**SUPPLEMENTARY MATERIALS**

**Data & materials online:** https://osf.io/6h2rf/?view\_only=ec3d24689fd640fbad41a3a747514e6b

**Study 1**

**Stimuli creation**

We collected White and Black facial stimuli from the Average Faces database (DeBruine & Tiddeman, 2016), Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), Face Database (Minear & Park, 2004), and Color FERET database (Phillips, Wechsler, Huang, & Rauss, 1998). Several databases were used to obtain enough high-resolution photos of faces of different races and both genders.

Once we had a large set of high-resolution photos (*N*= 238), we pretested them with Mturk workers (*N* = 174; each worker rated 30 randomly presented faces). All 238 faces in this pre-morphing set were categorized as White or Black over 95% of the time, and were given scores within .5 of the extremes in a 7-point scale ranging from Completely Black to Completely White. All faces were in color to increase ecological validity (e.g., Stepanova & Strube, 2009).

For Study 1, we included 5 levels of Black-to-White phenotypicality: 0% Black, 25% Black, 50% Black, 75% Black, and 100% Black. A subset of 160 White and Black faces was selected and divided into 5 groups of 32 faces that were matched on gender (50%), racial phenotypicality, age, attractiveness, and database of origin. Faces were cropped in a standardized way to remove less stable or irrelevant features (e.g., hair, background). Pairs of faces within each of these groups were then morphed to different degrees using the online Psychomorph software (DeBruine & Tiddeman, 2016). Two of the groups contained faces of the same race (both Black or both White), which were morphed to 50% in pairs, resulting in 8 Black and 8 White faces. This step controlled for main effects of morphing, since even our monoracial faces were the result of a morphing algorithm. For the other three groups, a White and a Black face were morphed to be 25% White and 75% Black, 50% White/Black, and 75% White and 25% Black.

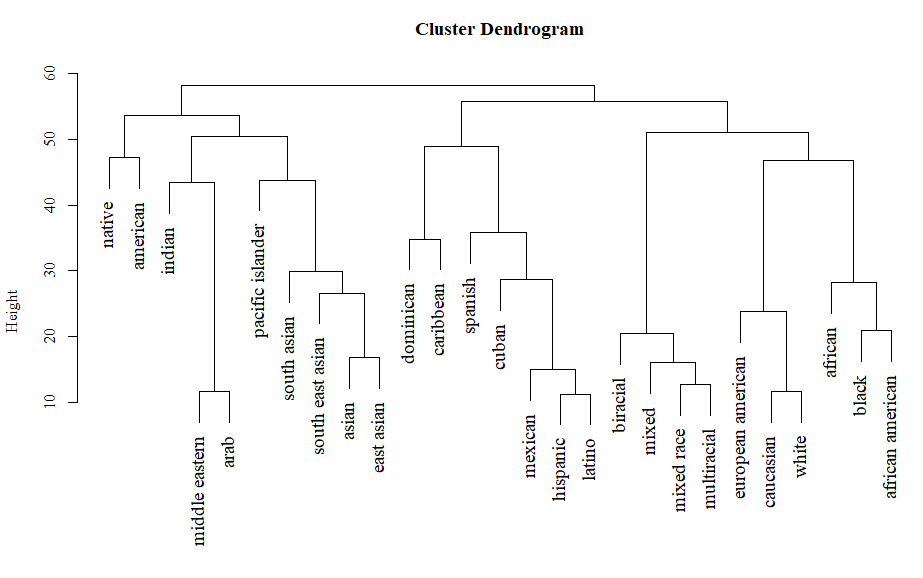
**Categorization speed manipulation**

We included a between-subjects manipulation of categorization speed that asked participants to either make categorizations as fast as possible or after careful deliberation. A previous study (Peery & Bodenhausen, 2008) using this manipulation suggests that speeded categorizations lead to more Black categorizations while deliberative categorizations lead to Multiracial categorizations. However, Peery and Bodenhausen used forced-choice tasks, and changed tasks simultaneously with categorization speed, leading us to suspect that the results would not apply to a free-response task and would be better explained by task differences than by categorization speed. Since this manipulation may provide additional insights into how the use of forced-choice tasks could affect inferences, we included it in the study. However, given the conflict between previous results and our untested expectations about task effects, we consider these results exploratory.

**Results**

The following R packages were used for the main analyses: lme4 (Bates et al., 2014), lmerTest (Kuznetsova et al., 2016), lsmeans (Lenth, 2016), and car (Fox & Weisberg, 2011). See online repository for code.

**Sorting task.** Results were analyzed using hierarchical clustering (complete method; see Figure S1). We used these data to code the free-response task, resulting in Black, White, Multiracial, Hispanic, Middle-Eastern, and Other category codes. A small number of responses from Study 1 (2%) were labels not included in the sorting task, which the researchers coded at their discretion (see Table S1), although excluding these responses did not change results. A cluster analysis using only the data from participants in the dichotomous, trichotomous, and polytomous tasks (i.e., the participants for whom the coding based on the cluster analysis did not apply) was nearly identical and all of the same clusters for relevant categories emerged.



*Figure S1.* Hierarchical cluster analysis for racial group labels. More similar labels join sooner on the hierarchy.

Table S1

*Participants’ Responses With Their Corresponding Coding and Frequency.*

|  |  |  |
| --- | --- | --- |
| Response | Coded race | Frequency |
| African descent | Black | 4 |
| African | Black | 82 |
| African American | Black | 696 |
| Black | Black | 1910 |
| European | White | 1 |
| Irish | White | 1 |
| Dutch | White | 6 |
| German | White | 7 |
| Italian | White | 7 |
| Swedish | White | 10 |
| French | White | 15 |
| Caucasian | White | 497 |
| White | White | 2050 |
| Half White half Black | Multiracial | 8 |
| Mixed | Multiracial | 39 |
| Biracial | Multiracial | 41 |
| Dominican | Alternative - Hispanic | 4 |
| Puerto Rican | Alternative - Hispanic | 6 |
| Mexican | Alternative - Hispanic | 18 |
| Latina | Alternative - Hispanic | 26 |
| Spanish | Alternative - Hispanic | 36 |
| Latino | Alternative - Hispanic | 50 |
| Hispanic | Alternative -Hispanic | 356 |
| Muslim | Alternative - Middle eastern | 1 |
| Persian | Alternative - Middle eastern | 1 |
| Egyptian | Alternative - Middle eastern | 6 |
| Arab | Alternative - Middle eastern | 16 |
| Middle eastern | Alternative - Middle eastern | 106 |
| Central Asian | Alternative - Other | 1 |
| Filipino | Alternative - Other | 1 |
| Mediterranean | Alternative - Other | 1 |
| Ethiopian | Alternative - Other | 2 |
| Other | Alternative - Other | 2 |
| American | Alternative - Other | 3 |
| Brazilian | Alternative - Other | 3 |
| Canadian | Alternative - Other | 3 |
| Eastern European | Alternative - Other | 3 |
| Pacific islander | Alternative - Other | 3 |
| Native American | Alternative - Other | 4 |
| South Asian | Alternative - Other | 5 |
| Asian | Alternative - Other | 9 |
| Brown | Alternative - Other | 21 |
| Indian | Alternative - Other | 32 |

**Black categorizations.** We start by reporting results for the Black categorization data, which include all tasks. In addition to the tasks described in the main text, these analyses include the dichotomous task, which is the most widely used (see Nicolas & Skinner, 2017), and prompted participants to make either White or Black categorizations using the “s” and “l” keys (see online repository for exact wording of all instructions). We found a significant effect of task, *χ*²(3) = 60.19, *p* < .001, and phenotypicality, *χ*²(1) = 665.72, *p* < .001, with no interaction, *χ*²(3) = 1.88, *p* = .597.

Next, we focused on the pattern for the average of the 25%, 50%, and 75% faces. Our results indicated that for the mixed-race faces, the dichotomous task overestimated the number of Black categorizations (vs. free-response: *Z* = 4.22, *p* <.001). On the other hand, the trichotomous task underestimated the number of Black categorizations in comparison to the free-response tasks (*Z* = -4.13, *p* < .001). The polytomous and the free-response tasks did not differ in the number of Black categorizations (*Z* = 0.27, *p* = .992).

**White categorizations.** Next, we report the results for White categorizations, noting that there will be some redundancy for the dichotomous task results, given that White categorizations will be 1 – Black categorizations. As in the previous model, we found an effect of phenotypicality, *χ*²(1) = 679.44, *p* < .001, and task, *χ*²(3) = 89.58, *p* < .001, with no interaction, *χ*²(9) = 6.77, *p* = .080.

Again, to further break down these results, we focused on the average of the middle three levels of phenotypicality and compared across tasks. These results indicated that the dichotomous task overestimated the number of White categorizations (vs. free-response: *Z* = 5.41, *p* <.001). In contrast, the trichotomous task underestimated the number of White categorizations in comparison to the free-response task (*Z* = -4.12, *p* < .001). The polytomous and the free-response tasks did not differ in the number of White categorizations (*Z* = -1.00, *p* = .750).

**Alternative categorizations.** If we consider the polytomous task’s “Neither Black nor White” configuration as a proxy for alternative categorizations, we can compare this rate to the free-response task’s. This analysis revealed the expected significant quadratic effect (*b* = -0.40, *Z* = -9.37, *p* < .001), no significant effect of task, *χ*²(1) = 0.17, *p* = .684, and a significant interaction, *χ*²(2) = 8.45, *p* = .015. A contrast between the tasks for the ambiguous faces (25%, 50%, 75%, as in other analyses) was not significant (*Z* = 1.03, *p* = .302). The interaction reflected a significant difference for 25% Black faces, which were slightly more likely to considered neither Black nor White in the polytomous (10%) than alternative in the free-response task (5%; *Z* = 2.12, *p* = .034).

**Other categorization results.** In the Free-response task, categorizations as Black or alternative did not significantly differ for the 50% faces (Z = -0.92, p = .360), and neither did categorizations as White or alternative (Z = - 0.37, p = .711).

**Categorization speed manipulation.** To explore the effect of categorization speed, we did a manipulation check by regressing (untransformed, minus < 300 and > 3000 ms responses) reaction times on speed condition (dummy coded) and confirmed its effectiveness, with faster responses in the fast (*M* = 878.32, SD = 466.07) than slow (*M* = 977.77, SD = 535.69) condition, *t*(27969) = -16.64, *p* < .001. However, there were no effects of categorization speed on categorizations (absolute *Z*s < 1.42, *p*s > .16).

These exploratory results suggest that it is possible previous significant effects attributed to categorization speed (e.g., Peery & Bodenhausen, 2008) were the result of task changes (see our Study 1 task effects) and not categorization speed. We note that there were other differences between our studies and previous research (e.g., we did not include ancestry information about targets; but see Study 2’s Multiracial salience manipulation) that could also help explain these differences. Additionally, stronger manipulations of categorization speed (e.g., limiting response times in the task rather than manipulating instructions; Chen & Hamilton, 2012) might show significant effects. However, we believe our results are in line with our argument that features of task should be carefully considered in studies of categorization of mixed-features stimuli.

**Reaction times.** Reaction times for the forced-choice tasks were recorded at key press, and for the free-response task at the time the participant started speaking. Reaction times (reported only for the fast condition, but same pattern is found for deliberative condition response times) were processed in the following way: first, reaction times below 300 ms and above 3000 ms were removed given our assumption that these values were most likely be the result of incorrect triggers from the microphone (either the participant made a non-verbal noise, did not speak loudly enough, etc.). Additionally, we removed values greater than 2 standard deviations above or below the mean, and log transformed the data. All results are exploratory.

Participants responded faster to simpler tasks (i.e., those with fewer response options) than to more complex tasks (i.e., those with more response options or the free naming task), *F*(3, 90.35) = 15.49, *p* < .001. We note that comparability between the free-naming and the other tasks is limited given potential differences between motor and speech responses. Across tasks, the response times fit a quadratic pattern well, with responses to targets at the center of the racial continuum taking longer to categorize, *b* = -.05, SE = .005, *t*(91.09) = -10.24, *p* < .001. This finding suggests that 50% faces were the most racially ambiguous to our participants.

**Study 2**

**Stimuli Creation**

Photographs of real faces were drawn from two sources: the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), and a series of real Black-White Multiracial faces collected by Pauker, Ambady, and Freeman (2013). Faces from the Chicago Face Database were selected to be unambiguously Black or White. We selected our stimuli to be among the better resolution photographs in the Pauker et al. (2013) stimuli set. Inclusion of real faces addresses multiple issues with the creation of morphed stimuli in our previous study, including potential confounds associated with digital manipulation, our decision to match pre-morphed faces on gender and other characteristics that may not reflect how racial mixing operates in real life, and, relatedly, making our findings more relevant to actual mixed-race targets by increasing ecological validity.

**Results**

**Other categorization results.** We found that alternative categorizations (84%) were significantly more common than White categorizations (16%), χ²(1) = 20.50, p < .001. Alternative categorizations (94%) were also more common than Black categorizations (6%), χ²(1) = 61.35, p < .001.

**Exploratory measures.** A genetic overlap measure asked participants to indicate what percentage of genes (0% - 100%) different combinations of races had in common. We found that the perceived genetic overlap between Whites and Black-White biracial (*M* = 68.1, *SE* = 21.5) was larger than between Whites and Hispanics (*M* = 60.9, *SE* = 29), *t*(112) = -3.99, *p* < .001. Similarly, perceived genetic overlap between Blacks and Black-White biracials (*M* = 68.0, *SE* = 22.8) was larger than between Blacks and Hispanics (*M* = 52.7, *SE* = 32.3), *t*(108) = -8.59, *p* < .001. Hispanics were seen as more genetically similar to Whites than to Blacks, *t*(108) = -4.68, *p* < .001.

We included questions testing the frequency with which participants made distinctions between ethnicity and race when thinking about others (“How frequently do you make a distinction between ethnicity and race when you think of others?”; 5-point scale ranging from never to always), asking whether they knew the distinction between ethnicity and race (Yes or No), and if they indicated that they did, asking for an explanation of the distinction (“Do you know what the difference between ethnicity and race is? If so, please let us know briefly what the distinction is.”). Results from these questionnaires indicated that only 53% of our respondents indicated knowing the definition of ethnicity. Of these, only 36.7% (*n* = 22; 19% of total sample) reported making a distinction between race and ethnicity when thinking about others relatively frequently (above the midpoint our 5-point scale ranging from never to always).

**Reaction times.** In Study 2, the same pattern of racial phenotypicality as in Study 1 emerged, with the 50% faces taking longer to be categorized (*M* = 7.33, *SE* = .02) than both the White (*M* = 6.96, *SE* = .02) and Black (*M* = 7.05, *SE* = .02) faces, *p*s< .001.

**References**

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology, 48*, 152–164. doi:10.1016/j.jesp.2011.10.005

DeBruine, L. M. & Tiddeman, B. P. (2016, June 13). WebMorph. Retrieved from http://www.webmorph.org

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression.* Thousand Oaks CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Kuznetsova, A., Brockhoff, P. B., & Bojesen Christensesn, R. H. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. Retrieved from: https://CRAN.R-project.org/package=lmerTest

Lenth, R., V. (2016). Least-Squares Means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. doi:10.18637/jss.v069.i01

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. doi:10.3758/s13428-014-0532-5

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633. doi:10.3758/BF03206543

Nicolas, G., & Skinner, A. L. (2017). Constructing race: How people categorize others and themselves in racial terms. In H. Cohen, & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (2nd ed., pp. 607–635). California: Elsevier Science

Pauker, K., Ambady, N., & Freeman, J. B. (2013). The power of identity to motivate face memory in biracial individuals. *Social Cognition, 31*, 780–791. doi:10.1521/soco.2013.31.6.780

Peery, D., & Bodenhausen, G. (2008). Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science, 19*, 973–977. doi:10.1111/j.1467-9280.2008.02185.x

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing, 16*, 295–306. doi:10.1016/S0262-8856(97)00070-X

Stepanova, E. V, & Strube, M. J. (2009). Making of a face: role of facial physiognomy, skin tone, and color presentation mode in evaluations of racial typicality. *The Journal of Social Psychology, 149*, 66–81. doi:10.3200/SOCP.149.1.66-81