Research Article

Psychological Science
1–16

DOI: 10.1177/0956797620920360
www.psychologicalscience.org/PS

# The Development of Intersectional Social Prototypes

TC

**Ryan F. Lei**[1,2], **Rachel A. Leshin**[1], **and Marjorie Rhodes**[1]
[1]Department of Psychology, New York University, and [2]Department of Psychology, Haverford College

**Abstract**

Race and gender information overlap to shape adults' representations of social categories. This overlap can lead to the psychological "invisibility" of people whose race and gender identities are perceived to have conflicting stereotypes. In the present research ($N = 249$), we examined when race begins to bias representations of gender across development. In Study 1, a speeded categorization task revealed that children were slower to categorize Black women as women, relative to their speed of categorizing White and Asian women as women and Black men as men. Children were also more likely to miscategorize Black women as men and less likely to stereotype Black women as feminine. Study 2 replicated these findings and provided evidence of a developmental shift in categorization speed. An omnibus analysis provided a high-powered test of this developmental hypothesis, revealing that target race begins biasing children's gender categorization around age 5 years. Implications for the development of social-category representation are discussed.

**Keywords**

race, gender, intersectionality, prototypes, development, social cognition, open data, open materials, preregistered

Received 7/11/19; Revision accepted 1/28/20

Black feminist theorists have long noted that Black women are excluded from feminist movements in the United States, illustrating a form of psychological invisibility (Crenshaw, 1989; Hull, Bell Scott, & Smith, 1982). This form of invisibility stems from a failure to account for intersectionality—the mutual reinforcement of intersecting systems of oppression (Purdie-Vaughns & Eibach, 2008). For example, the fight for gender equality often focuses on White women, whereas the fight for racial equality often focuses on Black men—thus discounting the unique discrimination faced by Black women.

Asian men can also experience a similar form of psychological invisibility (Galinsky, Hall, & Cuddy, 2013; Johnson, Freeman, & Pauker, 2012; Schug, Alt, & Klauer, 2015). For example, social media awareness movements such as #StarringJohnCho highlight the underrepresentation of Asian men in film (e.g., La Force, 2018). Indeed, a recent analysis of popular culture magazines found that Asian men and Black women were both underrepresented relative to Asian women and Black men, respectively (Schug, Alt, Lu, Gosin, & Fay, 2017).

The invisibility that Black women and Asian men often face is argued to stem from their being viewed as nonprototypical of either their race or gender categories—a form of invisibility termed *intersectional invisibility* (Purdie-Vaughns & Eibach, 2008). Intersectional invisibility manifests in basic cognitive processes (e.g., Goff, Thomas, & Jackson, 2008; Johnson et al., 2012). For example, in memory-confusion paradigms in which participants observe ostensible conversations among Black men, Black women, White men, and White women and are then asked to recall who said what, people are more likely to misremember whether a Black woman had been part of the conversation and to misattribute her statements than they are for other targets (Sesko & Biernat, 2010). Similar work has replicated these results for Asian men (Schug et al., 2015).

In the present work, we considered the possibility that intersectional invisibility might emerge early in development because of basic features of people's conceptual representations. A key feature of the conceptual

**Corresponding Author:**
Ryan F. Lei, Haverford College, Department of Psychology, 370 Lancaster Ave., Haverford, PA 19043
E-mail: rlei1@haverford.edu

representations of everyday categories is that exemplars vary in *prototypicality*. Some exemplars are viewed as highly representative of their category (e.g., *robins* for the category of bird), and others are viewed as more atypical (e.g., *penguins*; Rosch & Mervis, 1975). Category prototypes play a central role in numerous cognitive processes, including being recognized more quickly and recalled more easily as category members, remembered more readily, and thought to provide more generalizable information (for a review, see Mervis & Rosch, 1981). In the realm of social prototypes, overlapping stereotype content between the categories of *Black* and *male*, as well as of *Asian* and *female*, shape representations of these categories (Johnson et al., 2012). Because race and gender stereotypes develop early in childhood (Kinzler, Shutts, & Correll, 2010), here we tested the hypothesis that biased prototypes are an early emerging feature of conceptual structure, specifically that children also view Asian and White women as more representative of *women* than Black women, and Black and White men as more representative of *men* than Asian men.

We considered this early-emerging "gendered-race" hypothesis (Johnson et al., 2012) in the context of several developmental alternatives. One alternative is that children's gender categorizations are unbiased by race information because children often pay more attention to gender than race (Rhodes & Gelman, 2009; Shutts, Roben, & Spelke, 2013). Such a finding would suggest that young children first represent gender and race separately and develop integrated representations much later. As another alternative, race information may shape children's categorizations, but not in a manner akin to adult gendered-race prototypes. In this regard, we considered three possibilities. First, children might categorize both males and females faster and more accurately when they are of the same race as the participating child (consistent with *own-race bias* effects, e.g., Anzures et al., 2013). Second, children may categorize out-group males faster and more accurately than other groups (consistent with *out-group male* hypotheses, which suggest preferential processing of potential threats; e.g., Perszyk, Lei, Bodenhausen, Richeson, & Waxman, 2019). Third, children may categorize people who match themselves with respect to both gender and race faster and more accurately (consistent with preferential processing of people most "like me"; e.g., Meltzoff, 2013). Additionally, we considered how children's environments might shape these processes given that the diversity of individuals' neighborhoods can influence person perception from infancy onward (e.g., Mandalaywala, Ranger-Murdock, Amodio, & Rhodes, 2019). For a final, exploratory hypothesis, we examined the extent to which any potential effects may be due to differences in gender-linked facial features across racial groups.

To test these hypotheses, we used both a speeded categorization task (e.g., Bauer & Cox, 1998; Zarate & Smith, 1990) and an explicit stereotyping task. We tested how race biases gender categorization (rather than how gender biases racial categorization) because gender is highly salient to children and use of gender labels is commonplace throughout childhood (Liben & Bigler, 2002; Martin, Ruble, & Szkrybalo, 2002). We also included a sample of adult participants to validate our modified procedure and replicate prior gendered-race findings with adults using real faces, compared with the computer-generated stimuli used in previous adult studies, as well as with a verbal response (instead of button presses).

We preregistered our hypotheses and procedures on AsPredicted at http://aspredicted.org/blind.php?x=zw38pd, and our data and analysis scripts are available on the Open Science Framework at https://osf.io/qnye7. A video of our procedure is available to registered users at Databrary.org (http://doi.org/10.17910/b7.883).

## Study 1

### Method

***Participants.*** We recruited 127 children from the Children's Museum of Manhattan, a nonprofit organization in New York City.[1] Children ranged from 3 to 8 years old (age: $M$ = 5.74 years, $SD$ = 1.18), and our sample was mostly gender balanced (51% girls, 49% boys). Participants identified as White (46%), Asian (27%), or Black (26%). We determined our sample size from previous work (Johnson et al., 2012), which included approximately 100 participants using a conceptually similar paradigm. All participants completed the study using Inquisit 5 reaction time (RT) software (Millisecond, 2016), but we retained only those participants for whom we were able to code data more precisely using Datavyu coding software (Lingeman, Freeman, & Adolph, 2014). Because Datavyu coding required additional parental consent for video recording, our Datavyu sample was slightly smaller ($n$ = 102 children; mean age = 5.82 years; 52% girls; 46% White, 27% Asian, and 26% Black). We used Datavyu to code these data instead of the RT data provided by Inquisit because (a) Datavyu coding reflected the most accurate representation of children's RTs since the voice-to-key relay from Inquisit often picked up ambient noise from other children and adults in the museum testing space, and (b) we were able to code for categorization errors as an additional dependent variable.

***Adult comparison sample.*** Our adult comparison sample consisted of 63 Black ($n$ = 28; 39% female) and White ($n$ = 35; 46% female) adults (age: $M$ = 38.32 years, $SD$ = 14.56). We used the raw RTs from Inquisit (i.e., the direct voice-to-key relay) for adults because their testing space was much more controlled and thus less susceptible to environmental interferences. Adult participants were recruited via an online advertisement and paid $25 for their time.

***Materials and procedure.***

*Speeded categorization task.* Children were told that they were going to play a game to help the experimenter identify what was shown in different pictures. The experimenter explained that the pictures had been in a box but were dropped and now all mixed up. The child's job was to help the experimenter sort the pictures into the appropriate groups. This introduction served to set up the adapted speeded categorization task. For the task itself, children were asked to provide their categorization response verbally (e.g., Bauer & Cox, 1998) instead of by pressing a computer key (as is commonly done in adult studies) to reduce motor demands. Inquisit recorded the time (in milliseconds) to voice onset, or the amount of time that passed from when the image appeared after the fixation cross to when the participant made a sound. The speeded categorization task consisted of three parts—training, validation trials, and critical trials—and our dependent measures of interest were speed of categorization and categorization errors.

*Training.* To familiarize children with the process of categorizing stimuli into one of two dichotomous categories, we had children first complete a training exercise using physical pictures of rocks and trees. Children were told to classify the pictures as either rocks or trees, highlighting the broad category. Some images were more prototypical exemplars (e.g., a maple tree), whereas others were less prototypical (e.g., a bonsai tree). Participants completed six object categorizations (three of each type) and were instructed to classify each as quickly as possible. No feedback was provided throughout the task.

*Validation.* After finishing the training trials, children were told that they would perform the same activity but on the computer. Because a verbal speeded categorization task had not yet been used with children, we included a block of practice trials to validate the adapted task. Specifically, we wanted to ensure that this measure would reflect differences in categorization speed as a function of prototypicality. We used animal stimuli (i.e., pictures of typical and atypical birds and fish) for these validation trials because young children's graded representations of these animal categories are similar in many ways to those of adults (e.g., Foster-Hanson & Rhodes, 2019).

In this validation task, participants were told that they would be helping to identify whether each animal shown on the computer was a bird or a fish. We included six prototypical (e.g., goldfish for "fish," robin for "bird") and six nonprototypical (e.g., peacock for "bird," puffer fish for "fish") exemplars. The experimenter added that even if the child could recognize the picture as a more specific animal (e.g., turkey), the child should still use either the "bird" or "fish" label. As with the object-categorization trials, children were told to go "as fast as you can" and received no feedback.

*Experimental trials.* Immediately following the validation trials, children completed the critical experimental trials with people's faces. We used adult faces drawn from the Multi-Racial Mega-Resolution (MR2) database (Strohminger et al., 2016) that varied by both gender (male or female) and race (White, Asian, or Black). We randomly selected four faces for each race–gender combination. We used faces from this database because each face pictured had hair pulled tightly back, minimizing cues such as hair texture or length. Experimenters instructed children to classify each person as a boy or girl as quickly as possible. We chose "boy" and "girl" category labels instead of "man" and "woman" to equate the number of syllables and ease of word production. After children offered a verbal categorization, the computer screen advanced to the next image. Pictures remained on the screen for a total of 2 s; if no verbal categorization was made in that time, the program advanced to the next image after a short intertrial interval. Children saw two blocks of 24 pictures each presented in random order (for a total of 48 trials); there was a 30-s break between blocks.

***Datavyu coding of RT data.*** After the study was completed, two research assistants independently used Datavyu to code videos for the subset of participants for whom videos were available. Data-cleaning procedures were preregistered and are available at osf.io/3d92n/.

*RT coding.* Videos were coded at the level of the trial. Each participant video contained up to 60 trials (12 validation trials and 48 experimental trials), depending on how much of the study children completed. To code for RT, two research assistants watched each trial at one-eighth speed and coded the exact moment at which each stimulus appeared on screen. Then, research assistants increased the time to half speed and coded the moment of the child's first utterance of a response. We calculated RTs by taking the difference between stimulus onset and when the participant uttered a response. One research assistant watched all videos, and the other watched 25% of the videos for reliability coding. Overall, reliability was very good ($r$ = .85). We retained trials from the primary coder and then cleaned these RT data by removing trials with incorrect responses (so that only correct responses were maintained for the RT measure) as well as trials with RTs of less than 300 ms (which we classified as guesses, following Zarate & Smith, 1990).

*Miscategorization coding.* In addition to coding for RT, two research assistants also coded the accuracy of children's responses. Responses were coded as inaccurate if children responded with the incorrect dichotomous categorization (e.g., "boy" instead of "girl"). If children responded with a synonym of the category labels that we provided (e.g., "lady" instead of "girl"), responses were coded as correct. Reliability between coders was very good (κ = .85). When there were discrepancies between the primary and secondary coders, we retained the primary coder's judgments and used these for all subsequent data analysis.

***Explicit stereotype measure.*** After children completed the speeded categorization task, they completed a measure of gender stereotyping. Children first heard a story about an office with a new boss who needed help identifying all of the different people in the office. The new boss had notes describing various people in the office but no information about to whom the notes referred. On each trial, children were presented with a description of a stereotype in child-friendly language (e.g., for *gullible*, children heard, "This person believes everything others tell them, even if it's not true") and shown an array of six different faces—one for each race–gender combination. Children were asked to point to the face they thought embodied the trait described. Faces for this task were drawn from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) and were matched on racial prototypicality within trial. The traits included in this task consisted of six masculine and six feminine stereotypes (half of which were positively valenced and half of which were negatively valenced) obtained from a survey of the literature (Devine & Elliot, 1995; Holt & Ellis, 1998; Prentice & Carranza, 2002; Spence & Buckner, 2000; Spence, Heleich, & Stapp, 1973). The three positive feminine traits were "empathetic," "helpful," and "nice"; the three negative feminine traits were "shy," "gullible," and "subordinate." The three positive masculine stereotypes were "leaderlike," "athletic," and "competitive"; the three negative masculine stereotypes were "dominant," "mean," and "risk taking." We computed separate indices for positive feminine, negative feminine, positive masculine, and negative masculine traits. Scores ranged from 0 to 1, reflecting the proportion of times that children chose each target type for each trait type.

***Modifications for adult sample.*** Adult participants completed the same measures as children with a few modifications. First, adults saw a wider range of nonprototypical stimuli (e.g., sharks, seahorse, kiwi) during the validation trials to account for their greater experience with category variability. Second, adults were tested in an isolated room, which eliminated ambient noise from influencing RT in the speeded categorization task. Finally, we did not include a measure of categorization errors for adults because adult participants made virtually no errors.

***Parent questionnaire.*** While children were completing the study, parents filled out a questionnaire to provide a sense of the child's social environment. Specifically, parents were asked to report how many of the child's 10 closest friends were White, Black, Asian, Middle Eastern, or Hispanic. Of primary interest was the proportion of children's social network that was White, Black, or Asian. We also asked parents to provide their home zip code, which we used to pull neighborhood racial-diversity data via the U.S. Census Bureau (https://data.census.gov/cedsci/). Other measures included a six-item questionnaire about parents' multicultural or color-blind beliefs and demographic variables (e.g., parent race, gender, age, and political ideology).[2]

***Measuring faces for gender-linked facial-feature exploratory hypothesis.*** An additional exploratory hypothesis that could account for a gendered-race pattern of results is that there are real differences in gender-linked facial features between Black women and White and Asian women (i.e., a bottom-up process; Johnson et al., 2012). That is, perhaps one reason that Black women may be slower to be recognized as women is because they do, in fact, have more masculine features. To address this hypothesis, we used the ratings dimension guide provided by Ma et al. (2015) for the Chicago Face Database and measured dimensions they identified as related to gender—specifically, cheekbone prominence, "heartshapedness" of the face, eye shape, eye size, face length, and chin length. Ma et al. reported that face length and chin length negatively loaded on a gender factor, whereas the rest positively loaded on a gender factor. Following procedures outlined by Ma and colleagues, we had two coders rate each face independently. In line with Ma et al.'s results, our overall reliability between coders was good ($r$s ≥ .75). If individual face measurements differed by more than 20% from the mean of all faces, we flagged and remeasured those. Finally, we took the average of the two coders' measurements and used that for analyses.

***Analytic strategy.*** Per our preregistration, we tested for an interaction between target race (contrast coded as −1, 0, and 1 for Black, White, and Asian, respectively) and target gender (−0.5 and 0.5 for male and female, respectively). The same contrast codes were used for the child's own gender and racial group memberships. For exploratory analyses examining whether age interacted with target gender and target race for both dependent variables (RT and categorization error), age was treated as a

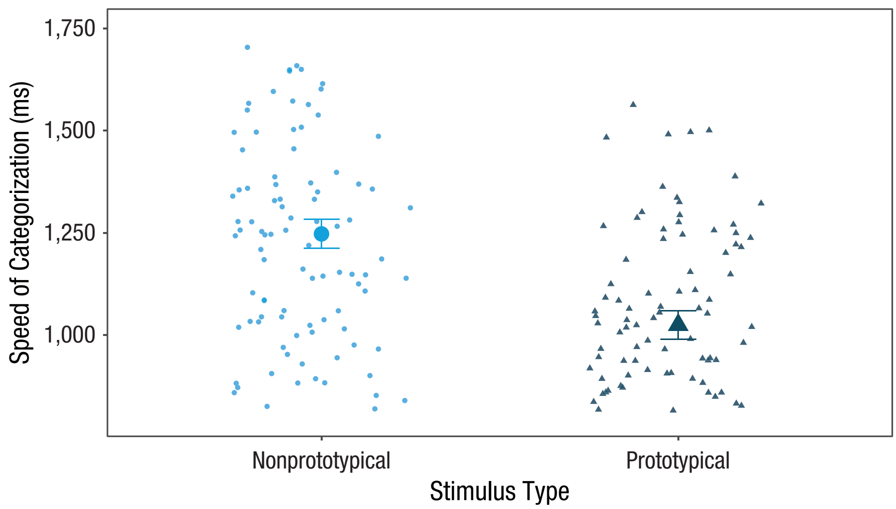


**Fig. 1.** Children's categorization speed for prototypical and nonprototypical animal exemplars (Study 1). Small shapes represent individual means, and large shapes represent group means. Error bars represent 95% confidence intervals.

continuous variable and centered. We report unstandardized coefficients to aid in the interpretation of results because they represent RT differences in milliseconds.

***Deviations from the preregistration.*** Although we aimed for a sample of 120 children with equal representation of White, Asian, and Black monoracial children, we were unable to reach the planned sample for Black and Asian children because of logistical changes by the museum testing site. Thus, the final sample for whom we had video-coded data consisted of 45 White children, 32 Asian children, and 24 Black children.

For our data analyses, we made a few changes to the planned protocol. First, following advice from experts in the field who use RTs to study conceptual structure, we opted to remove all error trials from the RT analyses and examine them separately. Analyses adhering to our preregistration (i.e., retaining and recoding error trials) are available in the Supplemental Material available online. Second, we opted to use linear mixed models instead of analyses of variance to account for the nested structure of the data. Within these linear mixed models, target race and gender, participant race and gender, and age were all entered as fixed effects, and we included a random intercept for participant to account for the repeated nature of the data. Third, for error trial data, we ran a negative binomial multilevel model because of overdispersion in the data. Fourth, we recoded our data to reflect own-race versus other-race RTs as a more precise and tractable way of testing our group-based hypotheses, rather than using either the four-way interaction specified or the particular contrast codes indicated for the out-group target male or "like-me" hypotheses. Finally, we coded our stereotype data on a scale ranging from 0 to 1 (i.e., reflecting the proportion of times children picked a given target for a given trait) to be more easily interpretable than our prespecified coding scheme.

## Results

***Validation-trial RTs.*** To validate our adapted speeded categorization task, we confirmed that children were faster to categorize prototypical animal stimuli (e.g., to categorize a robin as a bird; $M$ = 1,024 ms, $SE$ = 18) than to categorize nonprototypical stimuli (e.g., to categorize a chicken as a bird; $M$ = 1,247 ms, $SE$ = 18), $b$ = −227, $SE$ = 25, $t$(97.14) = −9.00, $p$ < .001, 95% confidence interval (CI) = [−278.30, −177.14] (see Fig. 1). Adults also showed the expected pattern; they were faster to categorize prototypical ($M$ = 947 ms, $SE$ = 7) than nonprototypical ($M$ = 1,087 ms, $SE$ = 7) animal exemplars, $b$ = −140, $SE$ = 35, $t$(62) = −3.92, $p$ < .001, 95% CI = [−211.13, −69.64].

***Gender-categorization RTs.*** On critical trials, in which children categorized human faces by gender, children were generally faster to categorize male faces than female faces, $b$ = −68, $SE$ = 13, $t$(478.29) = −5.36, $p$ < .001, 95% CI = [−92.83, −43.13], and faster to categorize Asian and White faces than Black faces, $b$ = −32, $SE$ = 8, $t$(477.61) = −4.12, $p$ < .001, 95% CI = [−47.26, −16.81]. These main effects were qualified by a significant interaction between target gender and target race, supporting our hypothesis, $b$ = 38, $SE$ = 16, $t$(477.61) = 2.47, $p$ = .014, 95% CI = [7.98, 68.88] (see left panel of Fig. 2). Given our central question of whether children's gender prototypes differ as a

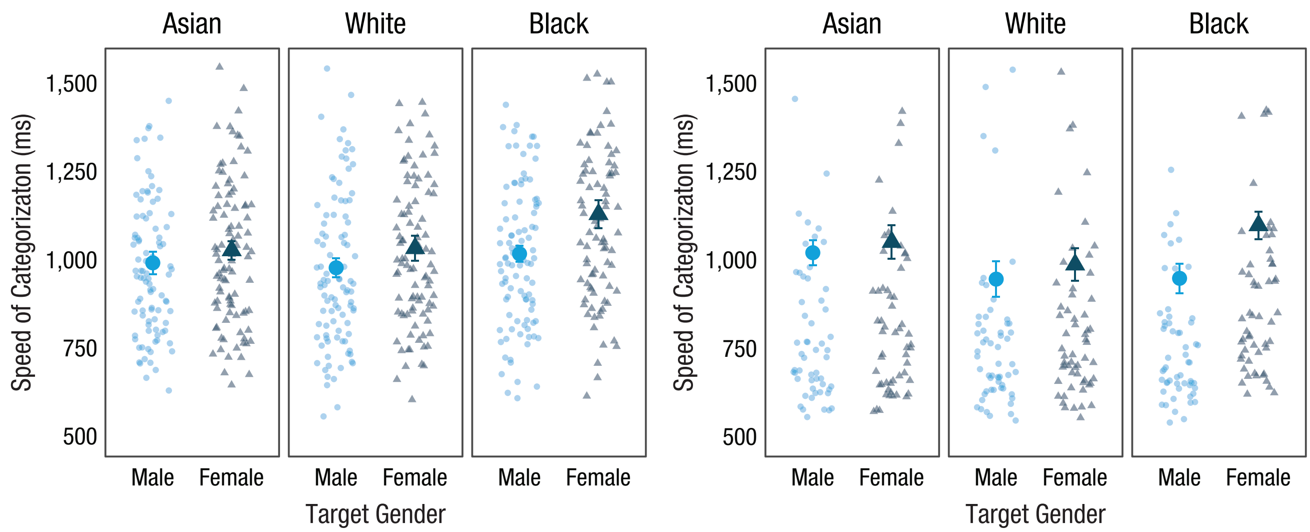


**Fig. 2.** Categorization speed for Asian, White, and Black targets as a function of target gender, separately for child (left panel) and adult (right panel) participants (Study 1). Small shapes represent individual means, and large shapes represent group means. Error bars indicate 95% confidence intervals.

function of target race, we parsed this interaction by looking at the effect of target race on target gender (for analyses parsing this interaction by focusing on the effect of target gender on target race, see the Supplemental Material). For female targets, Black women (*M* = 1,134 ms, *SE* = 24) were categorized significantly slower than either White women (*M* = 1,036 ms, *SE* = 23) or Asian women (*M* = 1,030 ms, *SE* = 23; both *p*s < .001). Results were inconsistent with our hypotheses about the psychological invisibility of Asian men; there were no differences in categorization speed as a function of the target's race for men (all *p*s > .15). Finally, there was no significant three-way interaction with age, *b* = −21, *SE* = 13, *t*(469.67) = −1.61, *p* = .108, 95% CI = [−46.22, 4.20].

The results of the adult participants broadly mirrored those of the children as well as of adults in previous research (see right panel of Fig. 2). That is, adults were generally faster to categorize male faces than female faces, *b* = −74, *SE* = 18, *t*(312) = −4.07, *p* < .001, 95% CI = [−109.23, −38.31], but the effect of target gender varied by target race, interaction: *b* = −60, *SE* = 22, *t*(312) = −2.70, *p* = .007, 95% CI = [−103.26, −16.40]. Analysis of the simple effects revealed that adult participants were slower to categorize Black women as women relative to White women (mean difference = 111 ms, *SE* = 31), *t*(124) = 3.60, *p* = .006, and relative to Black men as men (mean difference = 150 ms, *SE* = 30.90), *t*(62) = 4.86, *p* < .001, but not relative to Asian women (mean difference = 47 ms, *SE* = 31), *t*(124) = 1.53, *p* = .127. Because the distribution for adult RTs was positively skewed, we also log-transformed the RTs for analyses. Doing so yielded the same main effect of target gender, *b* = −0.03, *SE* = 0.01, *t*(312) = 4.77, *p* < .001, 95% CI = [0.02, 0.05], and interaction between target gender and target race, *b* = 0.03, *SE* = 0.01, *t*(312) = 3.17, *p* = .002, 95% CI = [0.01, 0.04].

***Gender miscategorizations.*** Results for children's gender miscategorizations revealed that children were more likely to miscategorize women than men, *b* = 1.29, *SE* = 0.12, *t*(581) = 10.67, *p* < .001, 95% CI = [1.06, 1.53], and to miscategorize the gender of Black stimuli compared with White or Asian stimuli, *b* = 0.20, *SE* = 0.07, *t*(581) = 2.78, *p* = .006, 95% CI = [0.06, 0.35]. Importantly, these main effects were qualified by a significant interaction between target race and target gender, *b* = 0.56, *SE* = 0.15, *t*(581) = 3.81, *p* < .001, 95% CI = [0.27, 0.85] (see Fig. 3). As hypothesized, children were significantly more likely to miscategorize Black women as men (*M* = 2.14, *SE* = 0.15) compared with Asian (*M* = 0.93, *SE* = 0.09) and White (*M* = 0.77, *SE* = 0.09) women (*p*s < .001). Children were also significantly more likely to miscategorize the gender of Black women than Black men (*p* < .001). Again, inconsistent with hypotheses about Asian men, there was no significant effect for male targets (*p*s > .25). Finally, there was no higher order interaction with age (*p* = .85).

***Identity-based explanations.*** We had preregistered a plan to test how a variety of intersectional processes other than those reflected in the gendered-race hypothesis evaluated above could also shape children's categorization behavior—including the possibility that children would be faster to categorize faces of their own race, faster to categorize out-group men in particular, or faster

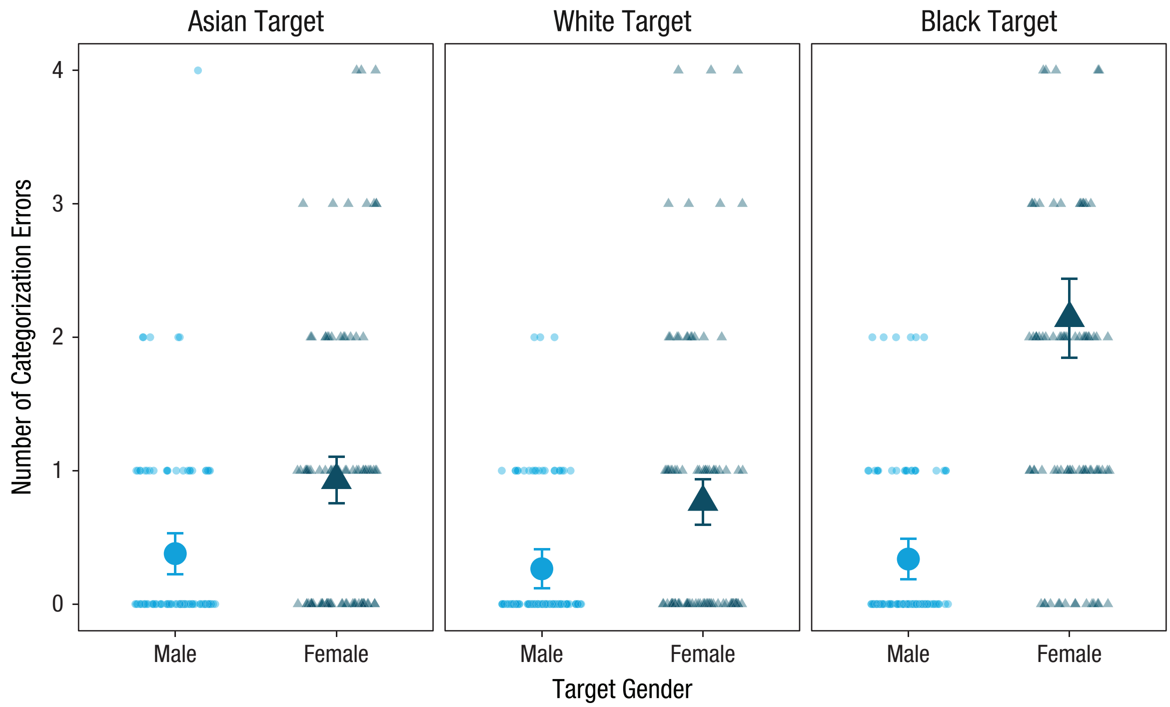


**Fig. 3.** Children's categorization errors for Asian, White, and Black stimuli as a function of target gender (Study 1). Small shapes represent individual means, and large shapes represent group means. Error bars represent 95% confidence intervals.

to categorize faces that matched their own in terms of both race and gender. Analyses did not support the possibility that any of these alternative processes shaped the speed of children's categorization decisions on this task (for model details and additional tests by participant race, see the Supplemental Material). We did find that children made fewer categorization errors for their own racial group, $b = -0.39$, $SE = 0.13$, $t(385) = -2.94$, $p = .003$, 95% CI = [−0.65, −0.13], and for male targets overall, $b = -1.50$, $SE = 0.20$, $t(385) = -7.54$, $p < .001$, 95% CI = [−1.88, −1.11]; however, there was no interaction between the two factors, $b = 0.44$, $SE = 0.28$, $t(391) = 1.57$, $p = .116$, 95% CI = [−0.11, 0.10]. Thus, identity-based processes cannot account for our finding that Black women in particular were miscategorized at higher rates than their same-race or same-gender counterparts in our central analyses.

***Differences in gender-linked facial-features explanation (not preregistered).*** To test the exploratory hypothesis that differences in gender-linked facial features between Black female faces and White or Asian female faces might account for our results, we conducted a linear mixed model predicting facial measurements as a function of target race and including random intercepts for both stimulus number and dimension measured. Results revealed no significant effect of target race, $b = -5.72$, $SE = 3.82$, $t(62) = -1.50$, $p = .139$, 95% CI = [−13.25, 1.82].

***Stereotype task.*** We also assessed whether children's perceptions of gender-stereotypic traits would be biased by the race of the target to examine whether gendered-race effects would emerge on a more deliberative, explicit task. We analyzed masculine and feminine traits separately because our hypotheses are focused on graded representations of gender categories by race. For feminine stereotypes, results revealed a significant effect of race; Black women were chosen less often than Asian women, who were in turn chosen less often than White women, $b = -0.04$, $SE = 0.01$, $t(464) = -2.89$, $p = .004$, 95% CI = [−0.07, −0.01]. This main effect of target race was qualified by an interaction between target race and valence, $b = -0.06$, $SE = 0.03$, $t(464) = -2.28$, $p = .023$, 95% CI = [−0.12, −0.01] (see Fig. 4a): Specifically, Black women ($M = .04$, $SE = .03$) were particularly unlikely to be chosen for positive feminine traits (i.e., nice, empathetic, helpful) compared with White ($M = .38$, $SE = .03$) or Asian ($M = .19$, $SE = .03$) women ($ps < .001$).

For masculine stereotypes, we found a significant effect of race; Asian men were chosen less often than

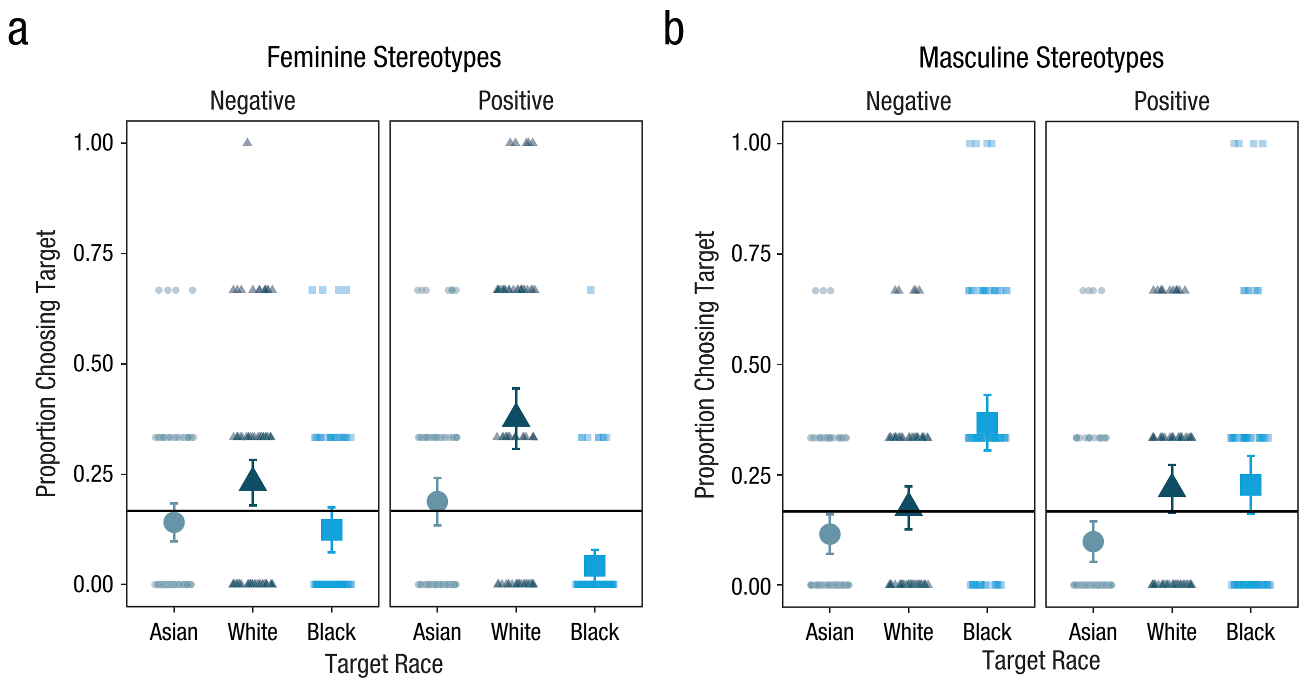


**Fig. 4.** Proportion of children who chose negative and positive traits for Black, White, and Asian women and men, separately for feminine (a) and masculine (b) stereotypes (Study 1). Small shapes represent individual means, and large shapes represent group means. The solid line represents chance levels of selecting a particular exemplar. Error bars indicate 95% confidence intervals.

White men, who were in turn chosen less often than Black men, $b = -0.10$, $SE = 0.01$, $t(464) = -7.18$, $p < .001$, 95% CI = [−0.12, −0.07]. This main effect of target race was also qualified by an interaction between target race and valence, $b = -0.06$, $SE = 0.03$, $t(464) = -2.34$, $p = .020$, 95% CI = [−0.11, −0.01] (see Fig. 4b). We broke down this interaction by trait valence and found that Asian men ($M = .10$, $SE = .03$) were less likely to be picked for positive masculine traits (leaderlike, athletic, competitive) than either White ($M = .22$, $SE = .03$) or Black ($M = .23$, $SE = .03$) men ($ps = .004$ and .002, respectively). However, a different pattern emerged for negative masculine traits (dominant, mean, risk taking). For negative masculine traits, Black men ($M = .37$, $SE = .03$) were chosen more often than White ($M = .18$, $SE = .03$) or Asian ($M = .12$, $SE = .03$) men (both $ps < .001$). Finally, we examined whether there were correlations between children's selection of Asian, White, or Black men and women for these gender stereotypes and either their speed of categorization or the propensity to miscategorize these targets, but we found no significant results (all $ps \geq .13$).

## Discussion

In Study 1, we investigated the hypothesis that children's representations of gender were biased by race. We found evidence that children's representations of women, specifically, were influenced by race; in the speeded categorization task, children were slower to categorize Black women as women (relative to White or Asian women) and more likely to miscategorize Black women than White or Asian women. Children were also less likely to pick Black women for positive feminine stereotypes. Moreover, these effects could not be explained by racial differences in gender-linked facial features alone, although it is possible that children incorporate both lower level information such as facial features along with higher level stereotype knowledge. We also acknowledge that there could be differences in the population that were not present in our stimulus set—differences that children may be familiar with.

In contrast, we did not find consistent evidence that children's representations of men were biased by race. Categorization speeds were equivalent in the speeded categorization task, and children were unlikely to miscategorize men regardless of racial identity. Only in the stereotype data did we find support for the hypothesis that children's representation of men was biased by race; Asian men were less likely (and Black men were more likely) than White men to be picked for masculine stereotypes.

Although this first study suggests that race biases children's representations of gender, we found no evidence that the child's own racial background shaped the development of these intersectional prototypes. This suggests that shared societal cues (e.g., children's media; Greenberg & Mastro, 2008) might be powerful enough to overwhelm identity-dependent pathways to social-category representation. Indirect support for this idea comes from work on social attitudes; children generally have preferences for high-status groups, regardless of their own group membership (e.g., Newheiser, Dunham, Merrill, Hoosain, & Olson, 2014). From this perspective, it is perhaps not surprising that children's intersectional social prototypes are consensual.

To test these ideas, we ran a replication study with biracial children. We examined biracial children because they have been found to show more flexibility in their social-categorization behavior than monoracial children in some tasks (Gaither, 2015) and are repeatedly exposed to people of different backgrounds in their local environment. Thus, if exposure to variability translates to a broader scope of gender representation, then biracial children should be faster to categorize women and men of all racial backgrounds relative to monoracial children. We explored these possibilities in Study 2.

## Study 2

### Method

***Participants.*** We recruited 93 children from the Children's Museum of Manhattan. Children ranged from 4 to 7.99 years old (age: $M$ = 5.66 years, $SD$ = 1.14), were roughly gender balanced (54% girls, 43% boys, 4% unreported), and were either monoracial White (49%) or biracial (51%). The racial composition of the biracial sample was as follows: White/Asian (36%), White/Black (21%), White/Hispanic (13%), Black/Hispanic (11%), broadly biracial (11%), and other (8%). We recognize that biracial children from different backgrounds have highly variable experiences, and thus, it is somewhat limiting to consider them as a single group. We did so here, as has commonly been done in previous work (e.g., Gaither, Sommers, & Ambady, 2013; Roberts & Gelman, 2017), because the cognitive flexibility that is of primary interest has been demonstrated to be something unique to biracial people as a broad group (i.e., being able to switch between identities) and not specific to the particular racial-group combination (e.g., Gaither, 2015). Of the 93 participants, we had Datavyu consent and videos for 84 participants, which comprised our final usable sample (mean age = 5.56 years; 54% girls; 52% biracial).

***Materials and procedure.*** Materials and procedures were exactly the same as in Study 1.

***Analytic strategy.*** Our analysis plan was the same as the revised analysis plan from Study 1 (i.e., incorporating all deviations from our preregistration). However, to examine whether children's identity moderated any effects, we compared monoracial (White) children (coded as −0.5) with biracial children (coded as 0.5).

### Results

***Validation-trial RTs.*** Results replicated those of Study 1: Children were faster to categorize prototypical animal stimuli (e.g., to categorize a robin as a bird; $M$ = 1,181 ms, $SE$ = 36) compared with nonprototypical animal stimuli (e.g., to categorize a chicken as a bird; $M$ = 1,398 ms, $SD$ = 36), $b$ = −217, $SE$ = 37, $t$(79.50) = −5.91, $p$ < .001, 95% CI = [−289.52, −144.78].

***Gender-categorization RTs.*** On the speeded categorization trials, children again generally categorized male faces faster than female faces, $b$ = −99, $SE$ = 14, $t$(414.72) = 6.90, $p$ < .001, 95% CI = [−126.51, −70.58], and Asian and White faces faster than Black faces, $b$ = −31, $SE$ = 9, $t$(414.77) = −3.59, $p$ < .001, 95% CI = [−48.60, −14.32]. As in Study 1, these main effects were qualified by a significant interaction between target gender and target race, $b$ = 39, $SE$ = 18, $t$(414.77) = 2.24, $p$ = .025, 95% CI = [5.02, 73.59] (see Fig. 5). We again parsed this interaction by looking at the effect of target race on target gender. For female targets, Black women ($M$ = 1,301, $SE$ = 26) were categorized more slowly than either White women ($M$ = 1,176, $SE$ = 26) or Asian women ($M$ = 1,198, $SE$ = 26; both $p$s < .001), with no difference in categorization speeds for the latter two targets ($p$ > .25). For male targets, there were no differences in the categorization speed as a function of the target's race (all $p$s ≥ .23).

***Exploratory moderation analyses for categorization speed.*** Finally, we explored two potential moderators of this interaction between target race and target gender: (a) whether the child was monoracial (White) or biracial and (b) the child's age (as in Study 1). We first turned to whether the two-way interaction between target gender and target race was moderated by the child's biracial status; results revealed no significant three-way interaction, $b$ = −20, $SE$ = 35, $t$(411.69) = −0.56, $p$ = .58, 95% CI = [−88.17, 48.89]. We also explored whether there might be a four-way interaction among all specified factors, but results indicated that there was not, $b$ = 17, $SE$ = 37, $t$(405.80) = 0.46, $p$ = .65, 95% CI = [−54.22, 87.84].

We turned next to the effects of the child's age; results revealed a significant three-way interaction among target race, target gender, and the child's age, $b$ = −39, $SE$ = 17, $t$(411.72) = −2.29, $p$ = .022, 95% CI = [−72.67, −5.87] (see Fig. 6). When we unpacked this three-way interaction by target gender, results revealed a significant two-way interaction between target race and the child's age for female targets, $b$ = −33, $SE$ = 13,


**Fig. 5.** Children's categorization speed for Asian, White, and Black stimuli as a function of target gender (Study 2). Small shapes represent individual means, and large shapes represent group means. Error bars indicate 95% confidence intervals.

$t$(160.90) = −2.62, $p$ = .010, 95% CI = [−58.50, −8.36]; as shown in Figure 6, children categorized targets more quickly with age, but this age-related improvement was attenuated for Black women relative to White and Asian women. For male targets, there was no such two-way interaction, $b$ = −5, $SE$ = 11, $t$(166) = −0.51, $p$ = .61, 95% CI = [−26.43, 15.59].

***Gender miscategorizations.*** When we examined children's gender miscategorizations, results replicated the interaction between target gender and target race observed in Study 1, $b$ = −0.80, $SE$ = 0.15, $t$(504) = −5.29, $p$ < .001, 95% CI = [−1.10, −0.51]; children were much more likely to miscategorize Black women as men ($M$ = 2.41, $SE$ = 0.18) than to miscategorize Asian ($M$ = 0.75, $SE$ = 0.09) and White ($M$ = 0.78, $SE$ = 0.10) women as men ($p$s < .001). Children were also significantly more likely to miscategorize the gender of Black women than Black men ($p$ < .001).

***Exploratory moderation analyses for gender miscategorization.*** As with the categorization-speed data, we again tested two candidate moderators: the child's biracial status and age. When we looked at potential effects of the child's biracial status, results revealed no significant three-way interaction, $b$ = 0.12, $SE$ = 0.30, $t$(500) = 0.41, $p$ = .69, 95% CI = [−0.47, 0.72]. We turned next to potential effects of the child's age; results revealed no higher order interaction with age, $b$ = −0.08, $SE$ = 0.18, $t$(500) = −0.47, $p$ = .64, 95% CI = [−0.43, 0.26], nor any four-way interaction between all specified factors, $b$ = 0.34, $SE$ = 0.37, $t$(492) = 0.92, $p$ = .36, 95% CI = [−0.38, 1.05].

***Stereotype task.*** As in Study 1, we analyzed masculine and feminine traits separately. For feminine stereotypes, results revealed only a significant effect of race; White women ($M$ = .26, $SE$ = .02) were chosen more often than either Asian ($M$ = .16, $SE$ = .02) or Black ($M$ = .15, $SE$ = .02) women, $b$ = 0.10, $SE$ = 0.03, $t$(507) = 3.69, $p$ < .001, 95% CI = [0.05, 0.15]. Unlike in Study 1, there was no interaction with trait valence, $b$ = 0.03, $SE$ = 0.03, $t$(506) = 1.16, $p$ = .246, 95% CI = [−0.02, 0.08]. For masculine stereotypes, there was also a significant effect of race; Asian men ($M$ = .11, $SE$ = .02) were chosen less often than either White ($M$ = .20, $SE$ = .02) or Black ($M$ = .21, $SE$ = .02) men, $b$ = −0.05, $SE$ = 0.01, $t$(506) = 3.80, $p$ < .001, 95% CI = [−0.07, −0.02]. Again, unlike in Study 1, there was no interaction with trait valence, $b$ = 0.02, $SE$ = 0.03, $t$(506) = 0.63, $p$ = .53, 95% CI = [−0.03, 0.06]. Finally, we also examined whether there were any higher order interactions with age or biracial status, but none emerged, $b$ = 0.01, $SE$ = 0.02, $t$(502) = 0.58, $p$ = .56, 95% CI = [−0.03, 0.06], and $b$ = 0.002, $SE$ = 0.05, $t$(502) = 0.04, $p$ = .97, 95% CI = [−.10, .10], respectively.

## Omnibus Analyses

To provide the highest powered test of our developmental hypotheses (i.e., change across age, which was

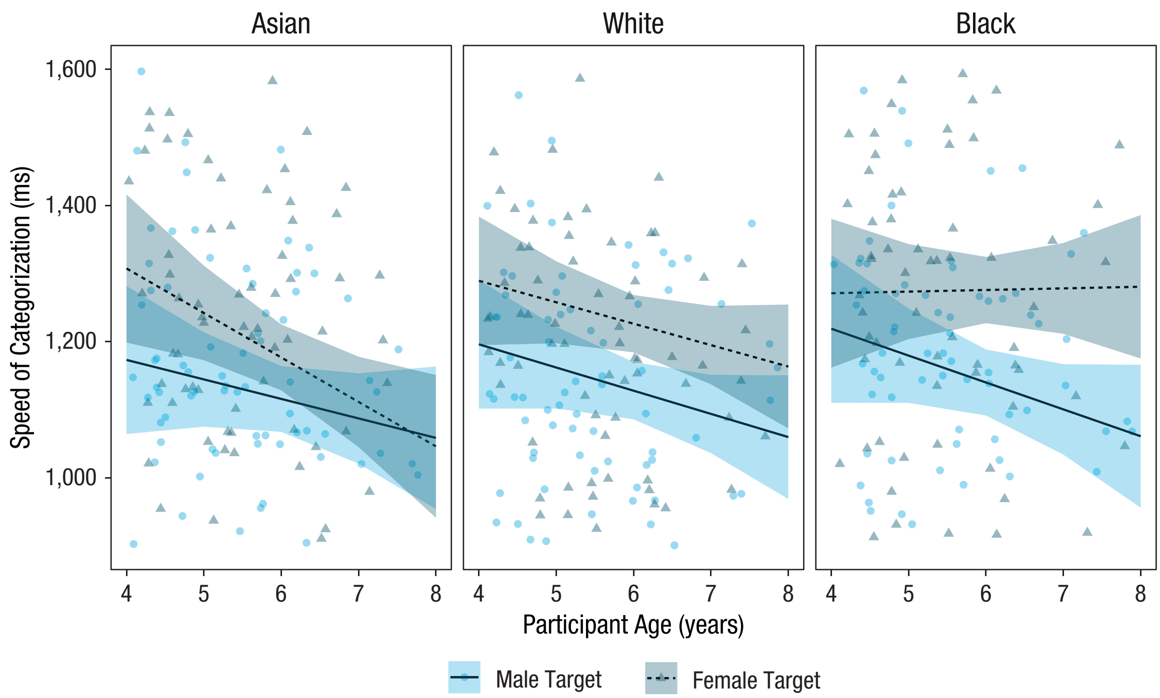


**Fig. 6.** Children's categorization speed for Asian, White, and Black targets as a function of participant age and target gender (Study 2). Circles and triangles represent individual means for male and female targets, respectively. Slopes are best-fitting regressions, and shaded regions are 95% confidence bands.

found in Study 2 but not Study 1) and a stronger test of the role of valence on the social stereotypes measured in our explicit task, we combined the data from children from Study 1 and Study 2 ($N$ = 186). For all analyses, we checked to see whether there were any higher order interactions by study, but none were significant ($p$s > .10). Thus, all subsequent analyses were collapsed across studies.

## Gender-categorization RTs

Results replicated those of Studies 1 and 2: Children were faster to categorize male faces than female faces, $b = -82$, $SE = 9$, $t(896.01) = -8.67$, $p < .001$, 95% CI = [−100.90, −63.69], and to categorize Asian and White faces faster than Black faces, $b = -32$, $SE = 6$, $t(895.40) = -5.47$, $p < .001$, 95% CI = [−43.23, −20.43]. We also replicated the interaction between target gender and target race, $b = 39$, $SE = 12$, $t(895.40) = 3.35$, $p < .001$, 95% CI = [16.15, 61.75]. Again, children were slower to categorize Black women as women ($M$ = 1,211 ms, $SE$ = 18) than to categorize Asian women ($M$ = 1,108 ms, $SE$ = 18) and White women as women ($M$ = 1,101 ms, $SE$ = 18; both $p$s < .001).

We next examined whether the two-way interaction between target race and target gender was further moderated by the child's age. Results revealed a significant three-way interaction, $b = -27$, $SE = 10$, $t(887.25) = -2.67$, $p = .008$, 95% CI = [−47.66, −7.33] (see Fig. 7). As in Studies 1 and 2, we first examined this three-way interaction as a function of target gender (for analyses as a function of target race, see the Supplemental Material). For female targets, there was a significant two-way interaction between target race and the age of the participant, $b = -18$, $SE = 8$, $t(352.40) = -2.21$, $p = .028$, 95% CI = [−33.04, −1.95]. Although older children generally were faster at categorizing faces than younger children, this improvement in categorization speed was attenuated for Black women. For male targets, there was no significant two-way interaction between target race and age, $b = 10$, $SE = 6$, $t(358.94) = 1.59$, $p = .11$, 95% CI = [−2.29, 22.21].

## Gender miscategorizations

Results replicated those of Studies 1 and 2: Children were more likely to miscategorize the gender of Black women specifically ($M$ = 2.26, $SE$ = 0.12), relative to both Asian women ($M$ = 0.85, $SE$ = 0.06), White women ($M$ = 0.78, $SE$ = 0.06), and Black men ($M$ = 0.31, $SE$ = 0.06), $b = 0.67$, $SE = 0.10$, $t(1091) = 6.51$, $p < .001$, 95% CI = [0.47, 0.87]. We also explored whether this propensity to miscategorize Black women might change with age but found no three-way interaction among

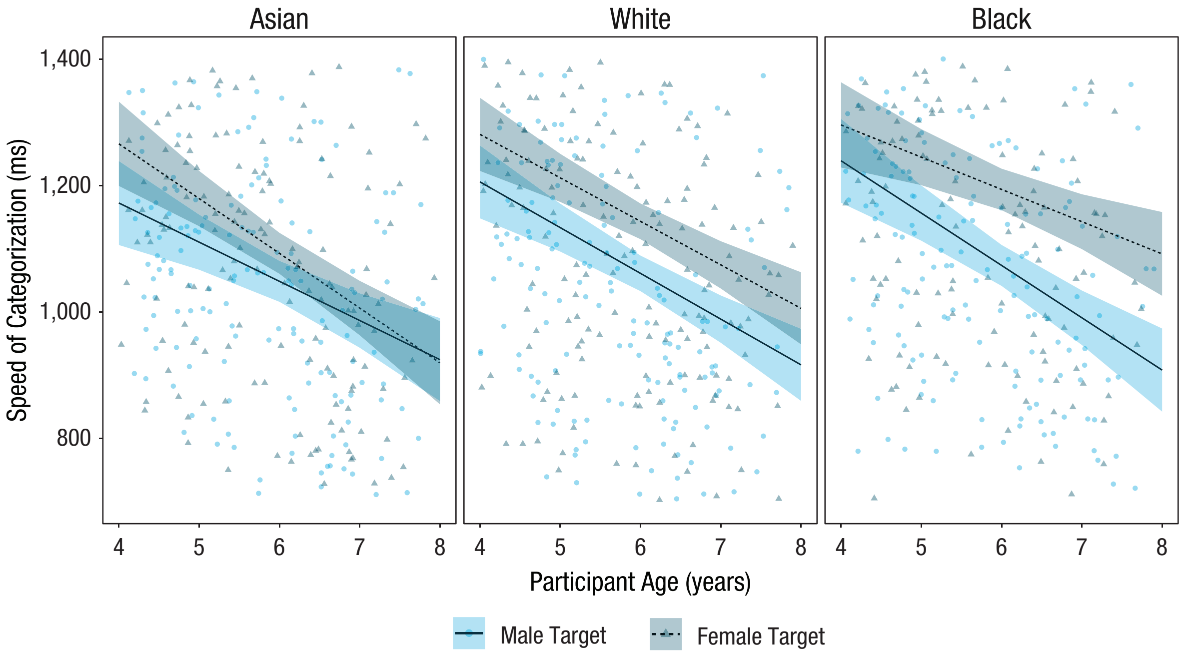


**Fig. 7.** Children's categorization speed for Asian, White, and Black targets as a function of participant age and target gender (omnibus analysis). Circles and triangles represent individual means for male and female targets, respectively. Slopes are best-fitting regressions, and shaded regions are 95% confidence bands.

target race, target gender, and the child's age, $b = -0.03$, $SE = 0.10$, $t(1081) = -0.32$, $p = .75$, 95% CI = [−0.22, 0.17].

## Environmental and parental correlates

Because the combined data set represented the largest number of responses for parent beliefs and environmental correlates (depending on the question, $n$s = 135–162), we opted to examine these variables here. We focused on children's social networks and neighborhood diversity in particular because previous work has demonstrated that these factors influence children's beliefs about race and gender (e.g., Mandalaywala et al., 2019). None of these variables moderated children's categorization speeds and error rates; however, there were some significant correlations. We summarize these patterns below (for all pairwise correlations, see the Supplemental Material).

In general, children's RTs to all stimuli were moderately to strongly correlated ($r$s = .45–.70). More notably, children's friendship networks were weakly correlated with their RTs (see Fig. 8). Specifically, children with a greater percentage of Black friends tended to be faster at categorizing people of all kinds ($r$s = −.10 to −.27), whereas children with a greater percentage of White friends tended to be slower at categorizing people of all kinds ($r$s = .16 to .27). Interestingly, this difference seems to be about friendships with White and Black children specifically, as the percentage of Asian friends in a child's social network did not significantly correlate with children's categorization speeds ($r$s = −.11 to .01). We also observed significant correlations between the diversity of children's neighborhood context and their categorization speeds. Specifically, children who live in neighborhood contexts with more White than Black people tended to be slower at categorizing targets ($r$s = .16 to .20). Finally, we also examined whether any of these environmental factors correlated with children's propensity to miscategorize targets but found no significant pairwise correlations.

## Stereotype task

We again analyzed masculine and feminine traits separately. For feminine stereotypes, results replicated those of Study 1 and showed a significant interaction between target race and trait valence, $b = 0.05$, $SE = 0.02$, $t(974) = 2.41$, $p = .016$, 95% CI = [0.01, 0.09]. For positive traits, Black women ($M = .09$, $SE = .02$) were chosen less often than Asian women ($M = .18$, $SE = .02$), who in turn were chosen less often than White women ($M = .33$, $SE = .02$; $p$s ≤ .003). For negative traits, both Black women ($M = .15$, $SE = .02$) and Asian women ($M = .14$, $SE = .02$) were chosen less often than White women ($M = .23$, $SE = .02$; both $p$s ≤ .006).

For masculine stereotypes, results again replicated those of Study 1 and also showed a significant interaction

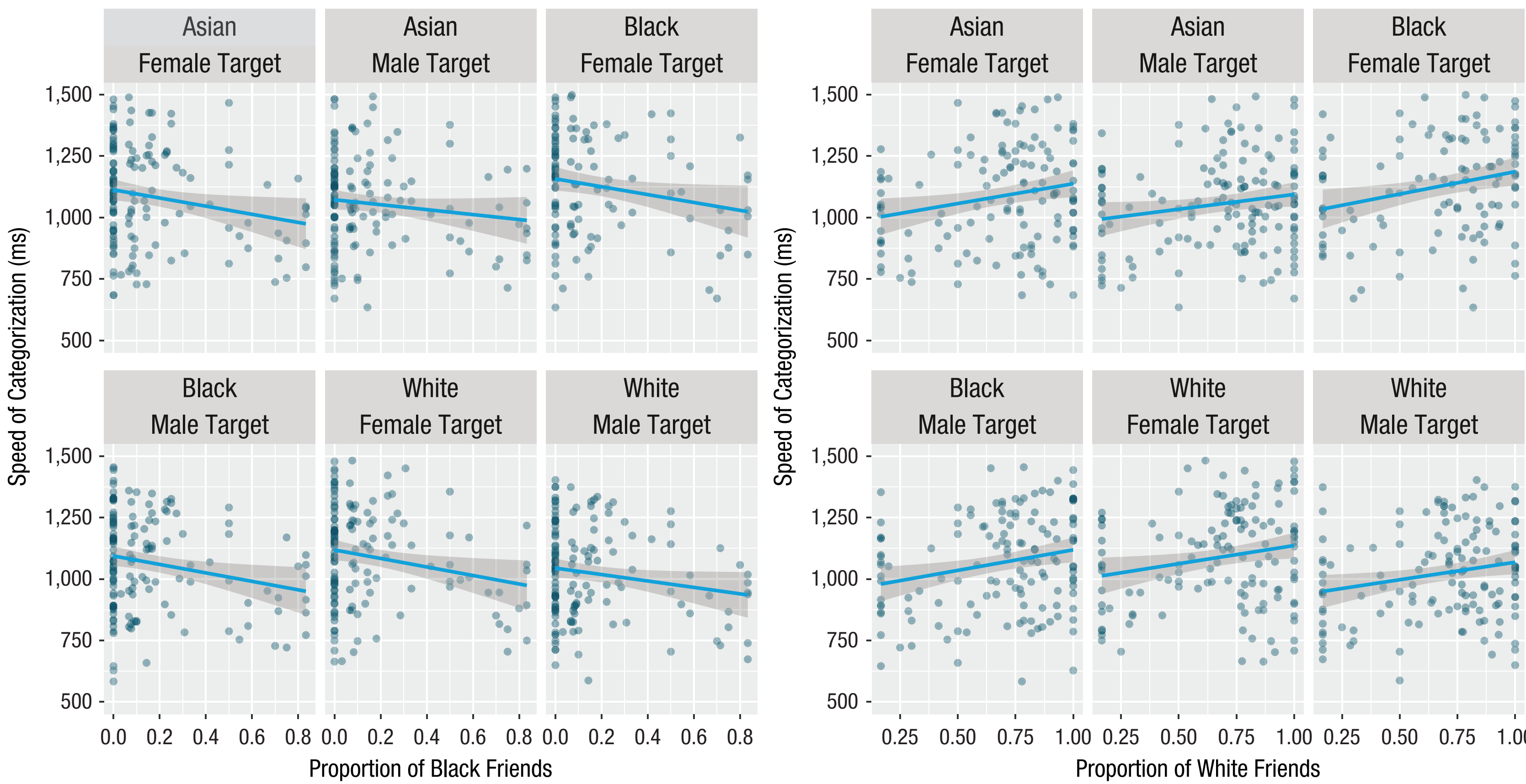


**Fig. 8.** Correlations between categorization speed and the proportion of children's friends who are Black (left panel) and White (right panel; omnibus analysis). Results are shown separately for each combination of target race and gender. Lines indicate best-fitting regressions, and shaded regions are 95% confidence bands.

between target race and valence, $b$ = 0.04, $SE$ = 0.02, $t$(974) = 2.08, $p$ = .038, 95% CI = [0.002, 0.07]. For positive traits, Asian men ($M$ = .10, $SE$ = .02) were chosen less often than either White ($M$ = .22, $SE$ = .02) or Black ($M$ = .20, $SE$ = .02) men ($p$s < .001). For negative traits, Black men ($M$ = .30, $SE$ = .02) were chosen more often than Asian men ($M$ = .12, $SE$ = .02) or White men ($M$ = .17, $SE$ = .02; both $p$s < .001).

## General Discussion

Across a variety of measures in two separate studies, children's representation of gender shared some structural similarities to that of adults (e.g., Johnson et al., 2012)—a conceptual structure reflecting gendered-race representations that may underlie the psychological invisibility of members of some social groups. Specifically, children were slower to categorize Black women as women, more likely to miscategorize them as men, and less likely to pair them with feminine traits—effects that are strikingly similar to previous work with adults (e.g., Goff et al., 2008). Notably, this psychological invisibility developed over age. The youngest children were equally fast at categorizing male and female faces regardless of race. Around age 5 years, however, children began showing gendered-race effects on the speeded categorization task (although children's tendency to miscategorize Black women did not vary across age).

Gender and race did not systematically interact to shape representations of men for children (cf. Johnson et al., 2012; Schug et al., 2015), with one exception. On the stereotyping task, children were less likely to choose Asian men for masculine traits and were particularly likely to choose Black men for negative masculine traits. One possible explanation for this divergence between men and women's faces is that they may invoke different processes. Children's general faster processing of male faces could reflect an *androcentric bias* (i.e., a tendency to view men as the prototypical *person*; Bailey, LaFrance, & Dovidio, 2019). Thus, children may more readily offer a male categorization (regardless of race) and have to correct this bias to produce a female categorization—a correction that may take longer for Black women. Another possibility is that the Asian–femininity association emerges later in development than the Black–masculinity association because it relies primarily on top-down stereotypic knowledge, whereas the Black–masculinity association may employ both bottom-up (e.g., facial cues) and top-down processes (Johnson et al., 2012). Future work should explore these possibilities.

Our results do not support the identity-dependent hypotheses (e.g., Anzures et al., 2013; Nesdale, 2004). Not all children were quicker to classify faces of their own race, nor did biracial children appear to have an advantage in categorization speed and accuracy, despite their greater flexibility on some tasks (Gaither, 2015) and broader range of racial exposure. These findings leave open the question of which specific forces shape children's representation of social categories. One possibility arises from the correlations with children's social networks. Perhaps exposure is not sufficient; children may need practice in classifying different exemplars into categories (as may happen implicitly when playing with peers). This classification process may be particularly influential with Black (vs. Asian) friends because Black faces provide greater variability from category prototypes in both skin tone and facial physiognomy (e.g., Dunham, Stepanova, Dotsch, & Todorov, 2015). Such classification processes also implicitly rely on bottom-up processes, which may play a greater role in daily life than observed in the current work. It may be that the more practice children have in incorporating this variability into their representations, the faster they are in general at categorizing people. Our measure of children's social networks was based on parent report, however; future work should further focus on how these features of children's daily lives influence their conceptual and social development with more objective indicators of children's experiences.

The present work adds to our theoretical understanding of how children develop social prototypes. Specifically, our findings suggest that children develop complex representations of gender that intersect with race; across early childhood, children begin to view Black women as less prototypical than Asian and White women across a range of dimensions (with more limited effects for their representations of Asian men). One implication of these results is that much of what we know about children's use of gender may reflect how children think about White women and men specifically. Such methodological choices could themselves reflect a manifestation of psychological invisibility (Fryberg & Townsend, 2008). Ultimately, these findings highlight a basic feature of conceptual representation that emerges in early childhood, which both reflects and may perpetuate the invisibility experienced by Black women and Asian men in interpersonal, social, and political contexts.

### Transparency

*Action Editor:* Erika E. Forbes
*Editor:* D. Stephen Lindsay
*Author Contributions*

R. F. Lei and M. Rhodes designed the research, and R. F. Lei and R. A. Leshin conducted the research and analyzed the data. All the authors wrote the manuscript and approved the final manuscript for submission.

*Declaration of Conflicting Interests*

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

*Funding*

This research was supported by a Scholars Award from the James S. McDonnell Foundation to M. Rhodes and by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award No. R01HD087672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

*Open Practices*

All data, materials, and analysis scripts have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/qnye7. The design and analysis plans were preregistered at http://aspredicted.org/blind.php?x=zw38pd. Deviations from the preregistration are noted in the text. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620920360. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

## ORCID iD

Ryan F. Lei https://orcid.org/0000-0002-7954-2368

## Acknowledgments

We thank Magnolia Byrne and Jerolynn Liu for assistance with data collection and the Children's Museum of Manhattan for partnering in this research.

## Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620920360

## Notes

1. The data collected for a pilot sample (noted in our preregistration) were not included in the reported sample but, rather, served to ensure that the task was understandable to young children. The distinction between the pilot and main collection phases was decided in advance.
2. We asked parents to fill out this survey in both Study 1 and Study 2 and opted to examine potential correlates only in our omnibus analysis (depending on the question, *N*s = 135–162) to obtain the highest powered estimate of effects.

## References

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the other-race effect. *Current Directions in Psychological Science*, *22*, 173–178.

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2019). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review*, *23*, 307–331.

Bauer, D., & Cox, W. M. (1998). Alcohol-related words are distracting to both alcohol abusers and non-abusers in the Stroop colour-naming task. *Addiction*, *93*, 1539–1542. doi:10.1046/j.1360-0443.1998.9310153910.x

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum, 1989*, Article 8. Retrieved from https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8/

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, *21*, 1139–1150. doi:10.1177/01461672952111002

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science*, *18*, 469–483.

Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? Development changes in representation of biological categories. *Cognitive Psychology*, *110*, 1–15.

Fryberg, S. A., & Townsend, S. S. M. (2008). The psychology of invisibility. In G. Adams, M. Biernat, N. R. Branscombe, C. S. Crandall, & L. S. Wrightsman (Eds.), *Decade of behavior: Commemorating Brown: The social psychology of racism and discrimination* (pp. 173–193). Washington, DC: American Psychological Association.

Gaither, S. E. (2015). "Mixed" results: Multiracial research and identity explorations. *Current Directions in Psychological Science*, *24*, 114–119.

Gaither, S. E., Sommers, S. R., & Ambady, N. (2013). When the half affects the whole: Priming identity for biracial individuals in social interactions. *Journal of Experimental Social Psychology*, *49*, 368–371.

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, *24*, 498–506.

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). "Ain't I a woman?": Towards an intersectional approach to person perception and group-based harms. *Sex Roles*, *59*, 392–403. doi:10.1007/s11199-008-9505-4

Greenberg, B. S., & Mastro, D. E. (2008). Children, race, ethnicity, and media. In S. L. Calvert & B. J. Wilson (Eds.), *The handbook of children, media, and development* (pp. 74–97). Malden, MA: Blackwell.

Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem Sex-Role Inventory. *Sex Role*, *39*, 929–941. doi:10.1023/A:1018836923919

Hull, G. T., Bell Scott, P., & Smith, B. (Eds.). (1982). *But some of us are brave: All the women are White, all the Blacks are men: Black women's studies*. Old Westbury, NY: Feminist Press.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*, 116–131. doi:10.1037/a0025335

Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology*, *40*, 581–592.

La Force, T. (2018, November 6). Why do Asian-Americans remain largely unseen in film and television? *The New York Times*. Retrieved from https://www.nytimes.com/2018/11/06/t-magazine/asian-american-actors-representation.html

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, *67*(2), 1–147.

Lingeman, J., Freeman, C., & Adolph, K. E. (2014). Datavyu (Version 1.2.2) [Computer software]. Retrieved from http://datavyu.org

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

Mandalaywala, T. M., Ranger-Murdock, G., Amodio, D. M., & Rhodes, M. (2019). The nature and consequences of essentialist beliefs about race in early childhood. *Child Development*, *90*, e437–e453. doi:10.1111/cdev.13008

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*, 903–933. doi:10.1037/0033-2909.128.6.903

Meltzoff, A. N. (2013). Origins of social cognition: Bidirectional self-other mapping and the “like-me” hypothesis. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 139–144). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199890712.003.0025

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115. doi:10.1146/annurev.ps.32.020181.000513

Millisecond. (2016). Inquisit 5 [Computer software]. Retrieved from http://www.millisecond.com

Nesdale, D. (2004). Social identity processes and children’s ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 219–245). New York, NY: Psychology Press.

Newheiser, A. K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, *50*, 1081–1090.

Perszyk, D. R., Lei, R. F., Bodenhausen, G. V., Richeson, J. A., & Waxman, S. R. (2019). Bias at the intersection of race and gender: Evidence from preschool-aged children. *Developmental Science*, *22*(3), Article e12788. doi:10.1111/desc.12788

Prentice, D. A., & Carranza, E. (2002). What women and men shouldn’t be, should be, are allowed to be, and don’t have to be. *Psychology of Women Quarterly*, *26*, 269–281. doi:10.1111/1471-6402.t01-1-00066

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*, 377–391.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*, 244–274. doi:10.1016/j.cogpsych.2009.05.001

Roberts, S. O., & Gelman, S. A. (2017). Multiracial children’s and adults’ categorizations of multiracial individuals. *Journal of Cognition and Development*, *18*(1), 1–15.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. doi:10.1016/0010-0285(75)90024-9

Schug, J., Alt, N. P., & Klauer, K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology*, *56*, 121–125. doi:10.1016/j.jesp.2014.09.012

Schug, J., Alt, N. P., Lu, P. S., Gosin, M., & Fay, J. L. (2017). Gendered race in mass media: Invisibility of Asian men and Black women in popular magazines. *Psychology of Popular Media Culture*, *6*, 222–236.

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*, 356–360.

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children’s use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *1*, 35–62. doi:10.1080/15248372.2011.638686

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly*, *24*, 44–53. doi:10.1111/j.1471-6402.2000.tb01021.x

Spence, J. T., Heleich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, *24*, 219–220. doi:10.3758/BF03329252

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, *48*, 1197–1204. doi:10.3758/s13428-015-0641-9

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