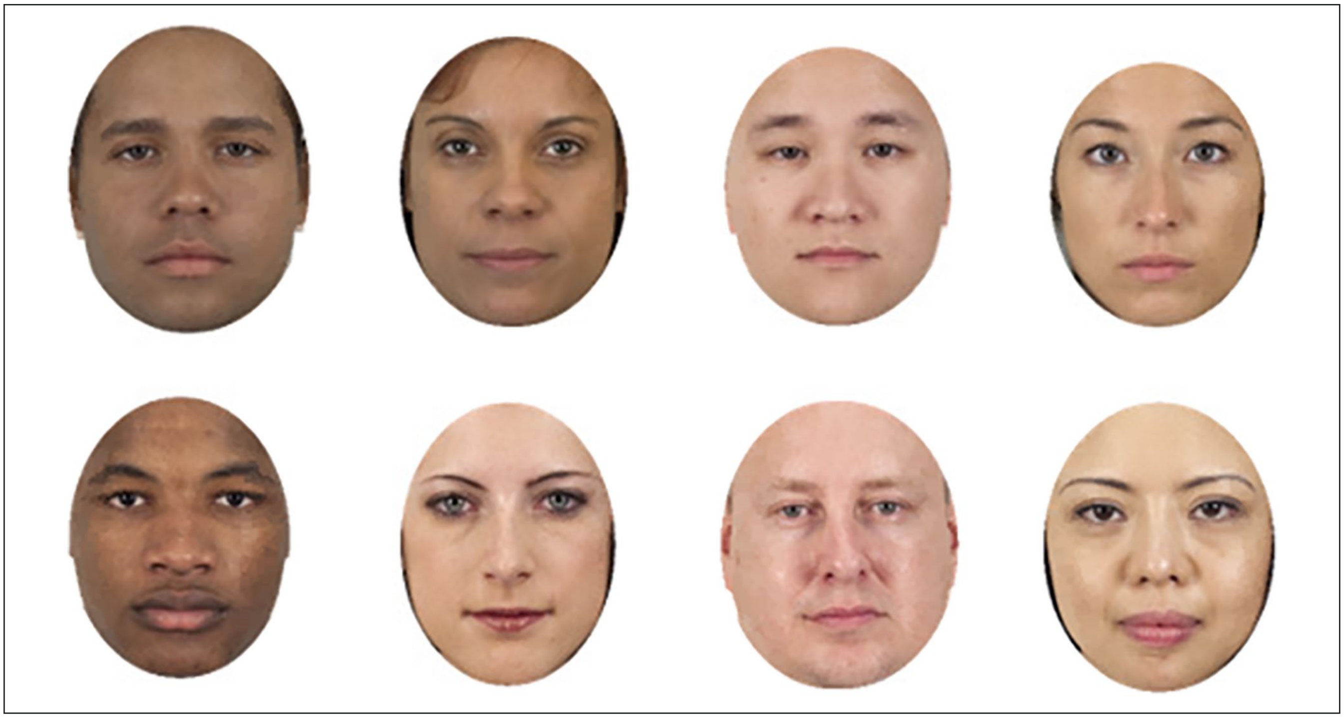

Figure 1. Stimuli used in Studies 1 and 2: Cropped morphed faces above and cropped monoracial (distracter) faces below.

Procedure and Design

Participants completed the study online. After giving consent, participants were told that they would be viewing pictures of people from different ethnic and racial backgrounds and that their task was to categorize them. Participants were randomly assigned to either categorize morphed multiracial faces or oval-masked, real multiracial faces. To conceal the primary objective of the study, we embedded multiracial faces among an equal number of self-identified monoracial distracters taken from the CFD. The monoracial distracters were oval-masked, to match the multiracial faces. The same set of real monoracial distracters were used for both conditions of the experiment to maintain parity in the decision context across conditions. While we considered this to be an important facet of the design, this approach introduced a potential methodological weakness, given that the morphed multiracial faces, against a backdrop of real monoracial distracters, were the only morphed stimuli in their condition, whereas the real multiracial faces judged in the other condition were not the only real faces in that condition. This aspect of the design would cause a concern to the extent that the morphed multiracial faces stood out relative to the distracters in ways that the real multiracial faces did not, because such distinctiveness might drive participants' classifications of the morphs. We addressed this concern in the process of selecting pairs of faces for morphing, with the goal of producing a set of morphs that would not clearly stand out to participants relative to the monoracial distracters. These faces are available for review at https://osf.io/sa9cd/ and a representative sample of them can be seen in Figure 1.

Each participant categorized 176 faces. Targets were shown in random order and one at a time with the following prompt ("What race/ethnicity is this person?"). Participants selected from nine options: Asian, Black, Hispanic or Latino, Middle Eastern, Native American, Pacific Islander, White, Biracial or Multiracial, and Other. This set of options was developed by combining race options from the U.S. Census and emergent races identified in previous multiracial face perception studies (Chen, Pauker et al., 2018). Participants provided basic demographics, were debriefed, and were thanked for their participation.

Results and Discussion

We compared the multiracial categorizations of morphed multiracial and oval-masked, real multiracial faces using a mixed model logistic regression. This type of model allowed us to account for differences between participants and stimuli (i.e., random factors). This analysis can be obtained on our OSF registry (https://osf.io/sa9cd/). In our full model, we specified a random intercept for participants, a random intercept for stimuli, and a fixed effect for condition (dummy coded 1 for morphs, 0 for real multiracial individuals). We present Odds Ratios, which can be interpreted as effect sizes, and also present Cohen's *d* for ease of interpretation (Cohen, 2013; Tabachnick et al., 2007). Morphed multiracials (M = .21, SD = .41, 95% confidence interval [CI] [.20, .22]) were

Table 1. Means and Standard Deviations for Accurate Categorizations of Monoracial Stimuli in Study 1.

Monoracial stimuli	*M*	*SD*	95% CI
Black	$.97_a$	.01	[.97, .97]
White	$.91_a$	.05	[.91, .92]
Asian	$.68_b$	.17	[.66, .71]
Latino	$.42_c$	.09	[.41, .43]

Note. Means with different subscripts differ $p < .05$. CI = confidence interval.

categorized as multiracial more often than the oval-masked, real multiracials ($M = .14$, $SD = .34$, 95% CI [.13, .15]), odds ratio [OR] = 2.07, $p < .001$, $d = .40$. For comparison, we also examined categorization data of monoracial distracters by calculating the number of accurate racial categorizations for each monoracial group (see Table 1), which we defined as instances in which participants' categorization matched the targets' self-identification. Across the two conditions, Black ($M = .97$, $SD = .01$) and White ($M = .91$, $SD = .05$) targets were accurately categorized at high rates. Categorizations of Asian ($M = .68$, $SD = .17$), and Latino ($M = .42$, $SD = .09$) targets were much lower, but still consistent with previous research showing that participants are more accurate at classifying monoracial than multiracial individuals (Chen et al., 2014; Chen & Hamilton, 2012; Gaither et al., 2019).

Study 2

In Study 1, multiracial morphs elicited higher rates of classifications as multiracial than images of real biracial people. These data suggest that morphs are not a perceptual stand-in for real multiracial faces; contrary to what one might have expected, they suggest that morphed multiracials more closely resemble people's representations of what biracials/multiracials look like than even real mixed-race faces. We further investigate the question of why morphs would look more multiracial than actual multiracials by collecting subjective ratings in Study 3.

In Study 2, we sought to replicate the findings from Study 1 while examining the effect of hair presence on multiracial classifications. As noted, morphs are limited by the fact that a key physical feature—hair—must be cropped from the image. Hair texture, color, length, volume, and style communicate much about group membership (Maclin & Malpass, 2003) and eliminating hair may limit the external validity of any oval-masked face stimulus in the context of a race classification experiment. This limitation was imposed on the oval-masked, real multiracial faces in Study 1 for parity with the morph stimuli. Study 2 thus offers a replication of Study 1 and adds a third group who categorized images of real biracials with hair. This design affords an estimate of (a) how the presence of hair impacts race classifications of real multiracial faces, and (b) whether morphs and real multiracial faces with hair elicit similar levels of multiracial categorizations.

Method

Participants

Existing research suggests that mixed effects models involving both participant and stimuli random effects require 6,400 observations for a power of .80 (Adelman et al., 2014; Brysbaert & Stevens, 2018). As in the previous study, this led us to recruit a minimum of 73 participants per condition. Participants were 246 mTurk workers (154 male, 90 female, 2 other; $M_{age} = 35.4$, $SD = 10.3$). Participants self-identified as follows: 181 White, 21 Black, 16 Asian, 17 Latino, 9 Biracial/Multiracial, and 2 Other. We included a captcha to prevent bots from completing the study. Participants were compensated $3.

Materials

Materials were identical to those used in Study 1, but included non-oval-masked images of the 88 real multiracial individuals.

Procedure and Design

The procedure followed that of Study 1, with the sole exception that participants were randomly assigned to one of three possible conditions: morphed multiracial faces; oval-masked, real multiracial faces; and unmasked, real multiracial faces. The monoracial distracter faces were oval-masked in the first two conditions, but unmasked in the third condition to match the multiracial faces.

Results and Discussion

Once again, we conducted a mixed effects model to test for differences in multiracial categorizations between face types. In our model, we specified a random intercept for participants, a random intercept for stimuli, and two dummy codes (i.e., fixed effects) for condition. The dummy codes established the oval-masked, real multiracial faces as the baseline condition. There was a meaningful difference between the morphed multiracial faces ($M = .22$, $SD = .42$, 95% CI [.21, .23]) and the oval-masked, real multiracial faces ($M = .13$, $SD = .34$, 95% CI [.12, .14]), OR = 2.26, $p < .001$, $d = .45$. This model also revealed that unmasked, real multiracial faces ($M = .23$, $SD = .42$, 95% CI [.22, .24]) were categorized as multiracial significantly more than oval-masked, real multiracial faces, OR = 2.72, $p < .001$, $d = .55$. A second model, establishing the unmasked, real multiracial faces as the baseline, showed no difference between the morphs and unmasked, real multiracial faces, OR = 1.01, $p = .96$, $d = .01$. The finding that including hair for real biracial faces

Table 2. Means and Standard Deviations for Accurate Categorizations of Monoracial Stimuli in Study 2.

Monoracial stimuli	*M*	*SD*	95% CI
Black	$.96_a$	.02	[.96, .97]
White	$.90_b$	.05	[.90, .91]
Asian	$.68_c$	.16	[.66, .70]
Latino	$.46_d$	.11	[.44, .47]

Note. Means with different subscripts differ $p < .05$. CI = confidence interval.

cues multiracial categorizations is in line with previous research showing that hair is an important cue in race categorizations (Maclin & Malpass, 2003). Given the significance of hair in making racial classifications, it is surprising that morphs (which do not have hair) should be judged as multiracial at the same rate as real mixed-race people with the hair included. Clearly facial features alone communicate multiracialism for morphs. Which features shape perception for real mixed-race faces remains unknown.

To replicate our findings in Study 1, we analyzed the categorizations for the monoracial distracters (Table 2). As before, we found that Black ($M = .96$, $SD = .02$) and White ($M = .90$, $SD = .05$) monoracial targets were accurately categorized across all conditions. Again, Asian ($M = .68$, $SD = .16$) and Latino ($M = .46$, $SD = .11$) targets were accurately categorized at lower rates (see Table 2). Our findings with the monoracial targets replicate Study 1 and existing literature (Chen et al., 2014; Chen & Hamilton, 2012; Gaither et al., 2019).

Study 3

Studies 1 and 2 reveal that morphed multiracials are categorized as multiracial significantly more than oval-masked, real multiracials, but at comparable levels to unmasked, real multiracial faces. To understand this pattern of findings, Study 3 measured subjective face characteristics that might promote multiracial classifications. We consulted the literature and examined the stimuli for potentially relevant differences across morphs, multiracial faces with hair, and multiracial faces without hair. One documented difference concerns the amount of pixelation in morphed compared with unmorphed stimuli. Chen and Hamilton (2012) reported that morphs in their studies had less than half the pixel variation of real biracial face images. Moreover, lower pixelation corresponds with greater perceived attractiveness (Matts et al., 2007). Also, as we noted in the Introduction, morphs are more average looking, which corresponds with attractiveness. Based on these two pieces of evidence, we hypothesized that attractiveness might explain increased multiracial categorizations for the morph stimuli. We also identified several other characteristics that might differentiate morphs, including the fact that morphs may differ in terms of how unusual/distinctive, striking, emotionally expressive, and photoshopped they appear relative to real faces. Our goal was to obtain subjective ratings of each of our three stimulus types and test for mean-level differences on these characteristics, as well as test whether these subjective ratings statistically mediated categorization differences across stimulus type.

Method

Participants

As in the prior two studies, based on existing research and the number of stimuli available, we recruited a minimum of 73 participants per condition (Adelman et al., 2014; Brysbaert & Stevens, 2018). Participants were 1,001 (564 male, 434 female, 3 Other; $M_{age} = 36.04$, $SD = 11.26$) from mTurk. Participants self-identified as follows: 690 White, 127 Black, 81 Asian, 60 Latino, 36 Biracial/Multiracial, 6 who identified as Other, and 1 opted not to respond. We included a captcha to prevent bots from completing the study. Participants were compensated $1.25.

Materials

Materials and categorization data from Study 2 were used in the current study in a target-level analysis.

Procedure and Design

The study included 88 targets of each type: morphed multiracial faces; oval-masked, real multiracial faces; and unmasked, real multiracial faces. Participants rated only one type of face. Because completing all of the ratings for all 88 faces would be onerous for an individual participant, we presented each participant with 22 randomly selected faces. Faces were presented one at a time and participants rated each target in terms of how attractive, unusual/distinct, striking, neutral of facial expression, and "photoshopped" the face was (1 = *not at all*, 7 = *extremely*). Participants were instructed that "photoshopped" referred to how much software-editing was used on the face and were explicitly told that it did not refer to any type of cropping. After completing the ratings, participants provided basic demographic information, were debriefed, and were thanked for their participation.

Results and Discussion

Similar to the first two studies, mixed model logistic regressions were run simultaneously. The mixed model analysis can be found at our OSF registry (https://osf.io/sa9cd/). We refer readers to Table 3 for full descriptive statistics and inferential tests, but highlight a few key findings here. A series of mixed effects models revealed significant differences across the three face types in perceived attractiveness,

Table 3. Means and Standard Deviations for the Evaluations Made by Participants for Each Type of Stimuli.

	Stimuli					
	Morph		Unmasked real		Oval-masked real	
DV	*M*	*SD*	*M*	*SD*	*M*	*SD*
Attractive	3.53_a	1.75	3.40_{ab}	1.72	3.24_b	1.76
Unusual/distinct	3.30_a	1.72	3.27_a	1.71	3.14_a	1.73
Striking	3.13_a	1.76	2.98_a	1.72	2.93_a	1.77
Neutrality of facial expression	4.51_a	1.56	4.37_b	1.56	4.28_b	1.63
Photoshopped	2.76_a	1.81	2.37_b	1.68	2.36_b	1.72

Note. Within rows, means with subscripts differ $p < .05$.

neutrality of facial expression, and the degree to which faces were perceived as photoshopped. Notably, morphs were perceived as significantly more attractive, $B = 0.30$, $p = .03$, having more neutral facial expressions, $B = 0.23$, $p = .01$, and being more photoshopped, $B = 0.40$, $p < .001$, than the oval-masked, real multiracial stimuli. Morphs were also perceived as more unusual/distinct than the oval-masked, real multiracial individuals but this was only trending toward significance, $B = 0.17$, $p < .10$. In comparison, there was only one significant difference between the morphs and the unmasked, real multiracial faces, such that the morphs were perceived as more photoshopped than the real multiracial faces, $B = 0.38$, $p < .001$. There were no meaningful differences found between the two types of real multiracial stimuli, all *p*s $> .17$.

Next, we conducted a statistical mediation analysis (Hayes & Preacher, 2014) using two sets of dummy codes. The codes used for the first contrast were 1, 0, and 0 and for the second contrast the codes were 0, 0, and 1, for the morphed multiracial, oval-masked real multiracial, and unmasked real multiracial faces. These allowed us to compare morphed multiracial to oval-masked, real multiracial faces and unmasked, real multiracial to oval-masked, real multiracial faces. These codes were submitted into a stepwise regression and were entered in Block 1. Categorization data from Study 2 were used here to carry out the target-level analyses. First, we tested the unique indirect effects of how attractive, unusual/distinctive, striking, neutral, and photoshopped the faces were by entering them in Block 2. Of these, there was no evidence that ratings of how attractive, striking, neutral, and photoshopped related to multiracial categorization, *t*s ≤ -1.12, $p \geq .26$; however, unusual/distinctive ratings significantly related to multiracial classifications, $t(263) = 3.91$, $p < .001$, $d = .48$. What precisely this unusualness and distinctiveness connoted to participants is debatable, because we included dummy codes for both morphed multiracial, oval-masked real multiracial, and unmasked real multiracial faces, and what is "unusual" may differ across comparisons. For example, morphed faces may be unusual because they appear artificial and do not include hair, whereas unmasked, real multiracial faces may be judged as unusual or distinctive, because they are rare in the population and stand out in a statistical sense. However, we do note that recent research reports a relationship between unusualness and categorizations of a target as multiracial (Ma, Kantner, & Wittenbrink, 2020). This is a question meriting future investigation.

General Discussion

We tested whether morphing, a common method for creating multiracial face stimuli, produces faces that yield race classification patterns similar to those of the real multiracial face stimuli they presumably represent. We compared categorizations of morphs to real faces that were both oval-masked to match the morphs (i.e., did not include hair) and were not oval-masked (i.e., included hair). Three major patterns were observed: first, morphs were judged to be multiracial more often than real multiracial people when the real faces excluded hair; second, real faces with hair were judged to be multiracial more often than real faces without hair; third, morphs and real faces with hair were judged to be multiracial equally often. To investigate what might be driving these differences, we used an exploratory mediational analysis on subjective rating data collected in Study 3. This analysis revealed that unusualness/distinctiveness may serve as a useful cue for multiracial categorizations, leading to higher rates for morphs and real faces with hair.

Our findings suggest that the appropriateness of morphs as a perceptual representation of real multiracial faces depends on the researcher's goals (see Gaither et al., 2019, for related discussion of the use of real versus FaceGen-created multiracial face stimuli). If the purpose of morphs is to create faces that can stand in for real, mixed-race faces in research, these data suggest that morphs are *not* an equivalent class of stimuli, because they actually garner higher rates of multiracial classifications than real, mixed-race faces (when matched for lack of hair). However, if a researcher aims to operationalize multiracialism with a stimulus that best communicates multiracial category membership (i.e., that is most commonly viewed as multiracial), morphs do appear to be more effective than oval-masked, real multiracials and are comparable to unmasked, real multiracials. By this standard, these studies lend support to the continued use of facial morphing. If one's research goal is to use multiracial face stimuli that are high in *external validity*, real faces, rather than morphs, should be used. By contrast, if one's goal is to use stimuli that are high in *construct validity* (i.e., stimuli people will perceive as multiracial), morphs appear to be a satisfactory option.

Although we believe these results carry implications for the use of facial morphs to operationalize multiracialism, we acknowledge weaknesses in the current research that prompt additional investigation. First, we cannot isolate the effects of morphing multiracial faces from morphing more generally. In the current studies, we chose to embed the morphed and real multiracial faces among real monoracial faces. Participants assigned to the morphed multiracial faces condition may have used artifacts associated with morphing as cues to multiracialism, which presumably would be less likely to occur were the monoracial distracters also morphed. If this was the case, then this could exaggerate the categorizations of morphed faces as multiracial. Our choice to use real monoracial faces as foils was to maintain a consistent set of distracters across real and morphed face conditions; however, including morphed monoracial face foils (e.g., White morphs faces created from morphing two monoracial White targets; c.f., Nicolas et al., 2019) would allow us to parcel out the effect of morphing on multiracial categorizations. A second concern pertains to the comparability of morphed multiracial stimuli and real multiracial faces. Whereas we used monoracial parent faces that were high in racial prototypicality to create morphed faces, we have no information about the racial prototypicality of the biological parents of the real multiracial people whose faces were used in the current studies. We might safely assume that these real-life, biological parents were less racially prototypic than the exemplary faces used to create the morphs, simply because we sampled from the tail-end of racial prototypicality. As such, it is likely that the morphs more precisely occupy the mid-point between two monoracial faces than real multiracial individuals, which could have been perceptually more heterogeneous to perceivers. It is possible that the averageness of morphed faces signaled participants to categorize these faces as multiracial at higher rates. We do not view this interpretation as problematic per se, because perceivers may mentally represent multiracial faces as more average looking and use averageness as a cue to multiracial classification. The averageness and variability among the morphed versus the real multiracial faces raises interesting questions worthy of future investigation. We can envision follow-up studies that systematically vary the average morph contribution of each parent face as well as the variance in the set of face stimuli to explore how expectations about the averageness of the faces and the genetic variability of mixed-race faces contribute to different aspects of multiracial face perception (Lewis, 2010). On the other hand, it is also possible that the real multriracial people whose images were used in this study were more likely to volunteer to be photographed because they were more racially ambiguous. That is, mixed-race individuals who are more racially ambiguous may be more likely to identify as multiracial and come in for a study seeking multiracial people. We could envision interesting follow-up studies exploring how racial prototypicality impacts identification with one's race or racial categories.

The current research focuses on methodological issues, but we stress the theoretical ramifications of the findings. First, as we allude to above, any theory of multiracial face perception can be informed by (and must account for) the finding that observers classify more artificial than real multiracial faces as multiracial. To us, this suggests that there may be a significant disconnect between what multiracial faces actually look like and perceivers' mental representations or expectations of multiracial faces. This dissociation could be driven by perceiver features (e.g., familiarity with multiracial individuals, individual differences, racial or ethnic background), some of which have been investigated already in the literature (Chen, de Paula Couto, et al., 2018; Ho et al., 2013; Krosch et al., 2013), as well as stimulus factors (e.g., differences in which features or combinations of features multiracial faces possess), which are less investigated. Second, our studies demonstrate that hair is an important (but virtually unexamined) cue to multiracial status for observers. Hair has been examined as an important factor for some aspects of social categorization (Bigler et al., 1997; Martin & Macrae, 2007), but is understudied in the domain of race categorization. Third, the subjective "unusualness" of a face is an important predictor of its classification as multiracial. The apparent association between judgments of faces as multiracial and unusual has also been observed in other work using different procedures (Ma, Kantner, & Wittenbrink, 2020) suggesting the robustness of this relationship. Although the current data do not allow us to drill down into what "unusual" means for perceivers, this observation offers a potential avenue for future research.

One major concern for the field may be finding reasonable operationalizations for multiracialism. Here, morphs were equal to unmasked, real multiracial faces at eliciting multiracial classifications, but neither operationalization is especially effective, considering that these faces are only categorized as such between 20% and 25% of the time. By comparison, researchers who study similar processes using monoracial faces select faces that are perceived as intended by the researcher at least 90% of the time (Ma & Correll, 2011; Maddox & Gray, 2002). The low rates of multiracial categorization observed here and in other studies (Chen & Hamilton, 2012; Chen et al., 2014; Gaither et al., 2019) pose a threat to the construct validity of morphs, and even real multiracial faces. Furthermore, follow-up analyses suggest that it may be difficult to find stimuli that are unquestionably multiracial. Specifically, we conducted ancillary random effects analyses, which revealed significant variance components related to both stimuli and participants. Across the critical tests we report in these three studies, we found significant variance in multiracial classifications due to stimuli, Wald *Z*s $\geq$ 6.28, *p*s $\leq$ .001. This illustrates considerable differences across multiracial stimuli in the extent to which they are perceived as multiracial. To us, this suggests that researchers may need to conduct extensive pre-testing to identify targets that are the most likely to be perceptually multiracial. However, we also observed significant variation across participants, Wald *Z*s ≤ -15.63, *p*s $\leq$ .001, suggesting that perceptions are likely to vary widely between individuals. This finding is consistent with previous research documenting a host of individual differences that impact biracial categorizations (e.g., Chen et al., 2014; Chen, de Paula Couto, et al., 2018). Taken together, these results suggest that it may be very difficult to produce a set of multiracial face stimuli that are highly categorized as multiracial (if this is the researcher's goal), because these categorizations depend on the stimuli and on individual differences. Relatedly, it is important to note that the images of the real multiracial faces and the monoracial faces used to create the morphs were all gathered and normed in the United States, and that the participants in the current studies were drawn from the United States. It is almost certainly the case that there exist important cultural differences in what constitutes a multiracial face, such that the current results may be limited in their generalizability. Recent research by Chen and colleagues (2018), for example, demonstrates that perceivers in the United States and Brazil differ in multiracial face categorization and the features that drive categorization judgments. We believe more research documenting and exploring cross-cultural differences is merited and that it is important to take the cultural context of the current studies into consideration when interpreting the reported effects.

Relatedly, we point out that the choice sets used in these race-categorization tasks can greatly impact the results (Chen, Pauker, et al., 2018; Gaither et al., 2019). Recall from the Introduction, for example, a stimulus could be categorized as Black if there is not an option to categorize the same stimulus as multiracial. Likewise, that stimulus face could be categorized as Latino if permitted (Chen, Pauker, et al., 2018; Feliciano, 2016; Tskhay & Rule, 2015). Here, we included Hispanic/Latino and Middle Eastern, which are likely to draw categorizations away from multiracial to provide a conservative test of morphs. Had we used a restricted choice set and included only Black, White, and multiracial, we expect that multiracial classifications would have increased dramatically. However, we also propose that the labels provided can have top-down influences on the way that a target is perceived. Perhaps label information may offer clarification for participants struggling to make a categorization when the signal is ambiguous. It is also possible that choice sets impact other psychological processes (e.g., motivates participants to be accurate, causes participants to be less rigid about racial categories), which in turn, affects categorization. A free response procedure, in which participants make race categorization judgments in the absence of experimenter-provided response options, avoids these response set considerations (Chen, Pauker, et al., 2018; Nicolas et al., 2019).

Despite using a categorization task here, we acknowledge that the insights gained through explicit categorization tasks may not necessarily reveal how multiracial individuals are mentally represented. Indeed, recent data from our lab using Multidimensional Scaling procedures suggest that participants have categorical boundaries separating monoracial and multiracial individuals perceptually (Ma, Dunn, et al., 2020), and this may occur even though explicit categorization suggests no differentiation among stimuli. To us, this suggests an opportunity to explore what factors might move some of these perceptual processes from unconscious to conscious awareness.

Acknowledgments

We thank members of the Stereotyping and Intergroup Prejudice Lab and Jessica Shropshire.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Primary support for this work was provided by the National Science Foundation (#1749542) to the first two authors.

ORCID iD

Debbie S. Ma https://orcid.org/0000-0003-0585-6476

Supplemental Material

Supplemental material is available online with this article.

References

Abrosoft Co. (2018). *Fantamorph* (Version 5) [Computer software]. http://www.fantamorph.com/index.html

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., . . . Scaltritti, M. (2014). A behavioral database for masked form priming. *Behavior Research Methods*, *46*(4), 1052–1067.

Aspinall, P. J. (2018). What kind of mixed race/ethnicity data is needed for the 2020/21 global population census round: The cases of the UK, USA, and Canada. *Ethnic and Racial Studies*, *41*(11), 1990–2008. https://doi.org/10.1080/01419870.2017.1346267

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, *68*(3), 530–543. https://doi.org/10.1111/j.1467-8624.1997.tb01956.x

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9.

Chen, J. M., de Paula Couto, M. C. P., Sacco, A. M., & Dunham, Y. (2018). To be or not to be (black or multiracial or white) cultural variation in racial boundaries. *Social Psychological and Personality Science*, *9*(7), 763–772.

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, *48*, 152–164.

Chen, J. M., Moons, W. G., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. (2014). Motivation to control prejudice predicts categorization of multiracials. *Personality and Social Psychology Bulletin*, *40*, 590–603.

Chen, J. M., Norman, J. B., & Nam, Y. (2020). Broadening the stimulus set: Introducing the American Multiracial Faces Database. *Behavior Research Methods*, 1–19.

Chen, J. M., Pauker, K., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. (2018). Black + White = Not White: A minority bias in categorizations of Black-White multiracials. *Journal of Experimental Social Psychology*, *78*, 43–54. https://doi.org/10.1016/j.jesp.2018.05.002

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Feliciano, C. (2016). Shades of race: How phenotype and observer characteristics shape racial classification. *American Behavioral Scientist*, *60*(4), 390–419.

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2019). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social Psychology*, *159*, 592–610. https://doi.org/10.1080/00224545.2018.1538929

Gaither, S. E., Schultz, J. R., Pauker, K., Sommers, S. R., Maddox, K. B., & Ambady, N. (2013). Essentialist thinking predicts decrements in children's memory for racially-ambiguous faces. *Developmental Psychology*, *50*, 482–488.

Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin & Review*, *10*, 149–156.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Ho, A. K., Sidanius, J., Cuddy, A. J., & Banaji, M. R. (2013). Status boundary enforcement and the categorization of Black-White biracials. *Journal of Experimental Social Psychology*, *49*, 940–943.

Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. R. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology*, *100*(3), 492–506.

Krosch, A. R., Berntsen, L., Amodio, D. M., Jost, J. T., & Van Bavel, J. J. (2013). On the ideology of hypodescent: Political conservatism predicts categorization of racially ambiguous faces as Black. *Journal of Experimental Social Psychology*, *49*, 1196–1203.

Lewis, M. B. (2010). Why are mixed-race people perceived as more attractive? *Perception*, *39*(1), 136–138.

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*, 1122–1135.

Ma, D. S., Dunn, S., Benitez, J., & Kantner, J. (2020). Are biracials a cognitively distinct category? Mapping Biracials in psychological space. Manuscript Submitted for Publication.

Ma, D. S., Kantner, J., & Wittenbrink, B. (2020). Chicago Face Database: Multiracial expansion. *Behavior Research Methods*, 1–12.

Ma, D. S., Koltai, K., McManus, R. M., Bernhardt, A., Correll, J., & Wittenbrink, B. (2018). Race signaling features: Identifying markers of racial prototypicality among Asians, Blacks, Latinos, and Whites. *Social Cognition*, *36*(6), 603–625.

MacLin, O. H., & Malpass, R. S. (2003). The ambiguous-race face illusion. *Perception*, *32*(2), 249–252. https://doi.org/10.1068/p5046

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.

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*(5), 806–816. https://doi.org/10.1002/ejsp.445

Matts, P. J., Fink, B., Grammer, K., & Burquest, M. (2007). Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*, *57*(6), 977–984. https://doi.org/10.1016/j.jaad.2007.07.040

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35.

Morning, A. (2012). Multiraciality and census classification in global perspective. In *International perspectives on racial and ethnic mixedness and mixing* (pp. 10–22). Routledge. https://doi.org/10.4324/9780203118344

Nicolas, G., Skinner, A. L., & Dickter, C. L. (2019). Other than the sum: Hispanic and Middle Eastern categorizations of Black–White mixed-race faces. *Social Psychological and Personality Science*, *10*, 532–541.

Parker, K., Horowitz, J. M., Morin, R., & Lopez, M. H. (2015, June 11). *Multiracial in America: Proud, diverse and growing in numbers*. https://www.pewsocialtrends.org/2015/06/11/multi-racial-in-america/

Pauker, K., & Ambady, N. (2009). Multiracial faces: How categorization affects memory at the boundaries of race. *Journal of Science Issues*, *65*, 69–86.

Pauker, K., Ambady, N., & Freeman, J. B. (2013). The power of identity to motivate face memory in biracial individuals. *Social Cognition*, *31*, 780–791.

Peery, D., & Bodenhausen, G. V. (2008). Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science*, *19*, 973–977.

Steyvers, M. (1999). Morphing techniques for manipulating face images. *Behavior Research Methods, Instruments, & Computers*, *31*(2), 359–369.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481–498). Pearson.

Tskhay, K. O., & Rule, N. O. (2015). Semantic information influences race categorization from faces. *Personality and Social Psychology Bulletin*, *41*, 769–778.

Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, *94*(6), 1033–1047. https://doi.org/10.1037/0022-3514.94.6.1033