

I know your face but not where I saw you: Context memory is impaired for other-race faces

RUTH HORRY

University of Sussex, Brighton, England

AND

DANIEL B. WRIGHT

Florida International University, Miami, Florida

People are more likely to falsely identify a face of another race than a face of their own race. When witnesses make identifications, they often need to remember where they have previously encountered a face. Failure to remember the context of an encounter can result in unconscious transference and lead to misidentifications. Forty-five White participants were shown White and Black faces, each presented on one of five backgrounds. The participants had to identify these faces in an old/new recognition test. If participants stated that they had seen a face, they had to identify the context in which the face had originally appeared. Participants made more context errors with Black faces than with White faces. This shows that the own-race bias extends to context memory.

Cross-race face recognition has been studied systematically for almost four decades, since Malpass and Kravitz (1969) found that White and Black university students showed superior recognition accuracy for faces of their own race in comparison with faces of the other race. This effect, termed the *own-race bias* (ORB), has been replicated across several races and experimental paradigms (Ferguson, Rhodes, Lee, & Sriram, 2001; Wright, Boyd, & Tredoux, 2003). For reviews, see Bothwell, Brigham, and Malpass (1989), Meissner and Brigham (2001), and Sporer (2001).

When making identification decisions, participants can be correct in one of two ways. If participants say that they have seen a target before, this is a hit. If participants say that they have not seen a foil before, this is a correct rejection. Similarly, identification decisions can be incorrect in one of two ways. If participants decide that they have not seen a target, this is a miss. Finally, if participants decide that they have seen a foil before, this is a false alarm.

Research has shown that the ORB can be caused by both increased hit rates for own-race faces compared with other-race faces and with inflated false alarm rates for other-race faces compared with own-race faces (Meissner & Brigham, 2001). However, some studies have shown that the ORB is more pronounced for false alarm rates than for hit rates (Slone, Brigham, & Meissner, 2000). In other words, participants are more likely to wrongly identify a foil of another race as having been seen before than a foil of their own race. The real-world implications of this in terms of eyewitness memory are alarming, because a false alarm within the legal system can ultimately lead

to the wrongful conviction of an innocent person (Doyle, 2001).

When making an identification, witnesses must not only remember whether they have ever seen a specific individual, but also in what circumstance they encountered that individual. For example, did they see the person at the crime scene or the day before in some unrelated context? Or perhaps the person was an innocent bystander to the crime. The cognitive process involved in making such decisions is known as *source monitoring*, and failures in this process can lead to unconscious transference (see Lindsay, 2007, for a review). This is when a witness mistakes a familiar person from another time or place for the culprit (Read, Tollestrup, Hammersley, McFazden, & Christensen, 1990).

Source memory has been shown to be much less accurate than face recognition accuracy (Brown, Deffenbacher, & Sturgill, 1977). For example, participants may be unable to accurately remember whether a target individual was an assailant or a bystander at a crime (Read et al., 1990; Ross, Ceci, Dunning, & Togli, 1994), or may be unable to recall whether an individual was seen at a crime scene or later, in a photo array (Deffenbacher, Bornstein, & Penrod, 2006; Dysart, Lindsay, Hammond, & Dupuis, 2001).

The Innocence Project (www.innocenceproject.org) includes real-world cases where unconscious transference has led to the incarceration of innocent men. For example, in 1989, African-American Larry Bostic was convicted of sexual battery and assault in Florida. The victim identified Bostic as her attacker, but after he was exonerated through DNA evidence, she recanted her identification, claiming

R. Horry, r.horry@sussex.ac.uk

that she may have seen him in the neighborhood in the days before the attack.

Here, we examine how accurately White participants can identify White and Black targets and identify the contexts in which the targets appear. The targets were superimposed onto one of five contexts at encoding, and participants performed an old/new recognition test for the targets and a forced choice test for the contexts for their positive identifications. It was predicted that participants' recognition for White faces would be superior to recognition for Black faces, particularly for false alarm rates. It was also speculated that the ORB would extend to context memory. In other words, participants would be poorer at remembering on which background a Black face had appeared than a White face.

Although much is known about cross-race face recognition, little is known about the impact that race can have on other aspects of eyewitness memory. It is therefore important to explore any possible links between the own-race bias and memory phenomena such as unconscious transference if we are to ensure that minority-race suspects are not disadvantaged within the legal system.

METHOD

Participants and Design

Fifty-eight participants were selected from the psychology subject pool to take part in the experiment in return for £4 (~\$8 U.S.). Forty-six participants identified their ethnic origin as White. Non-White participants were excluded from the analysis.¹ Seventy-four percent of the included participants were female. Ages ranged from 15 to 56, with a mean age of 27. Data from 1 participant were lost due to a computer error.

A 2×5 within-subjects design was used. The first factor was race of face, the two levels of which were White and Black. The second factor was context, the five levels of which were a basketball court, a prison cell, a classroom, a factory, and a gray screen. Both of these factors varied within participants.

The dependent variables were recognition accuracy, confidence, and context memory.

Materials

One hundred sixty faces, 80 of each race, were obtained from Internet image searches. The faces were looking straight ahead with neutral expressions and had no distinguishing marks, glasses, or jewelry. The faces were cropped in Adobe Photoshop to remove necks, shoulders, and backgrounds.

The same photos were used at encoding and testing. This can present the possibility of low-level image recognition occurring, rather than higher level face recognition. However, the central aim in this research was to investigate context memory for own- and other-race faces, rather than the ORB for face recognition itself.

The faces were then superimposed on the five different context images. The size and position of the face in relation to the background were made constant across all stimuli. For this study, the term *context* refers to the background on which the face appeared at encoding, although it is acknowledged that the term *context* can include a host of other information, such as the time of encoding.

Examples of the stimuli are shown in Figure 1.

Procedure

Participants were tested in groups of up to 3 by a White experimenter. Each participant sat at an individual PC, separated from the other participants by large dividing screens. The experimenter

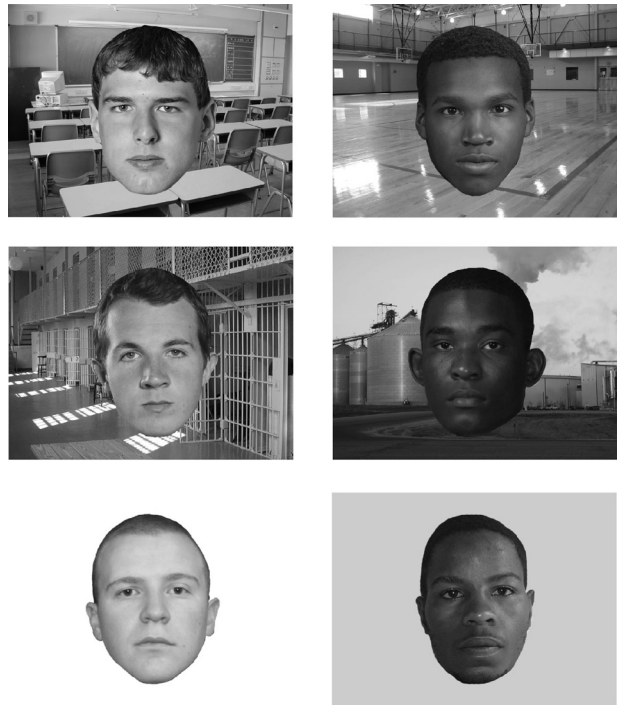


Figure 1. Examples of faces on all five encoding backgrounds and on a blank screen (for testing). Backgrounds from left to right, top to bottom, are classroom, basketball court, prison, factory, blank, and gray screen.

remained in the room throughout the experiment to ensure that there was no talking among participants. The experiment was created and run using E-Prime 1.1 for Windows.

Participants were informed that they would be performing a face recognition task and that they would need to pay close attention to both the faces and the backgrounds on which they were presented. Fifty faces of each race appeared as targets, with the remaining 30 presented as foils. Therefore, in the encoding phase, participants saw 10 faces of each race in each different context. Each face was presented for 5 sec, with a 1-sec interstimulus interval.

Testing began immediately after the encoding phase. The target faces, along with the foils, were presented sequentially on a blank white screen. First, the participants had to decide whether they recognized the face, by pressing "1" on the keyboard for *yes*, or "0" for *no*. They were then asked how confident they were on a 9-point scale (1 being *not confident at all* and 9 being *very confident*). If they had responded that they recognized the face, they had to decide in which context the face had appeared. This was a forced choice test, with no option to respond "don't know." The face remained on-screen throughout these questions. Participants were asked to be as accurate as possible, and to take as much time as they needed over their answers.

The experiment was split into four blocks, so that participants had the opportunity to rest and so that they did not have to remember too many faces at once. In each encoding block, 25 stimuli were presented. These stimuli consisted of faces of both races and all contexts and were presented in a random order. In each testing phase, the 25 faces were shown again, along with 15 foils. So, in total, participants viewed 100 targets and 60 foils.

Presentation of faces as either targets or foils was counterbalanced between participants. The contexts in which the faces appeared were also fully counterbalanced.

After the experiment, participants were thanked, debriefed, paid, and asked not to discuss the experiment with anyone.

RESULTS

Traditionally, with recognition memory data, signal detection theory measures such as d' are used to account for the sensitivity of responses—that is, how well a participant can distinguish a new stimulus from an old stimulus—on the basis of hit rates and false alarm rates. d' values were calculated for the White and Black faces across all five encoding contexts. False alarm rates could not be calculated separately for each context, because the foils did not appear in any encoding context. The overall false alarm rates for each race were used to calculate the d' values for each encoding context. C scores were also calculated, to check for any difference in response bias.

The d' values were then entered into a 2 (race of face) \times 5 (context) within-subjects ANOVA. Mean d' scores for White faces (1.94) were higher than those for Black faces (1.64) [$F(1,44) = 8.53, p = .005, \eta_p^2 = .16$]. We did not observe significant differences across contexts [$F(4,176) = 0.43, p = .78, \eta_p^2 = .01$], and context did not appear to interact with race [$F(4,176) = 1.18, p = .32, \eta_p^2 = .03$].

A second ANOVA was conducted on the C scores. The mean C for White faces was $-.19$, compared with a C of $-.004$ for Black faces, suggesting that participants were responding more conservatively to White faces than to Black ones. This was supported by the significant effect of race [$F(1,44) = 9.00, p = .004, \eta_p^2 = .17$]. There were no

observable differences across contexts [$F(4,176) = 0.43, p = .79, \eta_p^2 = .01$], and the interaction did not approach significance [$F(4,176) = 1.18, p = .32, \eta_p^2 = .03$].

Because the ORB is often more pronounced for false alarms than for hit rates, we also analyzed hit rates and false alarm rates separately. Figure 2 shows the hit rate for the five different contexts and the false alarms for the new faces, broken down by whether the faces are White or Black. The 95% between-subjects confidence intervals are also shown. The hit rates are approximately 80% for all contexts and for both races, whereas the false alarm rates are substantially lower for White faces (12.67%) than for Black ones (22.30%). t tests were run on the hits for each context and for false alarms between White and Black faces, and the 95% confidence intervals for these differences are shown in the inlaid graph. There were no significant differences for hits [maximum $t(44) = 1.45, p = .08$], but there was a substantial difference for false alarms [$t(44) = 5.60, p < .001, d = 0.84$]. To increase power for the analysis of hits, a 2×5 within-subjects ANOVA was conducted on the hit rates. As expected from viewing Figure 2, none of the effects were significant: main effect for race of face [$F(1,44) = 0.77, p = .39, \eta_p^2 = .02$]; main effect of context [$F(4,176) = 0.22, p = .93, \eta_p^2 = .01$]; and interaction [$F(4,176) = 0.54, p = .70, \eta_p^2 = .01$]. Thus, the own-race bias observed in the d' analysis was due almost entirely to inflated false alarm rates for Black faces.

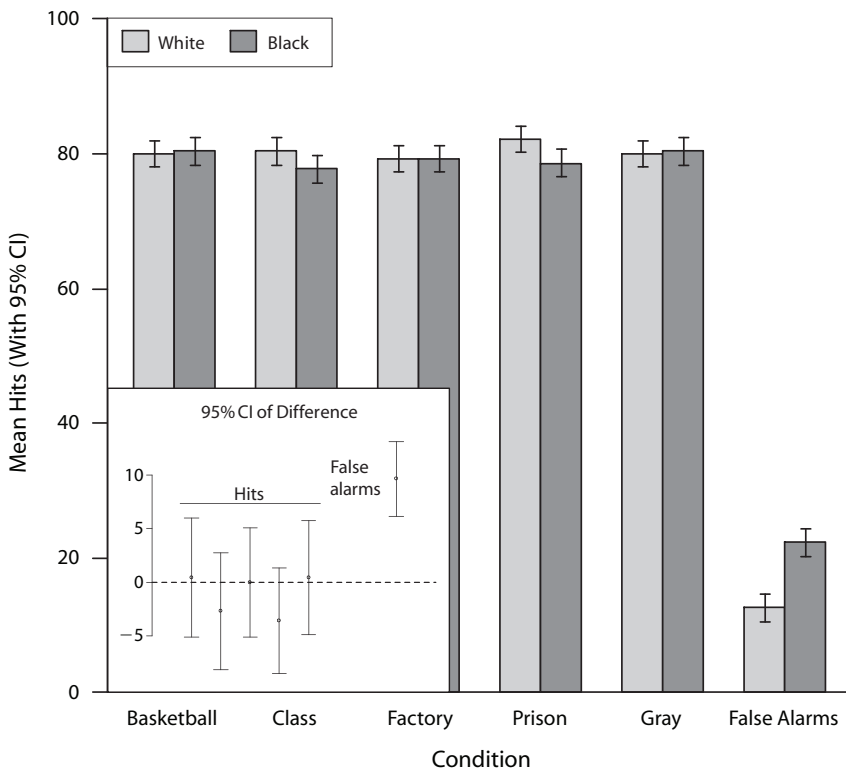


Figure 2. The hit rates for the five contexts for White and Black faces, and the false alarm rates for White and Black faces. The 95% confidence intervals are shown. The inlaid graph shows the confidence intervals between recognition for White and Black faces in each context. CI, confidence interval.

We felt that the nonsignificant differences in hit rates for Black and White faces could be hiding something more subtle. Participants were also asked to rate their confidence on a 1–9 rating scale. To distinguish between “yes” and “no” responses, confidence ratings for “no” responses were assigned negative values. This meant that confidence ratings could vary from -9 to $+9$, depending on the response type. So, for targets, positive confidence values reflect correct responses, and for foils, negative confidence values reflect correct responses.

Mean confidence ratings were obtained for each participant for targets and foils separately. The mean confidence rating for White targets (5.10) was higher than for Black targets (4.74), a statistically significant effect [$t(44) = 2.20, p = .03, \eta^2 = .05$]. The mean confidence rating for White foils (-5.76) was lower than for Black foils (-3.73), which was, again, a statistically significant effect [$t(44) = 6.65, p < .001, \eta^2 = .13$]. The effect for foils was larger than the effect for targets, supporting the findings of the accuracy analysis. However, the significant effect for targets suggests that participants were more confident in their correct responses to White targets.

Next, we examined whether memory for context differed for Black and White faces. Participants were asked only to identify the context if they said “old,” and therefore context judgments were made only for hit and false alarm trials. Because all context judgments for false alarm trials are in error, we analyze only the context judgments for hits. On 1 of 3,588 hits, the person did not record a context ($<.03\%$), so this trial is excluded.

Table 1 shows the percentage of context judgments for hits of Black and White faces originally shown in each context. The most important numbers in Table 1 are those along the diagonal in bold. These values represent the percentage of correct context responses for Black and White faces on each encoding background. For each of these, the percentage of correct context recognitions is higher for White faces. Overall, 52% of context judgments for Black hit trials were accurate, whereas 59% were accurate for White hit trials. Because the number of hits varies by person, we used multilevel modeling for statistical analysis,

so that inappropriate weight was not given to people with different numbers of hits (Wright & London, 2008). Because the response is binary—getting the context right or wrong—we used a multilevel logistic regression (Wright, 1997). The effect of race was statistically significant [$\chi^2(1) = 16.73, p < .001$], with the regression parameter estimated at $.29$ ($SE = .07$). There were no observable effects of the different contexts.

DISCUSSION

In line with previous research, participants were almost twice as likely to misidentify a Black foil than a White foil. This supports the findings of Meissner and Brigham (2001) that the ORB is due largely to higher false alarm rate for other-race faces than for own-race faces. Although participants were shown 160 faces, which does not mimic the encoding and testing circumstances of the typical crime, the results are consistent with a meta-analysis of other laboratory studies (Meissner & Brigham, 2001) and a field study with only a single target (Wright, Boyd, & Tredoux, 2001). The real-world implications are disconcerting, because eyewitnesses are more likely to misidentify an innocent suspect of another race than one of their own. This suggests that police may wish to use further precautions when dealing with cross-race identifications, such as using more foils in lineups (Wells & Olson, 2001).

That participants were no more accurate at identifying White targets than Black targets was reflected in the similar hit rates for faces of both races. Although several authors have found a larger ORB in false alarm rates than in hit rates, some authors have observed the ORB for hit rates. We used confidence measures, and these did show that participants were better with White faces.

The main aim of this study was to see how well participants could remember in which context they had seen faces of Black and White people. In support of the findings of Brown et al. (1977), memory for context was moderate. Participants were able to remember the context for 55% of the faces they correctly identified. This is still well above chance, since guessing would result in 20% accuracy.

Table 1
Percentage of Responses for Each Context of Hits
for White Faces and Black Faces

Response Context	Race of Face	Original Context				
		Basketball	Class	Factory	Prison	Gray
Basketball	White	57	7	11	10	8
	Black	49	8	12	8	12
Class	White	11	65	9	8	13
	Black	14	62	15	15	13
Factory	White	11	8	54	8	13
	Black	15	10	48	12	16
Prison	White	11	11	12	61	9
	Black	11	10	8	54	11
Gray	White	11	9	14	13	57
	Black	10	10	16	10	47
<i>n</i> for White faces		360	362	357	368	360
<i>n</i> for Black faces		361	350	353	354	362

Note—The diagonal (in bold) shows accurate context memories. Some column percentages do not sum to 100 because of rounding.

Participants could remember the context for White targets better than they could for Black targets. This is an important finding. It is the first demonstration that White people are better at remembering the context in which they saw White faces. This suggests that people are less likely to remember the circumstances in which they encountered an individual of a different race. Research into bystander misidentification (Read et al., 1990; Ross et al., 1994) and mugshot exposure (Deffenbacher et al., 2006; Dysart et al., 2001) has shown that people can and do make mistakes concerning the context in which a face has been encountered. This study shows that these transference errors may be more likely in cross-race identifications.

To date (April 7th, 2008), the Innocence Project has reported information about 215 people who have been falsely convicted and later exonerated from the American prison system. Most of these convictions were based on flawed eyewitness testimony. The problem of inaccurate eyewitness testimony has two sides: Not only can an innocent person be sent to prison, but the true culprit will remain free. Understanding the factors that can influence the accuracy of eyewitness memory is therefore of prime importance in ensuring that the legal system is as fair as possible.

AUTHOR NOTE

This research was funded by the Economic and Social Research Council. Correspondence concerning this article should be addressed to R. Horry, Department of Psychology, University of Sussex, Brighton BN1 9RH, England (e-mail: r.horry@sussex.ac.uk).

REFERENCES

- BOTHWELL, R. K., BRIGHAM, J. C., & MALPASS, R. S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin*, *15*, 19-25.
- BROWN, E., DEFFENBACHER, K., & STURGILL, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*, 311-318.
- DEFFENBACHER, K. A., BORNSTEIN, B. H., & PENROD, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law & Human Behavior*, *30*, 287-307.
- DOYLE, J. M. (2001). Discounting the error costs: Cross-racial false alarms in the culture of contemporary criminal justice. *Psychology, Public Policy, & Law*, *7*, 253-262.
- DYSART, J. E., LINDSAY, R. C. L., HAMMOND, R., & DUPUIS, P. (2001).

- Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. *Journal of Applied Psychology*, *86*, 1280-1284.
- FERGUSON, D. P., RHODES, G., LEE, K., & SRIRAM, N. (2001). "They all look alike to me": Prejudice and cross-race face recognition. *British Journal of Psychology*, *92*, 567-577.
- THE INNOCENCE PROJECT (www.innocenceproject.org). Accessed April 7, 2008.
- LINDSAY, D. S. (2007). Autobiographical memory, eyewitness reports, and public policy. *Canadian Psychology*, *48*, 57-66.
- MALPASS, R. S., & KRAVITZ, J. (1969). Recognition for faces of own and other race. *Journal of Personality & 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, & Law*, *7*, 3-35.
- READ, J. D., TOLLESTRUP, P., HAMMERSLEY, R., MCFAZDEN, E., & CHRISTENSEN, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology*, *4*, 3-31.
- ROSS, D. F., CECI, S. J., DUNNING, D., & TOGLIA, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar but innocent person. *Journal of Applied Psychology*, *79*, 918-930.
- SLONE, A. E., BRIGHAM, J. C., & MEISSNER, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic & Applied Social Psychology*, *22*, 71-84.
- SPORER, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, & Law*, *7*, 36-97.
- WELLS, G. L., & OLSON, E. A. (2001). The other-race effect in eyewitness identification: What do we do about it? *Psychology, Public Policy, & Law*, *7*, 230-246.
- WRIGHT, D. B. (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical & Statistical Psychology*, *50*, 21-29.
- WRIGHT, D. B., BOYD, C. E., & TREDoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, & Law*, *7*, 119-133.
- WRIGHT, D. B., BOYD, C. E., & TREDoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology*, *17*, 365-373.
- WRIGHT, D. B., & LONDON, K. (2008). *Modern regression techniques using R: A practical guide for students and researchers*. London: Sage.

NOTE

1. A comparable sample of Black participants could not be tested, due to the demographics of the university and the surrounding area. Approximately 3% of psychology undergraduates at Sussex University are Black, and approximately 0.3% of the population of Sussex identify themselves as Black.

(Manuscript received October 12, 2007;
revision accepted for publication December 30, 2007.)