Empirical Research Paper

Intergroup Context Moderates the Impact of White Americans' Identification on Racial Categorization of Ambiguous Faces

Personality and Social Psychology Bulletin
1–13

DOI: 10.1177/01461672231190264
journals.sagepub.com/home/pspb
Sage

Jacqueline M. Chen[1], Chanel Meyers[2], Kristin Pauker[3], Sarah E. Gaither[4], David L. Hamilton[5], and Jeffrey W. Sherman[6]

Abstract

We examined how the number of groups in a categorization task influences how White Americans categorize ambiguous faces. We investigated the strength of *identity-driven ingroup overexclusion*–wherein highly identified perceivers overexclude ambiguous members from the ingroup–proposing that, compared with dichotomous tasks (with only the ingroup and one outgroup), tasks with more outgroups attenuate identity-driven ingroup overexclusion (a dilution effect). Fourteen studies ($n = 4{,}001$) measured White Americans' racial identification and their categorizations of ambiguous faces and manipulated the categorization task to have two groups, three groups, or an unspecified number of groups (open-ended). In all three conditions, participants overexcluded faces from the White category on average. There was limited support for the dilution effect: identity-driven ingroup overexclusion was absent in the three-group task and only weakly supported in the open-ended task. The presence of multiple outgroups may dampen the impact of racial identity on race perceptions among White Americans.

Keywords

categorization, ingroup overexclusion, racial identity, racial ambiguity

Received September 29, 2022; revision accepted July 07, 2023

The social environment is complex and noisy, with multiple cues and identities guiding perceivers' judgments of others. Perceivers are equipped with various social perception mechanisms that help them rapidly make sense of the world. Social categorization is one such process (Allport, 1954; Tajfel, 1969), providing a framework that guides perceivers' impressions of and attitudes toward others (Brewer, 1988; Fiske & Neuberg, 1990). When individuals judge others to be ingroup members, they are more likely to judge them favorably and share resources with them. In contrast, when individuals judge others to be outgroup members, they are more likely to discriminate against them. Reliance on social categorization is ubiquitous and potent, and therefore an important question is how the mind compensates when a target's social group is ambiguous. We specifically consider the case of racially ambiguous faces that have a mixture of Black and White facial features, who are the most frequently studied type of racially ambiguous face to date (see Chen et al., 2021).

Racial categorization is a complex process by which perceivers place an exemplar into a racial group. There are different models of how perceivers make categorization decisions. Classic rule-based models argue that perceivers test a series of explicit if-then hypotheses to determine whether a target fits a category (e.g., Bruner et al., 1956; Nosofsky et al., 1994). A prototype-based model of categorization proposes that perceivers compare an exemplar (e.g., a single face) to different category prototypes to assess its degree of fit with each group (Posner & Keele, 1968; Rosch & Mervis, 1975), deciding on the category that fits the exemplar the best. Exemplar-based models propose that perceivers store individual exemplars and compare new exemplars

[1]The University of Utah, Salt Lake City, USA
[2]University of Oregon, Eugene, USA
[3]University of Hawai'i at Manoa, USA
[4]Duke University, Durham, NC, USA
[5]University of California, Santa Barbara, USA
[6]University of California, Davis, USA

Corresponding Author:
Jacqueline M. Chen, Department of Psychology, The University of Utah, 380 S 1530 E BEH S 502, Salt Lake City, UT 84112, USA.
Email: jacqueline.chen@psych.utah.edu

to those stored in memory (e.g., Nosofsky et al., 2011). Enough similarity precipitates a categorization of the target into the same category as other, related exemplars. Each approach to categorization has its limitations, and modern views of categorization tend to hybridize the approaches to account for diverse instances of category learning (see Love, 2017, for a review).

A central characteristic of modern views of categorization is the role of top-down effects—the influence of knowledge, expertise, and motivation—that are theorized to play a stronger role in the case of categorizing ambiguous targets compared with unambiguous ones (e.g., Freeman & Ambady, 2011). These models explicate the social construction of race; namely, that perceptions of others' race are imbued with meaning based on the perceiver's previous knowledge, experience, and motivations, which vary between individuals and cultures (e.g., Chen, de Paula Couto, et al., 2018). Nonetheless, it is not yet well understood when *particular* top-down processes will influence racial categorization, and when they will not. The present research investigates whether the impact of perceivers' racial identification on racial categorization depends on the intergroup context, in particular on the number of outgroups present. Specifically, we investigate this question by examining how White Americans categorize Black–White racially ambiguous faces by race, and how these processes change as a function of the categorization task.[1]

Extant research suggests that social categorization, and other intergroup processes, differ depending on how many outgroups are salient. Consequently, previous findings from contexts with only one ingroup and one outgroup may not generalize when one or more outgroups are added. Thus, because societies today are increasingly diverse, and perceivers are routinely in ingroups whose position is relative to multiple outgroups, it is imperative that we determine how this variable impacts intergroup processes. To do so, we examine the ecological validity of close-ended categorization tasks by comparing them with results from an open-ended categorization task, which is the most face-valid way to examine everyday social categorization processes.

Salience of Groups Impacts Intergroup Processes

Classic theories of intergroup psychology argue that group memberships foster social identities that play a major role in attitudes and behaviors (Tajfel & Turner, 1986; Turner et al., 1987). In these theoretical traditions, the salience of group identity is a major determinant of downstream intergroup processes; specifically, a salient group identity can lead to depersonalization, stereotyping, and outgroup bias (see Turner & Reynolds, 2011 for a review).

One way to increase the salience of an ingroup identity is to divide people into two groups—"us" and "them"—which facilitates social comparison, competition, and conflict (e.g., Brewer, 2001; Esses et al., 1998; Hartstone & Augoustinos, 1995; Sherif, 1966; Spielman, 2000). In fact, the minimal group paradigm (Tajfel et al., 1971), used to study intergroup processes for decades, typically divides participants into "us" and "them." It is relatively unknown how intergroup settings with more than two groups impact the salience of the ingroup and resulting intergroup processes.

We know of only two studies to date that have expanded the minimal group paradigm beyond the two-group context (Hartstone & Augoustinos, 1995; Spielman, 2000). In both papers, the researchers modified the minimal group paradigm to include two groups or three groups, and then they compared levels of intergroup bias (measured as preferential resource allocation to the ingroup over the outgroup[s]) between the conditions. Both studies found that participants in the three-group context exhibited little to no ingroup bias (and significantly less bias than participants in the two-group context). These findings suggest that intergroup processes may differ depending on how many groups there are.

Consistent with these findings, we propose that adding multiple outgroups to an intergroup context can reduce the impact of the perceiver's ingroup identity on social categorization. In other words, relative to the dichotomous intergroup context, adding more outgroups can reduce the salience of the ingroup identity and the impact of ingroup-based motivations.

Reduced Salience of the Ingroup May Impact Social Categorization

Our research investigates the effect of the number of groups on identity-driven racial categorization. From the social categorization literature, the process of *ingroup overexclusion* describes the phenomenon that perceivers are more likely to categorize ambiguous members as outgroup members than ingroup members (Castano et al., 2002; Leyens & Yzerbyt, 1992; Yzerbyt et al., 1995). While ingroup overexclusion describes a categorization pattern, it is also defined as an identity-driven process stemming from the desire to protect the ingroup from potentially undesirable outgroup members (Leyens & Yzerbyt, 1992; Yzerbyt et al., 1995). Evidence of *identity-driven ingroup overexclusion* comes from studies documenting that perceivers who are highly identified with their ingroup are more likely to engage in overexclusion (Castano et al., 2002; Knowles & Peng, 2005). Parallel to past studies on other intergroup processes (e.g., resource allocation, prejudice), studies documenting identity-driven ingroup overexclusion rely on paradigms that present ambiguous targets and asks perceivers to categorize them in a two-group task. To our knowledge, there is no research that documents identity-driven ingroup overexclusion when perceivers have more than one outgroup category available to them. Thus, we test whether the number of groups salient to the perceiver changes the influence of perceiver identification on social categorization.

Research has documented various perceiver-based motivations that can shift how racially ambiguous faces are categorized (see Pauker et al., 2018). One line of work has shown that White Americans tend to categorize racially ambiguous faces as Black more often than as White (e.g., Ho et al., 2011, 2013; Krosch & Amodio, 2014; Krosch et al., 2013; Peery & Bodenhausen, 2008), consistent with ingroup overexclusion. However, a recent meta-analysis of the published literature found that empirical support for ingroup overexclusion relies on a particular categorization task used (Young et al., 2021). Specifically, studies that document White Americans engage in ingroup overexclusion when categorizing racially ambiguous faces rely *exclusively* on two-group categorization tasks, in which the participant must decide whether the ambiguous target is Black or White (see also Gaither et al., 2019; Knowles & Peng, 2005). When studies provide additional categorization options, there is little to no support for ingroup overexclusion. For example, when the categorization task includes a third category option of "Multiracial," White American perceivers typically categorize the faces as Multiracial by a slim majority, and their remaining categorizations tend to be more frequently White than Black or White and Black in equal proportion (Chen & Hamilton, 2012; Gaither et al., 2018; Peery & Bodenhausen, 2008).

While Young et al. (2021) noted the discrepancy in results produced by two-group categorization tasks versus other categorization tasks, the available data could not provide an explanation for why this difference emerged. We argue that categorization tasks that include multiple outgroups either explicitly (e.g., adding a "Multiracial" category) or implicitly (e.g., through the use of open-ended responses) change the intergroup context; by increasing the numbers of outgroups salient, these conditions dilute the perceiver's social identity activated by a dichotomous intergroup context. As a result, we propose that intergroup contexts with more than two groups reduce the impact of social identification on racial categorization.

To test the impact of the number of groups on the extent to which identity underpins racial categorization, we investigate how strongly White Americans' racial identification predicted their categorizations of ambiguous faces across three tasks: a two-group categorization task (Black or White), a three-group categorization task (where the third option was Multiracial or Latino), and a completely unconstrained, open-ended categorization task (where participants generate a response to categorize the person's race). We hypothesized that participants' racial identification would only predict their ingroup overexclusion in the two-group categorization task and not in the three-group or open-ended categorization tasks. Consistent with previous work on ingroup overexclusion, we expected that White Americans' stronger racial identification would predict fewer categorizations of ambiguous faces as White (ingroup) in the two-group task.

In addition, ingroup overexclusion has been predominantly investigated among European samples and with respect to categorizing targets into national identities. The only studies to investigate ingroup overexclusion among White Americans include one study by Knowles and Peng (2005), which used an implicit measure of racial identification not directly comparable with the majority of the ingroup overexclusion literature, and a series of studies by Gaither et al. (2016), which focused on the impact of social exclusion on ingroup overexclusion. Therefore, the research reported here provides the most systematic study of White Americans' use of ingroup overexclusion to date.

Overview of Current Research

We tested our research question by meta-analyzing the entire contents of one lab's "file drawer" (Rosenthal, 1979). Analyzing the contents of researchers' file drawers is an important method for increasing transparency in the field (see Ioannidis, 2012), as it can result in more accurate estimates of effect sizes. In some cases, analysis of a lab's file drawer studies can provide valuable checks on false positives in the literature (see Lane et al., 2016).

Our meta-analysis included 14 studies (n = 4,001). Of these, 10 met the original criteria for inclusion, containing (a) a sample of White Americans, (b) a measure of racial categorization of ambiguous faces in two-group, three-group, and/or open-ended categorization tasks, and (c) at least one measure of racial identification. Studies could not include a manipulation that sought to impact the effect of the categorization task and/or relationship between racial identity and categorization. Four additional studies were run to address limitations of the previous 10 studies.

Because the raw data were available, a meta-analysis based on individual data points was preferred to the more typical approach of meta-analysis, which takes a weighted average of study effect sizes (Lambert et al., 2002; Stewart & Parmar, 1993). A meta-analysis with individual data points across multiple studies is also referred to as an integrative data analysis (see Curran & Hussong, 2009, for discussion of advantages; see ExTraMATCH Collaborative, 2004, for a prominent example using this approach). After combining the data, we examined whether White Americans' levels of racial identification predicted the likelihood that they categorized ambiguous faces as ingroup, and, critically, whether this association changed by categorization task condition. Because the data were categorical, we used logistic regression and estimated log odds to determine the relationship between racial identification strength and likelihood of making ingroup categorizations. An odds ratio significantly below 1.0 would indicate a negative relationship between racial identification and categorizations of the faces and would be evidence of ingroup overexclusion.

Table 1. Sample Characteristics for Studies 1 to 14.

Study	Year	*n*	Population	Gender (W, M, DNS)	Mean age (*SD*)	Categorization tasks (2, 3, OE)
1	2013	40	Undergraduate	30, 10	22.70 (5.40)	2, 3
2	2015	162	mTurk	77, 88	34.40 (11.03)	2, 3
3	2015	299	mTurk	143, 154, 2	34.18 (10.24)	2, OE
4	2016	440	mTurk	215, 223, 2	35.95 (11.38)	2, 3
5	2016	150	mTurk	96, 52, 2	37.95 (11.49)	2, 3, OE
6	2019	44	Undergraduate	27, 17	19.70 (1.92)	2, 3, OE
7	2019	175	mTurk	114, 61	37.96 (12.11)	2
8	2019	85	mTurk	48, 36, 1	39.24 (13.12)	2
9	2020	237	mTurk	138, 98, 1	37.43 (12.62)	2
10	2020	646	mTurk	296, 348, 1	40.09 (11.87)	2, 3M, 3L
11	2020	185	mTurk	88, 97	41.08 (13.51)	OE
12	2020	419	mTurk	216, 200, 3	41.69 (14.30)	2, OE
13	2020	396	mTurk	214, 180, 3	41.22 (13.30)	OE
14	2020	723	mTurk	403, 316, 3	41.08 (19.93)	2, 2L

Note. Samples only include White American participants. Gender response of "DNS" stands for "did not say" and includes *declined to state* and *other-identified* responses. Categorization task "OE" stands for an "open-ended" task in which participants respond to an open-ended question as to what race each target is. Categorization task "3L" indicates the three-group task with options Black, Latino, and White, and "3M" and "3" both refer to the three-group task with options: Black, Multiracial (or Biracial), and White. Categorization task "2L" indicates the two-group task with options: White and Latino.

Method

Survey measures, morphed stimuli, data, and code are publicly available (https://osf.io/9vnjp/). The analysis plan was not preregistered.

Study Inclusion Criteria

The first author included all available data sets from original studies that included White American perceivers, a categorization task of racially ambiguous faces, and at least one measure of racial identification. To be included, studies could not include an experimental manipulation intended to impact participants' responses in the categorization task. Some of the studies have small sample sizes because they were pilot tests for procedures or measures. Yet, we included all eligible data sets from the first author's lab to reduce selective reporting of studies that would reduce the validity of the results (Vosgerau et al., 2019). This search yielded 10 studies.

We ran four additional studies (Studies 11–14) that were also included in the omnibus analysis. Studies 11 to 13 were conducted to generalize the results to a new measure of racial identification. Prior to Study 11, all of the studies (1–10) used the same measure of racial identification, Luhtanen and Crocker's (1992) Collective Self-Esteem–Importance subscale that was adapted to measure racial identity. Studies 11–13 added a new measure of racial identification that was grounded in past research on White American identity specifically (Goren & Plaut, 2012; see Supplemental File for details on scale creation and validation). Study 14 was conducted to generalize results to another two-group task. Prior to Study 14, all studies (1–13) employed a two-group task with response options of either *White/Not White* or *White/Black*. Study 14 included *White/Black* and *White/Latino* between-subjects conditions to test whether category fit (Black being relatively worse fit than Latino) moderated the association between racial identification and ingroup categorization. Because there was no effect of the manipulation, the two conditions were combined.

Studies are numbered in chronological order. Refer to Table 1 for full details on each study sample and basic design. Table 1 provides sample characteristics after participant exclusions (see "Procedure" for details). We report all manipulations, measures, and exclusions for all studies.

Addressing Statistical Power

The main test of our hypothesis was the interaction between racial identification and categorization task (three-level factor) on ingroup categorizations. We planned to run a mixed effects logistic regression predicting ingroup categorizations from racial identification, categorization task, and the interaction term and with participants and stimuli as random factors. We used DeBruine and Barr's (2021) tutorial to simulate data with the present data's structure and calculate power using R (R Core Team, 2021). First, data sets were simulated with fixed and random effect parameter estimates based on the effects observed in the present data. Post hoc power indicated that all effects were powered at 1 except for the interaction between identification and task, which had post hoc power of .50. Ten simulations indicated that the minimum detectable effect at 80% power was somewhere between $b = 0.10$ and .15 for the interaction. Next, we determined power by simulating 496 data sets, fitting the model to each data set, and estimating how

many of the models rejected the null hypotheses. Results indicated that $b = 0.14$ was the minimum detectable effect for the focal interaction of task by racial identification for the current design, sample size, .80 power, and alpha = .05.

Materials

Stimuli. In Studies 1 to 4, participants viewed 40 prototypical Black faces, 40 racially ambiguous Black–White faces, and 40 prototypical White faces (20 men, 20 women per category). The racially ambiguous faces were created by morphing one Black and one White face together using Morpheus Photo Morpher (Version 3.17). The Black faces and White faces were created by morphing two Black faces or two White faces, respectively, together in the same method. This stimulus set was used in previous studies (Chen, Pauker, et al., 2018; Freeman et al., 2016). To shorten the duration of the study procedures, Study 6 used a subset of the faces from Studies 1 to 4. Specifically, we selected four ambiguous Black–White faces, four Black faces, and four White faces (two men and two women per category) for presentation in Study 6.[2] The complete set of 120 faces are available online (https://osf.io/9vnjp/).

The stimuli used for Studies 5 and 7 to 14 were exclusively racially ambiguous faces. These faces were of real Black–White biracial individuals (those who reported having one Black parent and one White parent, photographed in the lab against a white background with a neutral expression) that had been used and pretested for racial ambiguity in Gaither et al. (2019). The faces included four Black–White male faces and eight Black–White female faces, for a total of 12 faces.

Racial Identification. All 14 studies included the Collective Self-Esteem Importance subscale (CSE-Importance; Luhtanen & Crocker, 1992), and this was the measure of racial identification used in all analyses unless otherwise stated. This subscale was adapted to measure how important participants' racial/ethnic identity was to them (e.g., "My race/ethnicity is unimportant to my sense of what kind of person I am [reverse scored]" and "In general, belonging to my race/ethnicity is an important part of my self-image"). This measure was highly reliable across all of the studies (αs = .83–.88).

Studies 11 to 14 also included an original measure of White racial identification to provide convergent validity of the findings using CSE-Importance. The new measure was informed by Goren and Plaut's (2012) research on the dimensions of White identity and included three subscales that mapped onto these three dimensions: prideful identity (αs = .73–.77; e.g., "I am proud of my race"), power-cognizant identity (αs = .86-.91; "My race provides me with privileges and advantages"), and weak identity (αs = .66–.75; "I don't feel a strong sense of identity with my race"). Validation of the new measure and ancillary analyses with these measures are presented in the Supplemental File (see also, "Results: Follow-Up Tests For Robustness").

Procedure

In each study, participants learned that they would be categorizing faces by race. They viewed the faces in randomized order and their response options were manipulated by categorization task, as described in the next section and in Table 1. After all measurements, participants completed demographics and were thanked and debriefed. For the purposes of this research, only White-identifying participants were included in analyses.[3]

Only Study 1 was completed in a laboratory using Empirisoft's MediaLab and DirectRT programs, and participants completed racial identification measures in a prescreening survey before signing up for the study ($n = 40$).

Studies 2 to 14 were completed online using Qualtrics (www.Qualtrics.com). Participants provided informed consent before each study. In Studies 2, 9, 10 and 11, the racial identification items preceded the categorization task ($n = 1{,}230$). In Studies 4 to 6, the categorization task preceded the racial identification items ($n = 486$). Studies 3, 7, 8, and 12 to 14 counterbalanced the order of the categorization task and the racial identification items ($n = 1{,}527$).[4]

In Studies 12 and 13, after the categorization task, participants completed the Symbolic Racism scale and an exploratory item about their attitudes toward "Black Lives Matter."[5] Embedded in this block was an attention check question asking participants to select a certain response if they were reading the question. We excluded participants who did not pass this attention check ($n = 22$ in Study 12; $n = 6$ in Study 13).

In Study 14, 792 participants were recruited. After exclusions based on self-reported race and incomplete data, the final sample included 723 self-identified White Americans.

Design

Expanding on Table 1, Studies 1, 2, and 4 randomly assigned participants to the two-group (e.g., Black or White) or three-group (e.g., Black, White, or Multiracial) categorization task. Studies 3 and 12 randomly assigned participants to the two-group categorization task or the unconstrained, open-ended categorization task (e.g., What race is this person?). Studies 5 and 6 randomly assigned participants to the two-group, three-group, or unconstrained categorization task. Studies 7 to 9 only included the two-group categorization task. Study 10 had the two-group task and two versions of the three-group categorization task: (a) Black, Multiracial, or White and (b) Black, Latino, or White. Studies 11 and 13 only included the unconstrained categorization task.

Coding of Open-ended Responses

Participants' responses to in the open-ended categorization tasks were always coded by two raters who were unaware of the hypotheses and of participants' survey responses other

than their categorizations of faces. We created a coding scheme that would be inclusive of all participants' responses.

In Study 3, two White female coders used the following categories: Asian, Black, Latino/a, Middle Eastern, Multiracial, White, not race-related, and Other. Two different White female coders coded Studies 5, 6, 11, 12, and 13. In Studies 5, 6, 11, 12, and 13, responses were coded into the following categories: Asian, Black, Latinx, Middle Eastern, Multiracial, Native American and Pacific Islander, White, not race-related. (There was an increase in Native American/Pacific Islander categorizations beginning in 2016). Across the studies, coder agreement was extremely high (96%–99%), and one coder's responses were selected at random for use in analyses. In each study, we calculated the average number of categories used in the open-ended categorization task. We then tested whether the average was significantly greater than three using a one-sample *t*-test. In all cases, the average was significantly greater than three (M range: 3.14–4, ps $< .001$, d range: 0.94–1.30).

Results

Categorization Patterns by Task

Before testing for racial identification predicting ingroup overexclusion, it was helpful to know whether ingroup overexclusion occurred within each categorization task. Because the categorization task conditions differed in the number of non-White categories available to perceivers, we analyzed the baseline categorization patterns separately by categorization task.

Two-Category Task. We conducted a repeated measures analysis of variance (ANOVA) comparing the proportion of White categorizations and Black/Not White categorizations for all participants in the two-category condition. There was a significant effect of categorization type (White vs. non-White), $F(1, 2{,}028) = 670.18$, $p < .001$, $\eta_p^2 = .25$. Participants made significantly more non-White categorizations (Black or Not White responses; $M = 0.63$, $SD = 0.24$) compared with White categorizations ($M = 0.36$, $SD = 0.23$). Thus, ingroup overexclusion was supported.

Three-Category Task. We conducted a repeated measures ANOVA comparing the proportion of White, Multiracial, and Black categorizations made by participants in the three-category condition. There was a significant effect of categorization type, $F(2, 1{,}088) = 433.32$, $p < .001$, $\eta_p^2 = .44$. Participants made significantly more Multiracial categorizations ($M = 0.57$, $SD = 0.23$) than White categorizations ($M = 0.17$, $SD = 0.18$) or Black categorizations ($M = 0.24$, $SD = 0.19$), ps $< .001$. They also made significantly more Black categorizations than White categorizations, $p < .001$. These results are also consistent with ingroup overexclusion, in that both Black and Multiracial categorizations were significantly greater than White categorizations.

Open-Ended Task. We conducted a repeated measures ANOVA comparing the proportion of White, Multiracial, Latino, and Black categorizations made by participants in the open-ended conditions. There was a significant effect of categorization type, $F(3, 3{,}138) = 422.86$, $p < .001$, $\eta_p^2 = .29$. Follow-up comparisons indicated that Black ($M = 0.37$, $SD = 0.22$) was significantly more common than the other categorizations of faces, followed by Latino ($M = 0.29$, $SD = 0.22$), which was significantly more common than White ($M = 0.20$, $SD = 0.20$), all ps < 001. Multiracial was the least frequent categorization and was significantly less frequent than any other categorization ($M = 0.04$, $SD = 0.12$), all ps $< .001$. These findings support ingroup overexclusion in that it was relatively rare for participants to categorize faces as White (20% of responses) compared with non-White (80% of responses, varying in exact category selected).

Main Analysis: Predicting Ingroup Categorizations from Racial Identification and Categorization Task

We first structured the data in extra-long form, assigning each stimulus face and each participant with unique identifying numbers, before combining into a single file. Then, we estimated the mixed effects logistic regression models in R (R Core Team, 2021) using lme4 (Bates et al., 2015).

The logistic regression model predicted the binary outcome of ingroup (White) categorization (vs. a non-White categorization) from the participant-level fixed effects of categorization task (two dummy codes with the two-group task as the reference group), racial identification (grand-mean-centered), and the focal task by identification interaction (represented by the two multiplicative terms: identification by dummy code 1 and identification by dummy code 2). Random intercepts for face stimuli and participants were also included to estimate effects due to variation in stimuli and perceivers (Judd et al., 2012).[6]

The model showed a significant difference in ingroup categorizations between the two-group and three-group conditions, $b = -0.85$ ($SE = .09$), OR = 0.43 (95% CI: 0.36, 0.51), $p < .001$, and between the two-group and open-ended conditions, $b = -1.58$ ($SE = .08$), OR = 0.21 (95% CI: 0.18, 0.24), $p < .001$. The two-group condition increased the likelihood of ingroup categorizations relative to the other conditions. This is likely due to base rate differences, specifically the fact that ingroup categorizations were one of only two possible responses in the two-group condition, compared with being one of three possible responses in the three-group condition and one of theoretically unlimited possible responses in the open-ended condition.

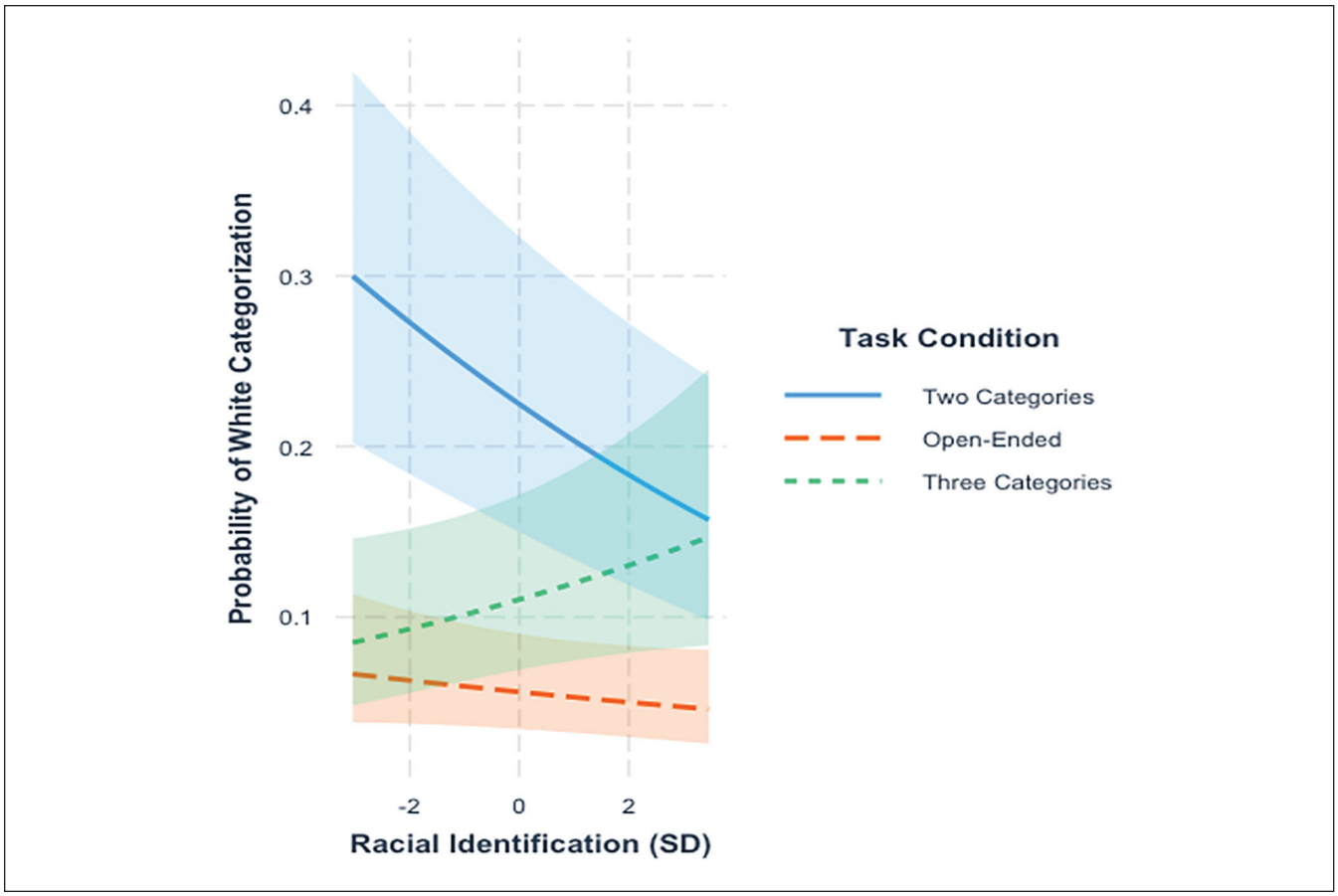

Figure 1. Probability of White Categorizations, With 95% Confidence Intervals, by Participants' Level of Racial Identification and Categorization Task Condition.

The model also revealed a significant interaction between racial identification and one categorization task dummy code (two- vs. three-group task), $b = .22$ ($SE = .06$), OR = 1.25 (95% CI: 1.11, 1.41), $p < .001$, indicating that the association between racial identification and ingroup categorization differed significantly between the two-group and three-group tasks. The association between racial identification and ingroup categorizations did not differ significantly between the two-group and open-ended conditions, $b = .07$ ($SE = .05$), OR = 1.07 (95% CI: 0.96, 1.19), $p = .21$. Then, we re-estimated the model with the three-group task as the reference group to estimate the racial identification by categorization task (three-group vs. open-ended). The interaction term was significant, $b = -.16$ ($SE = .07$), OR = 0.86 (95% CI: 0.74, 0.99), $p = .03$. Thus, the three-group task changed the association between racial identification and ingroup categorizations compared with the two-group task and open-ended task (see Figure 1).

Finally, we computed odds-ratios of racial identification predicting the likelihood of ingroup categorizations separately for each categorization task. In the two-group task, participants with stronger racial identification were less likely to make ingroup categorizations, OR = .88 (95% CI: .83, .93), $p < .001$, supporting ingroup overexclusion. Yet participants' racial identification did not significantly predict ingroup categorizations in the open-ended task, OR = 0.941 (95% CI: .86, 1.03), $p = .20$, or in the three-group task, OR = 1.10 (95% CI: .99, 1.23), $p = .08$. Therefore, although racial identification did predict fewer White categorizations in the two-group task, this slope was not significantly different from the slope in the open-ended condition, which was not significantly different from zero.

Follow-Up Tests for Robustness. We conducted several follow-up analyses to examine the robustness of our findings. The details of nonsignificant analyses are in the Supplemental File.

Robustness to Construct Operationalization. The first analysis re-ran the main analysis using a latent variable of racial identification that used confirmatory factor analysis to estimate a predicted value for each participant based on their responses to the CSE-import and two subscales of White identity that we had created for this research. The model only

revealed significant differences between task conditions, therefore providing no support for identity-driven ingroup overexclusion.

Robustness to Methodological Details. The next series of analyses tested for moderation of the focal effect—the racial identification by categorization task interaction—by the following variables, separately: (a) year of data collection (pre-2020 vs. 2020 or later) to account for variability due to the COVID-19 pandemic, (b) sample type (undergraduate vs. online), (c) facial stimulus set (including prototypical faces vs. not), and (d) stimulus gender (women vs. men). In each model, the new variable was included as a predictor and as a moderator. The primary test of robustness was the three-way interaction between the new variable, task, and racial identification. Models 1 to 2 did not qualify the main findings and are reported in the Supplemental File.

In Model 3, the logistic regression model predicted the binary outcome of ingroup (White) categorization (vs. a non-White categorization) from the stimulus-level fixed effect of facial stimulus set (prototypical faces excluded vs. included), participant-level fixed effects of categorization task (two dummy codes with the two-group task as the reference group), racial identification (grand-mean-centered), the two focal task by identification interaction terms, and all two-way and three-way interaction terms. Random intercepts for face stimuli and participants were also included to estimate effects due to variation in stimuli and perceivers (Judd et al., 2012). Stimulus set did moderate the task by racial identity interactions for the two-category vs. open-ended task, $b = -.29$, $z = -2.08$, $p = .04$. It did not moderate the two-group vs. three-group, $b = -.09$, $z = -.71$, $p = .48$, comparison. Simple slopes revealed that, in the open-ended task, racial identification did predict fewer White categorizations when prototypical faces were included, $b = -.37$, $SE = .11$, $p < .001$, but not when prototypical faces were excluded, $b = .001$, $SE = .05$, p group $= .05$. In the two-group task, there was a stronger association between racial identification when prototypical faces were included, $b = -.20$, $SE = .05$, $p < .001$, than when they were excluded, $b = -.11$, $SE = .03$, $p < .001$. In both three-group task conditions, identity did not predict categorization, ps $> .10$. Thus, including prototypical faces in the categorization task increased support for identity-driven ingroup overexclusion in both the two-category and open-ended categorization tasks. When prototypical faces were not included, the open-ended task did not provide support for ingroup overexclusion, consistent with the main analysis.

In Model 4, the logistic regression model was identical to Model 3 except that it included the stimulus-level fixed effect of gender (men vs. women) in lieu of the facial stimulus set factor. Stimulus gender moderated the interaction between racial identity and task (two-group vs. open-ended), $b = -0.14$, $SE = .05$, $z = -2.99$, $p = .003$. Simple slopes indicated that, consistent with the main analysis, racial identification predicted fewer White categorizations for female stimuli, $b = -.13$, $SE = .03$, $p < .01$, and male stimuli, $b = -.12$, $SE = .03$, $p < .01$ in the two-group task. In the open-ended task, racial identification did not predict White categorizations for female stimuli, $b = -.03$, $SE = .05$, $p = .61$, consistent with the main analysis. However, racial identification did predict fewer White categorizations for male stimuli, $b = -.14$, $SE = .04$, $p = .01$, in the open-ended task. (In the three-group task, racial identity did not predict White categorizations for either stimulus gender, ps $> .09$.)

Robustness to Chance-Level Responding. We conducted exploratory analyses using a dependent measure of White categorizations standardized across conditions. Specifically, participants' number of White categorizations was divided by the number of White categorizations that would be generated by chance (50% in the two-group condition, 33% in the three-group condition). As a result, we had an index of White categorizations that was corrected for chance levels and more directly comparable across conditions, with higher scores indicating relatively more inclusion in the White category.

Because there is no clear operationalization of "chance levels" in the open-ended tasks, we conducted two versions of this analysis, the first comparing only two-group and three-group tasks, and the second adding the open-ended categorization indices (uncorrected proportions of responses) using the Process macro (version 4.1; Hayes, 2022). In both cases, the results are consistent with the main analyses.

Two-Category vs. Three-Category Tasks, We regressed the corrected measure of White categorizations on White participants' racial identification, categorization task (two-group vs. three-group), and their interaction. Significant effects of identification, $b = -0.06$ ($-.10$, $-.02$) $SE = .02$, $p = .002$, and categorization task, $b = -0.35$ ($-.45$, $-.24$), $SE = .06$, $p < .001$, were qualified by a significant interaction, $b = .04$ (.01, .06), $SE = .01$, $p = .01$. Simple slopes indicated that racial identity was negatively associated with White categorizations in the two-group task, $b = -.03$ ($-.04$, $-.01$), $SE = .01$, $p = .001$, but not in the three-group task, $b = .01$ ($-.01$, .04), $SE = .01$, $p = .38$. Therefore, stronger racial identification predicted fewer White categorizations (more exclusion) in the two-group task but not in the three-group task. These findings remained significant when controlling for year of data collection.

All Three Tasks. We regressed the measure of White categorizations (corrected for chance levels in the closed-ended categorization tasks; proportion of responses that were "White" in open-ended task) on White participants' racial identification, categorization task (two-group, three-group, open-ended; dummy-coded with the two-group task as the reference group), and their interaction. The interactions of the dummy-coded task variables with racial identity were both significant: dummy code 1 (three-group task vs. others) by racial

identity (numbers) and dummy code 2 (open-ended task vs. others) by racial identity (numbers). Simple slopes revealed that racial identity predicted fewer White categorizations in the two-group task, $b = -.02$ (−.04, −.01), $SE = .007$, $p = .001$, but not in the three-group task, $b = -.01$ (−.01, .03), $SE = .01$, $p = .27$, or the open-ended task, $b = .01$ (−.01, .03), $SE = .009$, $p = .39$. When the year of data collection was included as a covariate in the model, the results were unchanged.

General Discussion

Using a meta-analytic approach, we provide the most comprehensive investigation of White Americans' ingroup overexclusion to date. Across 14 studies, we investigated the extent to which White Americans' level of racial identification predicted their categorization of racially ambiguous faces.

Our meta-analysis provided evidence that the categorization task moderated the role of White Americans' identification in their racial categorizations. Specifically, White Americans' level of racial identification predicted ingroup categorization differently depending on the categorization task used. The strongest result was that the three-group condition affected the relationship between racial identification and ingroup categorization differently than the two-group and open-ended conditions. In the three-group condition, there was no association between having stronger White identity and ingroup categorizations (if anything, there was a marginal pattern in the opposite direction). These results indicate that White Americans' racial identification levels did not predict their decision to include or exclude ambiguous individuals when they were in a three-group context. These findings indicate that adding an additional outgroup reduces identity-driven categorization, providing some support for the proposed dilution effect.

However, the hypothesized dilution effect was not completely supported. The association between racial identification and ingroup categorization in the open-ended categorization task, in which the number of groups is theoretically unlimited, was not significantly different from the two-group task. The dilution effect would have predicted a significantly reduced association between identification and ingroup categorization in the unconstrained intergroup context (open-ended task) relative to the dichotomous intergroup context.

Overall, our results provide lackluster support for the role of racial identification in ingroup overexclusion. Although White Americans with stronger racial identification were less likely to make ingroup categorizations in the two-category task (supporting the ingroup overexclusion effect), the effect was small and not significantly different from the null relationship between the two variables documented in the open-ended task. In fact, in both the three-group and open-ended categorization tasks, we documented no significant association between racial identification and categorization on average. Furthermore, analyses that used a latent variable approach to measuring racial identification produced even weaker support for ingroup overexclusion (see Supplemental File for details), revealing no significant association between identity and categorization across the entire sample.

Because White Americans' overexclusion of racially ambiguous faces was not consistently predicted by their racial identification, our results raise the question about what factors do explain the general tendency to categorize faces as outgroup over ingroup. Among White Americans, ingroup overexclusion of ambiguous faces are also consistent with the socially and historically significant pattern of *hypodescent*, which is the categorization rule by which mixed race individuals are considered to be members of their socially subordinated group membership (Davis, 1991; Ho et al., 2017; Peery & Bodenhausen, 2008). Also, to this day, Black–White biracial people are more likely to self-identify as Black (Pew, 2015). It is possible that the categorization pattern documented here reflects the historical legacy of hypodescent in the U.S. and the racial identification preferences of Black–White biracial people on average.

The follow-up analyses revealed additional nuance to our findings. Namely, identity-driven categorization was supported in the open-ended task for male faces (not female faces) and when prototypical faces were also included in the stimulus set. These factors had muted to null effects in the closed-ended categorization tasks. Thus, our results suggest that unconstrained categorization processes are more heterogeneous and multiply determined than categorization in closed-ended tasks. In the open-ended task, that ingroup overexclusion was only applied to male faces is consistent with the theory of gendered prejudice (McDonald et al., 2011), which proposes that prejudices are more frequently directed toward male group members than female group members (cf. Sidanius et al., 2018). Moreover, that inclusion of racially prototypical faces increased identity-driven ingroup overexclusion of ambiguous faces may be consistent with the idea that brief visual exposure can shift perceptual norms, such that viewing prototypical faces sharpened racial category boundaries whereas viewing only ambiguous faces may have dulled those boundaries (Lick & Johnson, 2014). That these findings were observed primarily in the open-ended categorization task increases our conviction that closed-ended categorization tasks can obscure meaningful categorization processes that could be observed in the real world.

There are limitations to this work that should be addressed in future research. One limitation is its focus on face perception. A large portion of the literature on ambiguous categorization uses nonvisual paradigms, such as presenting participants with a hypothetical target about whom information is manipulated and asking participants to categorize the target (e.g., Ho et al., 2017) or presents both visual and other autobiographical information (Nam & Chen, 2021). Our

findings only speak directly to the categorization processes engaged when perceivers are categorizing facial stimuli. It is reasonable to expect that the psychological processes involved in deciding how to weigh different racial ancestries in a generic target may differ from those involved in deciding how to racially categorize an ambiguous face. Yet, it is important to note that, in everyday life we routinely base our perceptions of a person on their face (Jaeger et al., 2019). Rarely do we have specific information about the person's ancestry, and even when both facial and ancestral information are provided to the perceiver, visual information is often weighed more heavily in perceivers' categorizations (Skinner & Nicolas, 2015).

Another important limitation of the present research was its focus on White Americans. While it was appropriate for our investigation to focus on White Americans because we hoped to shed light on discrepant published findings with this population, an important unanswered question concerns the extent to which our findings generalize beyond this group. For example, it is possible that American people of color are more likely to "carry over" race-based motivations to racial categorizations in unconstrained contexts. The possibility would further shed light on when perceiver attributes impact categorization processes and warrants future research (see also Ho et al., 2020).

Despite these limitations, the present work has important implications for the racial categorization and social identity literatures. With respect to Multiracial categorization, our research provides a clearer understanding of the documented discrepancy in this literature, namely, that how White American perceivers categorize ambiguous faces depends highly on the number of categories immediately accessible. In the real world, there are constraints on the categories available to perceivers based on situations (e.g., base rates of particular groups), experience (e.g., learning history, contact/exposure to diversity), and disposition (e.g., motivated attention). Our research suggests that narrowing or broadening the number of salient categories will impact how perceivers categorize ambiguous others.

Perhaps reflecting the natural variability in how many categories are available to perceivers, our findings indicate that neither the two-group nor three-group task are equivalent to the open-ended categorization context for White American perceivers. The three-group task increases the likelihood of Multiracial categorizations (see also Gaither et al., 2019; Nicolas et al., 2019; Peery & Bodenhausen, 2008), and it disrupts the relationship between White identity and categorization. Similarly, the two-group task may overestimate the strength of the relationship between White identity and categorization, as this condition was the only one in which a reliable relationship was found between the two variables. Thus, neither findings from the two-group nor three-group categorization tasks may adequately capture racial categorization in the real world. Thus, more research is needed to understand how racial categorization operates when perceivers do not have task constraints. The open-ended categorization task in the present set of studies gives researchers some insight into results that may be more reflective of real-world contexts. Additional research determining which racial categories naturally come to perceivers' minds when facing racially ambiguous faces (e.g., Chen, Pauker, et al., 2018) would provide a valuable foundation for future research in this area.

Consistent with previous findings (Chen, Pauker, et al., 2018; Nicolas et al., 2019), the three-group task increases the likelihood of the Multiracial categorization relative to the open-ended categorization task. Adding to the literature, our results also show that White Americans' racial identity did not predict their responses within the three-group context. We speculate that making the Multiracial category salient could reduce White Americans' motivation to protect the ingroup in a few different ways. First, presentation of an unusual or atypical category could increase surprise or the motivation for accuracy (see Crisp & Turner, 2011; Hutter et al., 2009), thereby increasing systematic processing of all faces. Second, presentation of the Multiracial category could prime participants with concepts of racial harmony, overlap, and interdependence (see Gaither et al., 2019; Pauker et al., 2017; Sanchez et al., 2015; Wilton et al., 2014), thereby reducing their motivation to protect the ingroup. Relatedly, it is possible that when perceivers' focus is shifted from a single dimension of identity (e.g., race) into multiple dimensions of identity (e.g., race and gender), there may be reduced impact of race-based motivations and beliefs. These possibilities represent important avenues for future research.

With respect to intergroup relations research more broadly, our findings suggest that the two-group categorization task might activate social identities in ways that other tasks do not. At best, our results provide limited support for identity-driven ingroup overexclusion. For some, the results presented here will question the underlying process of ingroup overexclusion, which has been conceptualized as a process rooted in ingroup identity from its discovery, when researchers described the effect as occurring "because their [perceivers'] social identity is put at stake" (Leyens & Yzerbyt, 1992, p. 551). Thus, without underlying evidence for the role of social identity in the categorization process, as was documented here in the three-group and open-ended conditions, there is weak support for the original conceptualization of ingroup overexclusion as an identity-driven phenomenon. Future research could help to clarify the generalizability of ingroup overexclusion using other measures of ingroup identification and group categorization.

The notion that the numbers of groups within an intergroup context can differentially activate social identity processes is an important idea that was proposed decades ago (e.g., Oakes, 1987). Yet to our knowledge, ours is one of the first direct tests of this assertion. Given the increasing diversity in modern societies, intergroup contexts with only two groups are becoming less frequent. It would therefore be fruitful for researchers to investigate how the number of

groups can shift the activation of social identity motivations, cognitions, and behaviors.

In conclusion, 14 studies revealed that the number of groups salient determines whether White Americans' racial identification influenced their categorizations of ambiguous faces. Our results raise important questions for the literatures on categorization, face perception, and intergroup relations, and highlight the important ways that methodological choices can impact scientific conclusions drawn.

Acknowledgments

The authors are grateful to Brian Baucom, Samantha Joel, Becky Neufeld, and Sabrina Thai for their guidance and feedback on the analyses in this manuscript.

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) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Jacqueline M. Chen https://orcid.org/0000-0001-9347-3816
Chanel Meyers https://orcid.org/0000-0002-7755-5413

Supplemental Material

Supplemental material is available online with this article.

Notes

1. Our focus on White Americans was driven by the fact that past research focusing on this population has discrepant findings (described below; see also Pauker et al., 2018; Young et al., 2020) that we will attempt to resolve.
2. The subset of stimuli were selected by the researchers to be even with respect to racial category (Biracial, Black, and White) and gender (men, women). These faces were chosen without any particular criteria in mind. The same subset of faces was shown to each participant.
3. Additional measures that were included in individual studies are reported in the Supplemental File.
4. We examined whether the order of the measures influenced the correlation between racial identification and categorization. Out of the 11 tests conducted, only one was significant: in Study 13, order moderated the correlation between CSE-Importance and categorization, $b = 0.49$, $p = .004$. When the categorization task was first, CSE-Importance did not predict categorizations, $b = -0.11$, $p = .34$. When racial identity was measured first, CSE-Importance did predict categorizations, $b = 0.38$, $p = .002$. Because this order effect was not replicated in Studies 3, 7, 8, or 12, we do not consider it to be reliable. Given the large number of null order effects, we collapsed across measure order when estimating the correlation between racial identification and ingroup categorizations in the main analyses.
5. Study 12 was run on May 31, 2020, at the height of protests against police brutality and for Black Lives Matter in the U.S. Study 13 was conducted on June 11, 2020, as the protests and momentum for Black Lives Matter continued. In Studies 12 and 13, after the categorization task, participants completed the Symbolic Racism scale and an exploratory item about their attitudes toward "Black Lives Matter." The BLM attitude item did not correlate with participants' categorizations of ambiguous faces. In Study 12, there was a three-way interaction between BLM attitudes, power-cognizant identity, and categorization task, $p = .01$. The interaction was driven by a significant power-cognizant identity by BLM attitude interaction in the two-group task, $p = .003$, that was not present in the open-ended condition, $p = .36$. In the two-category task, among participants who had more pro-BLM attitudes, power-cognizant identity predicted lower likelihood of categorizing ambiguous faces as White, $b = -0.55$ (95% CI: −.98, −.12), $SE = .22$, $p = .01$. This pattern is consistent with our hypothesis that social identities are more salient in the two-category contexts than in others.
6. Initially, we also included random slopes of stimuli and participants on categorization task in the models. However, these models failed to converge in R and in Stata. When models are forced to converge by lowering convergence criteria, estimates of standard error and statistical significance are no longer reliable (Raudenbush & Bryk, 2002). Thus, as recommended in cases of nonconvergence (Judd et al., 2012; Raudenbush & Bryk, 2002), we pruned the model to achieve convergence by focusing on the factors of theoretical significance.

References

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

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

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition: Vol. 1. A dual process model of impression formation* (pp. 1–36). Lawrence Erlbaum.

Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. Ashmore, L. Jussim, & D. Wilder (Eds.), *Social identity, intergroup conflict, and conflict reduction* (pp. 17–41). Oxford University Press.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. Wiley.

Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology*, *38*(3), 315–322. https://doi.org/10.1006/jesp.2001.1512

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. https://doi.org/10.1177/1948550617725149

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, *48*(1), 152–164. https://doi.org/10.1016/j.jesp.2011.10.005

Chen, J. M., Norman, J. B., & Nam, Y. (2021). Broadening the stimulus set: Introducing the American multiracial faces database. *Behavior Research Methods*, *53*, 371–389. https://doi.org/10.3758/s13428-020-01447-8

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.

Crisp, R. J., & Turner, R. N. (2011). Cognitive adaptation to the experience of social and cultural diversity. *Psychological Bulletin*, *137*(2), 242–266.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100.

Davis, F. J. (1991). *Who is black? One nation's definition*. Pennsylvania State University Press.

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–15.

Esses, V. M., Jackson, L. M., & Armstrong, T. L. (1998). Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of Social Issues*, *54*, 699–724.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *In Advances in Experimental Social Psychology*, *23*, 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*, *27*(4), 502–507. https://doi.org/10.1177/0956797615627418

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2018). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social Psychology*, *159*(5), 592–610. https://doi.org/10.1080/00224545.2018.1538929

Gaither, S. E., Pauker, K., Slepian, M. L., & Sommers, S. R. (2016). Social belonging motivates categorization of racially ambiguous faces. *Social Cognition*, *34*(2), 97–118. https://doi.org/10.1521/soco.2016.34.2.97

Gaither, S. E., Toosi, N. R., Babbitt, L. G., & Sommers, S. R. (2019). Exposure to biracial faces reduces colorblindness. *Personality and Social Psychology Bulletin*, *45*(1), 54–66. https://doi.org/10.1177/0146167218778012

Goren, M. J., & Plaut, V. C. (2012). Identity form matters: White racial identity and attitudes toward diversity. *Self and Identity*, *11*, 237–254. https://doi.org/10.1080/15298868.2011.556804

Hartstone, M., & Augoustinos, M. (1995). The minimal group paradigm: Categorization into two versus three groups. *European Journal of Social Psychology*, *25*(2), 179–193.

Ho, A. K., Kteily, N. S., & Chen, J. M. (2017). "You're one of us": Black Americans' use of hypodescent and its association with egalitarianism. *Journal of Personality and Social Psychology*, *113*(5), 753–768.

Ho, A. K., Kteily, N. S., & Chen, J. M. (2020). Introducing the sociopolitical motive × intergroup threat model to understand how monoracial perceivers' sociopolitical motives influence their categorization of multiracial people. *Personality and Social Psychology Review*, *24*(3), 260–286. https://doi.org/10.1177/1088868320917051

Ho, A. K., Sidanius, J., Cuddy, A. J. C., & Banaji, M. R. (2013). Status boundary enforcement and the categorization of Black-White multiracials. *Journal of Experimental Social Psychology*, *49*(5), 940–943. https://doi.org/10.1016/j.jesp.2013.04.010

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. https://doi.org/10.1037/a002156

Hutter, R. R., Crisp, R. J., Humphreys, G. W., Waters, G. M., & Moffitt, G. (2009). The dynamics of category conjunctions. *Group Processes & Intergroup Relations*, *12*(5), 673–686.

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*(6), 645–654.

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, *148*(6), 1008–1021. https://doi.org/10.1037/xge0000591

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.

Knowles, E. D., & Peng, K. (2005). White selves: Conceptualizing and measuring a dominant-group identity. *Journal of Personality and Social Psychology*, *89*(2), 223–241. https://doi.org/10.1037/0022-3514.89.2.223

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Science*, *111*(25), 9079–9084. https://doi.org/10.1073/pnas.1404448111

Krosch, A. R., Bernsten, 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*(6), 1196–1203. https://doi.org/10.1016/j.jesp.2013.05.009

Lambert, P. C., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, *55*(1), 86–94.

Lane, A., Luminet, O., Nave, G., & Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology*, *28*(4).

Leyens, J. P., & Yzerbyt, V. Y. (1992). The in-group overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, *22*(6), 549–569. https://doi.org/10.1002/ejsp.2420220604

Lick, D. J., & Johnson, K. L. (2014). Recalibrating gender perception: Face aftereffects and the perceptual underpinnings of gender-related biases. *Journal of Experimental Psychology: General*, *143*(3), 1259–1276.

Love, B. C. (2017). Concepts, meaning, and conceptual relationships. In S. E. F. Chipman (Ed.), *The Oxford handbook of cognitive science* (pp. 137–150). Oxford University Press.

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318. https://doi.org/10.1177/0146167292183006

McDonald, M. M., Navarrete, C. D., & Sidanius, J. (2011). Developing a theory of gendered prejudice: An evolutionary and social dominance perspective. In M. M. McDonald, C. D. Navarrete, & J. Sidanius (Eds.), *Social cognition, social identity, and intergroup relations* (pp. 192–223). Psychology Press.

Nam, Y., & Chen, J. M. (2021). Are you one of us: Investigating cultural differences in determining group membership. European *Journal of Social Psychology*, *51*(7), 1113–1137.

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*(4), 532–541. https://doi.org/10.1177/1948550618769591

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Oakes, P. J. (1987). The salience of social categories. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group* (pp. 117–141). Blackwell.

Pauker, K., Meyers, C., Sanchez, D. T., Gaither, S. E., & Young, D. M. (2018). A review of multiracial malleability: Identity, categorization, and shifting racial attitudes. *Social and Personality Psychology Compass*, *12*(6), Article e12392. https://doi.org/10.1111/spc3.12392

Peery, D., & Bodenhausen, G. (2008). Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science*, *19*(10), 973–977. https://doi.org/10.1111/j.1467-9280.2008.02185.x

Pew Research Center. (2015). Multiracial in America: Proud, Diverse and Growing in Numbers. Washington, D.C.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 241–248.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). SAGE.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Sanchez, D. T., Young, D. M., & Pauker, K. (2015). Exposure to racial ambiguity influences lay theories of race. *Social Psychological and Personality Science*, *6*(4), 382–390.

Sherif, M. (1966). *Group conflict and cooperation*. Routledge Kegan Paul.

Sidanius, J., Hudson, S. T. J., Davis, G., & Bergh, R. (2018). The theory of gendered prejudice: A social dominance and intersectionalist perspective. In Mintz & Terris (Eds.), *The Oxford handbook of behavioral political science* (pp. 1–35). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190634131.013.11

Skinner, A. L., & Nicolas, G. (2015). Looking Black or looking back? Using phenotype and ancestry to make racial categorizations. *Journal of Experimental Social Psychology*, *57*, 55–63.

Spielman, D. A. (2000). Young children, minimal groups, and dichotomous categorization. *Personality and Social Psychology Bulletin*, *26*(11), 1433–1441.

Stewart, L. A., & Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference? *The Lancet*, *341*(8842), 418–422.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, *25*, 79–97. https://doi.org/10.1111/j.1540-4560.1969.tb00620.x

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. *Psychology of Intergroup Relations*, *25*(1), 7–24. https://doi.org/10.1111/j.2044-8309.1986.tb00732.x

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Turner, J. C., & Reynolds, K. J. (2011). Self-categorization theory. *Handbook of Theories in Social Psychology*, *2*(1), 399–417.

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General*, *148*(9), 1628–1639. https://doi.org/10.1037/xge0000663

Wilton, L., Sanchez, D. T., & Giamo, L. (2014). Seeing similarity or distance? Racial identification moderates intergroup perception after biracial exposure. *Social Psychology*, *45*(2), 127–134. https://doi.org/10.1027/1864-9335/a000168

Young, D. M., Sanchez, D. T., Pauker, K., & Gaither, S. E. (2021). A meta-analytic review of hypodescent patterns in categorizing multiracial and racially ambiguous targets. *Personality and Social Psychology Bulletin*, *47*(5), 705–727. https://doi.org/10.1177/0146167220941321

Yzerbyt, V. Y., Leyens, J.-P., & Bellour, F. (1995). The in-group overexclusion effect: Identity concerns in decisions about group membership. *European Journal of Social Psychology*, *25*(1), 1–16. https://doi.org/10.1002/ejsp.2420250102