Mem Cogn (2013) 41:1021–1031
DOI 10.3758/s13421-013-0316-7

Same faces, different labels: Generating the cross-race effect in face memory with social category information

Kathleen L. Hourihan · Scott H. Fraundorf · Aaron S. Benjamin

Published online: 2 April 2013

Abstract Recognition of own-race faces is superior to recognition of other-race faces. In the present experiments, we explored the role of top-down social information in the encoding and recognition of racially ambiguous faces. Hispanic and African American participants studied and were tested on computer-generated ambiguous-race faces (composed of 50 % Hispanic and 50 % African American features; MacLin & Malpass, Psychology, Public Policy, and Law 7:98–118, 2001). In Experiment 1, the faces were randomly assigned to two study blocks. In each block, a group label was provided that indicated that those faces belonged to African American or to Hispanic individuals. Both participant groups exhibited superior memory for faces studied in the block with their own-race label. In Experiment 2, the faces were studied in a single block with no labels, but tested in two blocks in which labels were provided. Recognition performance was not influenced by the labeled race at test. Taken together, these results confirm the claim that purely top-down information can yield the well-documented cross-race effect in recognition, and additionally they suggest that the bias takes place at encoding rather than testing.

Keywords Recognition · Memory · Face processing

The cross-race effect (CRE) in memory (also referred to as the *other-race effect* or *own-race bias*) refers to the well-replicated finding that humans are better at remembering faces from their own racial group, relative to other groups (e.g., Malpass & Kravitz, 1969). There is also evidence for superior memory for faces from one's own age group (e.g., Rhodes & Anastasi, 2012), and perhaps from one's own gender (though this may be limited to females; see, e.g., Slone, Brigham, & Meissner, 2000). There is still no real consensus as to exactly why the CRE occurs.

The goal of the present work was to weigh in on the debates over the origins of this effect. According to one class of theories, the CRE is a consequence of perceptual expertise with faces that are characteristic of one's own group. The other class of theories proposes that the act of categorizing a face as belonging to another group entails changes to the way that we encode those faces. In the present experiments, we tested people's memory for the exact same faces that were paired with name information that categorized them as belonging to the same group as the participant or to another group. By holding the perceptual information constant across conditions, memory effects could be convincingly attributed to induced changes in the encoding process and not to perceptual expertise. When the name information was introduced during study, a CRE was observed; when it was introduced at test, a CRE was not observed. These results suggest that the social or cognitive biases introduced by group classification are sufficient to yield the CRE.

Competing accounts of the CRE

Perceptual-expertise accounts of the CRE (e.g., Rhodes, Lie, Ewing, Evangelista, & Tanaka, 2010; Valentine, 1991; Valentine, Chiroro, & Dixon, 1995) rely on the fact that most people have more experience perceiving, encoding, and remembering faces from their own group. We therefore

K. L. Hourihan (✉)
Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X9, Canada
e-mail: khourihan@mun.ca

S. H. Fraundorf · A. S. Benjamin
Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

simply lack the skill to properly remember other-race faces because we have not had sufficient experience to learn how to differentiate among other-race faces at the time of encoding. Alternatively, social–cognitive accounts of the CRE focus on differential encoding of social in-group and out-group faces (see, e.g., Sporer, 2001). According to these theories, faces are first classified (very rapidly; see, e.g., Levin, 1996, 2000) as belonging to either our own social in-group or a social out-group. In-group members are socially more important and tend to be encoded on the basis of individuating features, supporting superior memory performance subsequently. Faces identified as out-group members, however, are not processed in an individuating manner, and only group-identifying features tend to be encoded. This less-differentiating manner of encoding leads to poor subsequent recognition, with difficulty in discriminating previously encountered individuals from new individuals from the same out-group (Sporer, 2001). Recently, Hugenberg and colleagues (e.g., Hugenberg, Young, Bernstein, & Sacco, 2010; Young & Hugenberg, 2012) proposed the categorization–individuation model to explain how both motivation and experience interact in modulating the CRE (see also Young, Hugenberg, Bernstein, & Sacco, 2012, for a review of current theories of the CRE).

Support for social–cognitive theories has been obtained from studies in which researchers have used perceptually ambiguous face stimuli (e.g., MacLin & Malpass, 2001). When the social or racial group of a target face is ambiguous on the basis of the perceptual features inherent in the face (i.e., by computer generation of composite faces or morphing two pictures of faces from different races), the impact of top-down processing instantiated by contextual information can be examined with less influence from more automatic processing influences. That is, it is possible to examine the extent to which the perception and recognition of faces can be influenced by the viewer's a priori beliefs about whether the target face belongs to his or her own social in-group or an out-group.

Of critical interest for the present study is how recognition of ambiguous faces is influenced by various factors present at encoding. First, MacLin and Malpass (2001) showed that participants will rely on a single perceptual feature present in ambiguous faces—a hairstyle—in categorizing ambiguous faces, and that this influences subsequent recognition. Specifically, false alarms were significantly higher to new faces with the other-race hairstyle than to new faces with the own-race hairstyle, indicating that recognition decision processes are influenced by the perceived race of an ambiguous face, on the basis of a single perceptual feature. In considering nonperceptual, social factors, however, the evidence is somewhat mixed. Corneille, Huart, Becquart, and Brédart (2004, Exp. 1) presented Caucasian participants with a single face, morphed to be ambiguous between Caucasian and North African, along with social information consistent with one of the groups (i.e., either Caucasian or North African), for the purpose of an impression formation task. A subsequent surprise recognition test trial included the original ambiguous morphed face, along with less ambiguous morphed distractor faces (e.g., a morphed face composed of 40 % Caucasian and 60 % North African features). Although overall recognition was poor, participants were not systematically biased to select a distractor face that was consistent with the social information presented initially, as compared to one that was inconsistent with the social information. In Corneille et al.'s second experiment, they increased the number of stimuli presented at study and test, but also showed that recognition of ambiguous morphed faces was not influenced by social category labels presented at encoding. These findings might suggest that top-down, social category information does not influence face recognition.

However, opposing results were obtained by Pauker et al. (2009), who presented participants with both ambiguous and unambiguous White and Black faces, along with demographic information (including a race category label), and instructed the participants to learn the face–information pairing for a later test. Pauker et al. found that recognition of the ambiguous faces was poorer when they were paired with other-race labels at encoding than when they were paired with same-race labels. However, several issues limit the generalizability of Pauker et al.'s results. First, Pauker et al. presented both unambiguous and ambiguous faces in a single study phase, in randomized order. Rhodes et al. (2010) demonstrated that categorical perception of an ambiguous-race face is influenced by prior exposure to an unambiguous face. That is, Caucasian participants who first viewed an unambiguous Caucasian face were likely to perceive an ambiguous Asian/Caucasian face presented immediately afterward as more Asian, by contrast (and, conversely, the same ambiguous face would be perceived as being more Caucasian if it was preceded by an unambiguous Asian face). Indeed, the perceived race of ambiguous faces is more strongly influenced by the study context of unambiguous faces than by labels (Lie, 2004, cited in Rhodes et al., 2010). Although the random presentation order of the ambiguous and unambiguous faces made it unlikely that this contrastive perception effect was the driving force behind Pauker et al.'s findings, it is possible that the presence of unambiguous faces in the study phase had some uncontrolled influence on the encoding of the ambiguous faces.

Critically, Pauker et al.'s (2009) conclusion that ambiguous-race faces are recognized better when paired with an own-race label rested on an analysis that included both ambiguous and unambiguous faces. Although they did not find a significant main effect of face type (unambiguous vs. ambiguous), the reported benefit for own-race-labeled ambiguous faces was only present when performance was collapsed over the face type variable, so that performance for labeled ambiguous faces was combined with performance for unambiguous faces matching that race label. Although there was a numerical benefit for recognizing own-race-labeled ambiguous faces relative to other-race-labeled ambiguous faces (and this difference was relatively small in the White participant group), the comparison was not statistically significant. Thus, interpretation of the results of the Pauker et al. study relies on accepting several null results, and therefore their impact and generalizability are unclear.

In the present study, we presented a strong test of social–cognitive theories of the CRE by examining how recognition memory for racially ambiguous faces is influenced by the presence of social category information at either encoding or test. An additional strength of the present study is that both participant groups tested are from minority populations; the majority of studies examining the CRE have tested only participants from the majority group (e.g., Corneille et al., 2004; Hehman, Mania, & Gaertner, 2010; MacLin & Malpass, 2001; Rhodes et al., 2010; Young & Hugenberg, 2012) or have tested participants from the majority group and one minority group (e.g., Malpass & Kravitz, 1969; Pauker et al., 2009), thus confounding participant group with majority/minority status. By testing two different groups who are both minorities in their population, in the present study we examined whether a CRE would be observed in the absence of any possible majority/minority group membership confounds. In two experiments, Hispanic and African American participants studied a list of computer-generated faces designed to be racially ambiguous (i.e., 50 % African American and 50 % Hispanic; the faces had been used by MacLin & Malpass, 2001). In Experiment 1, we presented group label information prior to study, and in Experiment 2, we presented group label information only prior to recognition test trials. If the presence of race-specifying information influences the way in which ambiguous faces are initially perceived and encoded, we should observe the CRE in face recognition in Experiment 1, in which labels were presented prior to encoding and had an opportunity to affect whether ambiguous faces were encoded as own-race or other-race. However, if the CRE is driven entirely by differential perceptual expertise, we should not observe any systematic recognition differences based on the presented labels. Because the exact same ambiguous faces were randomly selected to be assigned to a label for each participant, whether a given face was perceived as own-race or other-race (or, indeed, as purely ambiguous) should be random, and there would be no reason to observe any systematic differences in recognition accuracy. Additionally, if recognition decisions (i.e., response criteria) are also affected by the presence of social category information, we might expect to see a similar pattern in Experiment 2, in which no category information was presented at study—only at test. In this case, overall recognition accuracy might be more influenced by the responses to new faces (e.g., by observing a more liberal response criterion when recognizing faces labeled as other-race; cf. MacLin & Malpass, 2001; see also Meissner & Brigham, 2001, who observed a small effect of response criterion in their meta-analysis of the CRE), given that all studied faces were encoded in the absence of any category information.

Experiment 1: Labels at study

In Experiment 1, African American and Hispanic participants studied two blocks of ambiguous-race faces for a recognition test. Prior to each block, a race category label was presented, and during the block, each studied face was also paired with a first name that was stereotypically associated with the category label. All faces were tested in a single recognition block, with no category or name information presented at test. We predicted that the social category information provided prior to study would influence how participants encoded the ambiguous faces, such that faces studied in the block with own-race labels would later be better recognized better than the faces studied in the block with other-race labels. That is, even though the exact same ambiguous face stimuli were randomly selected and assigned to the study blocks, Hispanic participants were expected to perform better with the faces studied in the block labeled as *Hispanic* than with the faces studied in the block labeled as *African American*; African American participants were expected to show the opposite pattern, mirroring the CRE typically observed in the recognition of unambiguous faces.

Method

Participants A group of 64 undergraduates from the University of Illinois at Urbana-Champaign participated in exchange for course credit or payment. In all, 32 of the participants self-identified as Hispanic, and 32 of the participants self-identified as African American on a demographics questionnaire. Participants who self-identified as both Hispanic and African American on the questionnaire were not included in the experiment. Half of the participants in each group were randomly assigned to one of the two presentation orders.

Materials Eighty computer-generated male faces were used in the experiment. The faces were selected from those used by MacLin and Malpass (2001), originally constructed using facial composite software (Faces 3.0). The faces were constructed to appear racially ambiguous, such that none of the facial features acted as racial markers. In MacLin and

Malpass's study, different hairstyles were added to the faces to serve as racial markers; the hairstyles were not included on the faces used in our experiment, to maintain the pure ambiguity of the actual face stimuli. The individual faces were randomly assigned for each participant to serve as study items (in either condition) or as distractor items at test.

At study, each face was presented along with a first name that was considered to be associated with that racial group (i.e., either Hispanic or African American). Twenty African American and 20 Hispanic first names were selected from the "African American Baby Boy Names" and "Hispanic Baby Boy Names" lists on an online baby name directory (www.babynames.org.uk). Selections from these lists were prerated by a group of 11 individuals for how strongly associated each name was with the two racial groups. Names were selected that were associated only with one of the two groups, and they were randomly selected for presentation with faces in the block with the corresponding group label.

Design The experiment had a 2 (participant group) × 2 (study label) mixed factorial design. Participant group was based on the participant's self-identified race, and study label was manipulated within participants. All participants viewed one block of Hispanic-labeled faces and one block of African-American-labeled faces; block order was counterbalanced across participants.

Procedure The participants were told that they would be shown a list of faces to try to learn for a later recognition test. They were informed that they would be shown each individual's name at the same time as the face to "help them learn the faces," but that the name would not be present at test. Participants were also asked to make predictions of how likely they thought it was that they would later recognize each face (cf. Hourihan, Benjamin, & Liu, 2012). Each block began with an instruction screen that appeared for 2,000 ms and indicated that participants would be shown a list of Hispanic/African American faces. A block of 20 study trials began immediately after this screen.

Each trial began with a male first name being presented centered at the bottom of the screen for 750 ms. The face then appeared at the center of the screen, and both the face and name were presented for 5,000 ms (the same presentation duration used by Pauker et al., 2009). After a 1,000-ms blank screen, participants were asked to predict their future recognition of the face that they had just studied, using a scale from 1 (*I am sure I will not remember this face*) to 9 (*I am sure I will remember this face*). After another 1,000-ms blank screen, the next trial began. After the first block of 20 faces had been presented, a new instruction screen appeared for 2,000 ms informing participants that they would now be shown a list of African American/Hispanic faces (whichever group label had not been presented for the first block). The next block of study trials proceeded in the same manner as the first.

Following the second block of study trials, the instructions for the recognition test appeared on the screen. Participants were informed that they would be shown one face at a time, and they were asked to press the "m" key if they believed that the face was one they had studied, and to press the "c" key if they believed that the face was one they had not studied. All 80 faces were presented on the recognition test (20 studied in the African American block, 20 studied in the Hispanic block, and 40 new faces), in random order. No names or descriptions were presented at test. The faces remained visible until participants pressed a key to respond, and then a 1,000-ms blank screen was presented prior to the next test trial. Following completion of the recognition test, participants were debriefed and thanked.

Analytic strategy As in many other studies of recognition memory, we adopted the framework of the theory of signal detection (Green & Swets, 1966; Macmillan & Creelman, 2005), which distinguishes *response criterion* (an overall tendency to judge an item as being studied or not studied, typically measured as c) from memory *discriminability* (an increased probability of responding "studied" when the item was actually studied, typically measured as d'). Traditionally, criterion and discriminability parameters are estimated by aggregating across items within each experimental condition for each participant. However, some items may be more memorable than others, and items might also differ in their baseline tendency to elicit "studied" or "not studied" responses. Such item-level variability would be lost by aggregating over items; thus, aggregation over items can lead to misestimation of the model parameters and provides no information about how effects of interest do or do not generalize across the population of items (e.g., Clark, 1973; Morey, Pratte, & Rouder, 2008). Consequently, we applied the technique of multilevel generalized linear modeling (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Wright, 1998). These models estimate the effect of experimental manipulations after simultaneously accounting for variability across both participants and items in discriminability and in response criterion. (For additional applications of multilevel generalized linear models to recognition memory, see Freeman, Heathcote, Chalmers, & Hockley, 2010; Wright, Horry, & Skagerberg, 2009.) Further details on the model and its fitting can be found in the Appendix.

Results and discussion

Recognition predictions The mean recognition predictions were numerically higher for faces studied in the own-race-labeled block ($M = 5.23$, $SD = 1.38$) than for faces studied in the other-race-labeled block ($M = 5.12$, $SD = 1.37$), replicating the results found by Hourihan et al. (2012). However, analyzing predicted performance as the dependent variable in the multilevel generalized model[1] revealed that this difference was not significantly different ($t = 1.92$, $p = .05$).

We measured the correlations between the predictions and recognition performance using the d_a measure (Benjamin & Diaz, 2008; Macmillan & Creelman, 2005; cf. Hourihan et al., 2012). An alternative way to assess these relations would have been to enter participants' prediction ratings as an additional variable into the multilevel generalized linear model of recognition memory reported below. However, such an approach would assume that participants' memory predictions were made on an interval scale—that is, that the difference between ratings of 1 and 2 must be the same as the difference between ratings of 4 and 5. As this assumption might not hold (Benjamin & Diaz, 2008), the accuracy of metamnemonic predictions is better assessed with d_a, which does not assume that predictions lie on an ordinal scale (Benjamin & Diaz, 2008). The d_a correlations between predictions and recognition performance were not significantly greater for own-race-labeled faces ($M = .165$, $SD = .608$) than for other-race-labeled faces ($M = .135$, $SD = .578$; $t(63) = 0.277$, $p = .783$); the accuracy of the recognition predictions was quite low and variable in general, unlike those observed by Hourihan et al. This was likely due to unfamiliarity with the sort of ambiguous, computer-generated faces used in the present study, as compared to the photographs of unambiguous faces used by Hourihan et al.

Recognition performance The mean proportions of "studied" responses in each condition are displayed in Fig. 1. We modeled participants' responses using the multilevel generalized linear models described above. The resulting model parameters describing the variability across the random effects (participants and test items) are displayed in Table 1, and the fixed-effect parameters describing the experimental manipulations are displayed in Table 2.

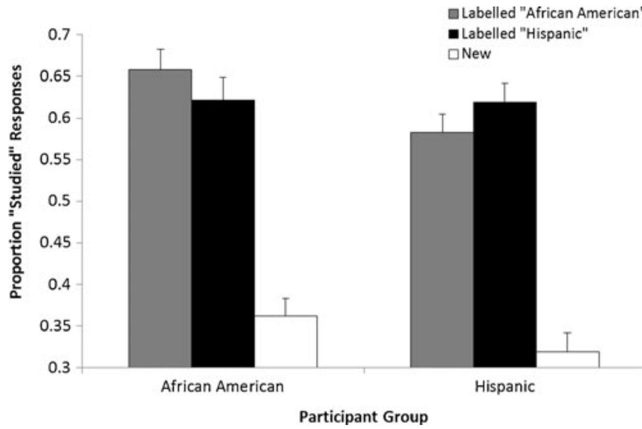

Fig. 1 Proportions of observed "studied" responses at test in Experiment 1, averaged across all items and participants in each group. Error bars represent one standard error of the mean

The random-effects portion of the model revealed substantial variability in the distribution of items, justifying an analytic technique that can capture such variability. The population of items varied just as much in discriminability as did the population of participants (estimated variance in $d' = 0.12$ for both), and items varied (estimated variance in $c = 0.11$) *more* than participants did (variance in $c = 0.04$) in response bias.

The theoretical interest, however, was in the fixed effects of participant group and category label. We first considered[2] participants' overall tendency to respond "studied"—that is, their response bias or criterion. Across all trials, participants' bias to respond "studied" or "not studied" did not differ significantly from chance. There was some evidence that African American participants were more apt to respond "studied" (i.e., to have a more liberal response criterion) than Hispanic participants, although this difference was only marginally significant.

Next, we considered participants' memory discriminability. Participants gave more "studied" responses to faces that had previously been studied than to new faces, indicating successful memory discrimination. Prior study status did not interact with participant group, indicating that overall

[1] Unlike the recognition memory decisions that were of primary interest, recognition predictions are not a binary outcome. Consequently, the probit link was not applied. In addition, because the predictions were a continuously varying variable, the model-predicted value for each observation could be accurate or inaccurate by varying degrees of error. Thus, the model of recognition predictions incorporated an additional residual term, e_{ijk}, directly analogous to that in a linear regression. In a multilevel generalized linear model with a continuous dependent variable, an inferential statistical test can be performed using the t statistic (Baayen et al., 2008). The degrees of freedom of a t statistic in a multilevel model are unclear (Baayen et al., 2008), but given a data set with thousands of observations, as in the present study, the t distribution has essentially converged to the z distribution, so the t statistic can just be treated as a z statistic (Baayen, 2008). In all other respects, the multilevel generalized linear model was the same as that applied to recognition memory.

[2] We assessed the statistical significance of the model parameters using the Wald z test, which assesses the expected change in model fit if the effect were removed from the model. An alternative means of hypothesis testing in generalized linear models is to actually fit two models, one with the effect of interest and one without, and to compare the model fits using a likelihood ratio test. Although the likelihood ratio test is often somewhat more reliable (Agresti, 2007), testing main effects by removing the effect from the model does not permit a sensible simultaneous test of both main effects and interactions, as is typically performed in an analysis of variance (ANOVA) context. We report the Wald z test for consistency with ANOVA analyses, but the tests of the critical interactions were unchanged when using the likelihood ratio test.

Table 1 Summary of random participant and item effects and correlations in a hierarchical probit model of "studied" responses in Experiment 1 (n = 5,120, log likelihood = –3,198)

Random Effect	s^2	Correlation
Participant		
Criterion (c)	0.04	–
Discriminability (d')	0.12	.32
Item		
Criterion (c)	0.11	–
Discriminability (d')	0.12	.47

memory performance did not differ reliably across the two participant groups.

Finally, we compared responses to faces studied with the *Hispanic* label versus those studied with the *African American* label. All of these faces were in truth studied and ideally should be judged as being studied; however, social–cognitive accounts of the CRE predict superior memory (more "studied" responses) for faces given an in-group label. This prediction was supported by a reliable interaction between the label at study and participant group. As can be seen in Fig. 1, for Hispanic participants, faces studied with the *Hispanic* label were remembered better than faces studied with the *African American* label; for African American participants, the reverse was true. We found no main effect of label after collapsing across the two participant groups: Faces labeled *African American* were not overall more memorable than faces labeled *Hispanic*, nor vice versa.

To summarize, we found that recognition of ambiguous-race faces was strongly influenced by social information presented at the time of encoding. Both Hispanic and African American participants were more likely to recognize ambiguous faces if they had been studied with an own-race label than if they had been studied with an other-race label. Critically, the faces were in fact exactly the same in all conditions. Our results therefore support the idea that social categorization is a driving factor in the production of the CRE; no systematic perceptual differences in the faces could have influenced encoding quality. Instead, it appears that the labels provided at encoding were sufficient in-group/out-group indicators to lead participants to encode faces in the own-race-labeled block in a manner that better supported later recognition, relative to the faces in the other-race-labeled block. This pattern of findings contradicts theoretical accounts of the CRE that rely solely on differential perceptual expertise with own- and other-race faces.

Experiment 2: Labels at test

Our first experiment showed that social category information presented at the time of encoding can strongly influence how ambiguous faces are later recognized. In the second experiment, the goal was to determine whether presenting social category information at test only would have an influence on the recognition of ambiguous-race faces initially studied in the absence of social category information: If ambiguous faces are studied in the absence of labels, will presenting labels at the time of test influence how recognition decisions are made? That is, will participants adopt different evidence criteria for responding "studied" when they are informed that faces are own-race rather than other-race? The procedure was similar to that of the first experiment, except that the social category labels and first names were not presented at study; all faces were studied in a single block. Recognition of the studied faces was tested in two blocks, and participants were only told at the beginning of the test that some of the faces that they had studied were African American and some were Hispanic, and that they would be tested separately on the two groups of faces. A category label was then presented prior to a test block, and

Table 2 Fixed-effect estimates for a hierarchical probit model of "studied" responses in Experiment 1

Fixed Effect	Estimate	*SE*	95 % CI	Wald z	p
Criterion (c)					
Overall	0.06	0.05	[–0.03, 0.16]	–1.30	.19
Participant group	–0.12	0.07	[–0.01, 0.25]	1.90	.06
Sensitivity (d')					
Overall	0.78	0.07	[0.65, 0.92]	11.36	<.001
Participant group	<0.01	0.11	[–0.21, 0.23]	<0.01	.99
Label	<0.01	0.05	[–0.10, 0.11]	0.09	.93
Participant Group × Label	0.21	0.10	[0.01, 0.42]	2.05	<.05

SE = standard error. 95 % CI = 95 % confidence interval of the parameter estimate, calculated as (1.96 × *SE*) in both directions of the point estimate. Note that, in Experiment 1, the labels could not influence criterion placement independent of study status, because the labels were manipulated only for the studied items.

first names were also presented along with each face. If social category information mainly influences how participants initially encode ambiguous faces, then presenting labels only at test should not result in differential recognition performance between the two test blocks. However, if recognition response criteria are also influenced by the presence of social category information, then response criteria might be more conservative for the block of faces tested with an own-race label (cf. MacLin & Malpass, 2001; Young, Bernstein, & Hugenberg, 2010).

Method

Participants A group of 66 individuals (33 African American and 33 Hispanic) were recruited in the same manner as in Experiment 1. In each group, 16 of the participants were randomly assigned to one of the two test orders, and 17 were assigned to the other test order. An additional three Hispanic and five African American participants completed the experiment but reported having participated in a similar experiment with computer-generated faces, and were therefore replaced (because they likely were aware a priori that the faces were intended to be racially ambiguous).

Materials The same 80 computer-generated, racially ambiguous faces used in Experiment 1 were used in Experiment 2. An additional 40 male first names (20 Hispanic and 20 African American) were obtained from the same source as in Experiment 1, in order to have sufficient names to present on all 80 test trials.

Design The experiment had a 2 (participant group) × 2 (test label) mixed factorial design. Participant group was determined by a participant's self-reported race, and test label was manipulated within participants. All participants viewed one test block of Hispanic-labeled faces and one test block of African-American-labeled faces; block order was counterbalanced across participants.

Procedure The procedure was similar to that of Experiment 1, except that in the study phase, all 40 faces were presented in a single study block, without any first names or group labels being presented. Participants did not provide recognition predictions in this experiment; because the faces were studied without names or labels, we did not expect any systematic variation in the prediction responses. At the beginning of the recognition test, participants were informed that some of the faces that they had studied were African American and some were Hispanic, and that they would be tested on the two groups of faces separately. Each test block began with an instruction screen that informed participants that they would now be tested on the Hispanic/African American faces that they had studied. The instructions remained visible until the participant pressed the space bar to begin the test trials. In each test block, a random 20 of the studied faces were presented along with 20 new faces, in random order. On each test trial, a face was presented at the center of the screen, along with a first name presented below the face. Following completion of the recognition test, participants were debriefed and thanked.

Results and discussion

The proportions of "studied" responses are displayed in Fig. 2 as a function of participant group, test label, and studied status (old or new). As in Experiment 1, we modeled participants' rates of "studied" responses using the multilevel generalized linear model described previously. The variability across the random effects of participants and items is displayed in Table 3, and the fixed-effect parameter estimates from the model are displayed in Table 4.

As in Experiment 1, the random-effects portion of the model revealed substantial variability across the population of items from which we sampled. In Experiment 2, items were just as variable as participants in discriminability (variance in d' = 0.09 for both), and more variable in criterion placement (item variance in c = 0.16, participant variance = 0.10).

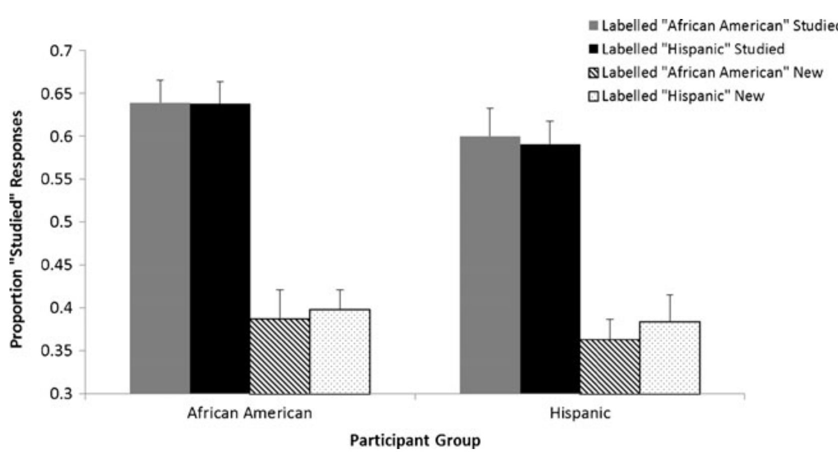

Fig. 2 Proportions of observed "studied" responses at test in Experiment 2, averaged across all items and participants in each group. Error bars represent one standard error of the mean

Table 3 Summary of random participant and item effects and correlations in a hierarchical probit model of "studied" responses in Experiment 2 (n = 5,280, log likelihood = –3,280)

Random Effect	s^2	Correlation
Participant		
Criterion (c)	0.10	–
Discriminability (d')	0.09	.20
Item		
Criterion (c)	0.16	–
Discriminability (d')	0.09	.65

We again began our analysis of the fixed effects by considering participants' preferences to respond "studied" versus "unstudied." As in Experiment 1, this preference did not differ significantly from chance, nor did it reliably differ between the two participant groups. Because the two labels were tested separately in Experiment 2, it was also possible that participants' criteria could have differed between the block labeled *Hispanic* and the block labeled *African American*. However, we found no such effect. Finally, and critically, no reliable Participant Group × Label interaction emerged, indicating that the amount of evidence required to call a face "studied" was not influenced by whether the labels at test were same-race or other-race.

We next turned to participants' ability to discriminate the previously studied faces from the new, unstudied faces. Overall, participants responded "studied" more to previously studied faces, indicating successful discrimination in memory. Memory sensitivity did not differ reliably between the two participant groups, nor did sensitivity differ as a function of whether a block was labeled *African American* or *Hispanic*.

In Experiment 1, in which the labels were presented during study, label interacted with participant group to determine how likely the face was to later be recognized. No such interaction was observed when the labels were presented at test in Experiment 2; in fact, the effect was numerically in the opposite direction.

To summarize, we found no evidence that recognition of ambiguous race faces is influenced by the presentation of social category labels at the time of test. Neither sensitivity nor response criterion was systematically influenced by the presence of the social category labels at test.

General discussion

In the present study, we demonstrated that the same ambiguous race faces can be encoded and remembered very differently, depending on the social category label presented at the time of encoding: When an ambiguous face was labeled as own-race at the time of encoding, it was later recognized better than when it was labeled as other-race. We observed this effect even though the same face stimuli were sampled for the own-race and other-race conditions in two different participant populations. This strongly supports the idea that social category information can influence face processing in a top-down manner and result in better recognition of ambiguous faces encoded as own-race, relative to faces encoded as other-race. This is the first unequivocal demonstration of such an effect with completely perceptually ambiguous faces. Our second experiment failed to show any influence of social category information presented at test only, consistent with the idea that the initial encoding of faces is what drives subsequent recognition, rather than differential response bias at test.

Our results would be difficult to account for by any theory of the CRE that relied principally on differential perceptual expertise with own-race and other-race faces (e.g., Rhodes et al., 2010; Valentine, 1991; Valentine et al., 1995). Because the same set of ambiguous-race faces was sampled in the two labeled study blocks, no systematic perceptual differences could have led to differential recognition performance. The only factor that differed systematically between the two study blocks was the social category label; this produced better memory for the faces labeled as

Table 4 Fixed-effect estimates for a hierarchical probit model of "studied" responses in Experiment 2

Fixed Effect	Estimate	SE	95 % CI	Wald z	p
Criterion (c)					
Overall	0.01	0.06	[–0.12, 0.13]	–0.07	.95
Participant group	–0.09	0.09	[–0.26, 0.08]	–1.07	.29
Label	–0.01	0.04	[–0.08, 0.06]	–0.34	.74
Participant Group × Label	–0.02	0.07	[–0.11, 0.18]	0.30	.76
Sensitivity (d')					
Overall	0.66	0.06	[0.54, 0.78]	10.68	<.001
Participant group	0.07	0.10	[–0.13, 0.28]	0.72	.47
Label	0.01	0.07	[–0.13, 0.16]	0.20	.84
Participant Group × Label	–0.07	0.15	[–0.36, 0.22]	–0.50	.62

SE = standard error. 95 % CI = 95 % confidence interval of the parameter estimate, calculated as (1.96 × SE) in both directions of the point estimate.

own-race for both groups of participants. Interestingly, this observation differs from the results of Rhodes et al. (2010), who found a CRE in face perception but not in recognition, using perceptually ambiguous faces. However, they presented faces for only 500 ms at study (as compared to our 5,000-ms presentation time); this short presentation time may not have been sufficient for top-down differences in encoding strategies to have an influence on later recognition memory.

Our results are more easily explained by social–cognitive accounts of the CRE (e.g., Sporer, 2001): In the absence of any perceptual cues to categorize ambiguous faces, the presented category labels led to superior encoding of faces labeled as own-race, relative to faces labeled as other-race. The fact that we observed this difference in recognition using the exact same faces across label conditions strengthens our findings.

Our findings augment the already substantial literature supporting social–cognitive accounts of the CRE, in which the categorization of faces as in-group or out-group members influences the quality of subsequent encoding and recognition (e.g., Sporer, 2001). Researchers have produced recognition effects similar to the CRE with own-race faces by pairing faces with social in-group versus social out-group contexts, such as own-university versus rival-university affiliation, or wealth versus poverty for relatively affluent participants (Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008). Furthermore, Hehman et al. (2010) did not observe a CRE when own- and other-race faces were grouped at study as belonging to participants' own university (i.e., their social in-group). Those studies demonstrate how perceptually identical (and same-race) faces can be encoded and recognized very differently on the basis of social category information that is entirely separate from face features.

Similarly, our results showed that contextual information (i.e., category labels) can influence the recognition of perceptually identical (ambiguous-race) faces. Moreover, several researchers have demonstrated that explicit instructions to pay careful attention to differentiating details of other-race faces at the time of encoding (Hugenberg, Miller, & Claypool, 2007; Tullis, Benjamin, & Liu, 2013) or to consider ambiguous faces as members of the social in-group (Pauker et al., 2009) can reduce or even eliminate the CRE. Coinciding with our observation that category labels influenced memory for ambiguous faces only when the labels were presented prior to study, and not just prior to test, Young et al. (2010) found that explicit instructions to differentiate other-race faces were only effective in reducing the CRE when the instructions were provided prior to encoding, and not only prior to test. As with our present results, these sorts of findings are difficult to explain by perceptual-expertise accounts that do not allow for some top-down control of encoding strategy.

Additionally, our results are consistent with the categorization–individuation model of the CRE (e.g., Hugenberg et al., 2010; Young & Hugenberg, 2012). This model accounts for the contributions and interaction of experience, motivation, and top-down social categorization in the observation (and magnitude) of the CRE. Given that the perceptual features of our stimuli were equivalent in all conditions, only the presented social category information could have influenced the differential encoding of faces. The categorization–individuation model allows for the inclusion of such situational factors exerting top-down influences on the tendency to categorize faces, resulting in the perception of greater within-category similarity for out-group faces. Our results present a new context in which purely verbal information can be shown to influence face categorization at encoding.

In conclusion, we have demonstrated that the encoding of ambiguous-race faces can be influenced by nonperceptual information: the presentation of social category labels. In the first known study to include two different minority group member, both Hispanic and African American participants recognized ambiguous faces better when the faces were studied with an own-race category label than when they were studied with an other-race category label, even when the same faces were randomized across conditions. We did not find evidence for a similar influence of category labels presented at test only; recognition decisions were only affected when the information was presented at face encoding. Our results are inconsistent with perceptual-expertise accounts of the CRE (e.g., Rhodes et al., 2010; Valentine, 1991), as applied to ambiguous faces, and are more supportive of theories that allow for the influence of top-down social categorization (e.g., Sporer, 2001; Young & Hugenberg, 2012) on face recognition.

Author Note This research was supported by Grant No. R01AG026263 to A.S.B. from the National Institutes of Health and by National Science Foundation Graduate Research Fellowship No. 2007053221 and a traineeship from National Institutes of Health Language Processing Training Program No. 5T32HD055272-13 to S.H.F. We are grateful to Scott Gronlund for his input on early stages of this work, and to members of the Human Memory and Cognition lab for their assistance with data collection.

Appendix

Participants' recognition memory judgments were modeled using multilevel generalized linear models. In these models, the unit of analysis is the response made on an individual trial (i.e., "studied" or "not studied"), rather than a proportion aggregated across items. To account for the fact that an individual recognition memory decision is a binary outcome (items can only be judged as being studied or not studied) rather than continuously varying, the dependent variable in the model was the cumulative distribution function of a "studied" response, also known as the *probit*[3] (Bliss, 1934). A probit model returns parameter estimates on the same scale as the d' measure used in detection-theoretic models of memory (DeCarlo, 1998), but with the advantage that the multilevel probit model can simultaneously incorporate participant and item variability. This model is displayed in Eq. A1.

$$\mathrm{probit}(\mathrm{Y_{ij}}) = \gamma_{00} + \gamma_{100}\mathrm{X_1} + \mathrm{u_{i0}} + \mathrm{u_{0j}} + \mathrm{u_{1i0}}\mathrm{X_1} + \mathrm{u_{10j}}\mathrm{X_1} + \gamma_{200}\mathrm{X_2} + \gamma_{300}\mathrm{X_3} + \ldots \quad (A1)$$

In this model, response criterion is captured by an intercept parameter γ_{00} that corresponds to the baseline rate of "studied" responses across all conditions; a positive parameter indicates a tendency to respond "studied" with greater than equal probability. This is the additive inverse of the more traditional criterion measure c (Macmillan & Creelman, 2005), for which a positive c indicates a conservative criterion and fewer "studied" responses; we report the response bias parameter transformed to c for consistency with the literature. Discrimination, or d', is captured by a parameter γ_{100} that describes the increase in the rate of "studied" responses when the item was actually studied ($X_1 = 0.5$), as compared to when it was not studied ($X_1 = -0.5$).

Participant and item variability were modeled through the inclusion of *random effects*, variables whose levels were sampled from a larger population. u_{i0} represents the baseline tendency for participant i to respond "studied" (parameterized as a deviation from the overall mean), while u_{0j} represents the baseline tendency for item j to elicit "studied" responses. Similarly, u_{1i0} and u_{10j} represent the discriminability of participant i and item j, respectively. Because the individual levels of a random effect are assumed not to be of interest (i.e., the goal of the experiment is not to report on the characteristics of Participant 3), random effects are fit using parameters that correspond not to particular sampled items or participants, but to the variance in the sampled population. Three parameters represented variance between participants in criterion placement, variance between participants in discrimination, and the correlation between participants' criterion placement and discrimination (e.g., $\sigma^2_{u_{i0}}$ the variance of the distribution of criteria across participants). Another three parameters represented variance between items in the same.

Each model also included *fixed effects* of interest that tested the influences of the experimentally manipulated variables on c and d'. For example, the fixed effects in Experiment 1 were the categorical variables of participant group (X_2), label type (*African American* or *Hispanic*; X_3), and their interaction (X_2X_3). These fixed effects are incorporated into the model as the effects γ_{200}, γ_{300}, and γ_{400} of categorical variables X_2, X_3, and X_2X_3. Several different coding schemes are available for categorical variables in multilevel generalized linear models (and other regression models); we centered each predictor variable at its mean, as this provides parameters that correspond to tests of the main effects and interactions of an ANOVA analysis.

All of the models reported were fit in the R software environment with Laplace estimation, using the lmer() function of the lme4 package (Bates, Maechler, & Bolker, 2011).

[3] An alternative dependent variable in binomial multilevel models is the *logit*, or log odds. The logit is highly correlated with the probit (Agresti, 2007), and we use the probit for consistency with detection-theoretic approaches to recognition memory.

References

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/jmla.2007.12.005

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Software] (R package version 0.99375-39). Retrieved from http://CRAN.R-project.org/package-lme4

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York, NY: Psychology Press.

Bliss, C. I. (1934). The method of probits. *Science, 79,* 38–39. doi:10.1126/science.79.2037.38

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359. doi:10.1016/S0022-5371(73)80014-3

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts toward more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology, 86,* 236–250. doi:10.1037/0022-3514.86.2.236

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186–206.

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62,* 1–18. doi:10.1016/j.jml.2009.09.004

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology, 46,* 445–448. doi:10.1016/j.jesp.2009.11.008

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition, 1,* 158–162. doi:10.1016/j.jarmac.2012.06.004

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a

solution to an insidious problem. *Journal of Experimental Social Psychology, 43,* 334–340. doi:10.1016/j.jesp.2006.02.010
Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization–individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117,* 1168–1187. doi:10.1037/a0020463
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007
Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1364–1382. doi:10.1037/0278-7393.22.6.1364
Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129,* 559–574. doi:10.1037/0096-3445.129.4.559
MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7,* 98–118. doi:10.1037/1076-8971.7.1.98
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330–334.
Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3–35. doi:10.1037/1076-8971.7.1.3
Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in *z*ROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology, 52,* 376–388. doi:10.1016/j.jmp.2008.02.001
Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96,* 798–810. doi:10.1037/a0013265
Rhodes, G., Lie, H. C., Ewing, E., Evangelista, E., & Tanaka, J. W. (2010). Does perceived race affect discrimination and recognition of ambiguous-race faces? A test of the sociocognitive hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 217–223. doi:10.1037/a0017680
Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138,* 146–174. doi:10.1037/a0025750
Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality and Social Psychology Bulletin, 34,* 260–274. doi:10.1177/0146167207310455
Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in whites. *Basic and Applied Social Psychology, 22,* 71–84.
Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7,* 36–97. doi:10.1037/1076-8971.7.1.36
Tullis, J. G., Benjamin, A. S., & Liu, X. (2013). *Self-pacing faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect*. Manuscript submitted for publication.
Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161–204.
Valentine, T., Chiroro, P., & Dixon, R. (1995). An account of the own-race bias and the contact hypothesis based on a "face space" model of face recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 69–91). New York, NY: Routledge.
Wright, D. B. (1998). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology, 12,* 339–357. doi:10.1002/(SICI)1099-0720(199808)12:4<339::AID-ACP571>3.0.CO;2-D
Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41,* 257–267. doi:10.3758/BRM.41.2.257
Young, S. G., Bernstein, M. J., & Hugenberg, K. (2010). When do own-group biases in face recognition occur? Encoding versus post-encoding. *Social Cognition, 28,* 240–250. doi:10.1521/soco.2010.28.2.240
Young, S. G., & Hugenberg, K. (2012). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science, 3,* 80–87. doi:10.1177/1948550611409759
Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review, 16,* 116–142. doi:10.1177/1088868311418987