



Changes in the own group bias across immediate and delayed recognition tasks

AUTHORS:

Colin Tredoux^{1,2}
Ahmed M. Megreya³
Alicia Nortje¹
Kate Kempen¹

AFFILIATIONS:

¹Department of Psychology, University of Cape Town, Cape Town, South Africa
²Cognition, Languages, Language, Ergonomics Laboratory, University of Toulouse – Jean Jaurès, Toulouse, France
³Department of Psychological Sciences, Qatar University, Doha, Qatar

CORRESPONDENCE TO:

Colin Tredoux

EMAIL:

colin.tredoux@uct.ac.za

DATES:

Received: 31 Aug. 2021

Revised: 25 Nov. 2022

Accepted: 25 Nov. 2022

Published: 29 Mar. 2023

HOW TO CITE:

Tredoux C, Megreya AM, Nortje A, Kempen K. Changes in the own group bias across immediate and delayed recognition tasks. *S Afr J Sci.* 2023;119(3/4), Art. #12126. <https://doi.org/10.17159/sajs.2023/12126>

ARTICLE INCLUDES:

- Peer review
- Supplementary material

DATA AVAILABILITY:

- Open data set
- All data included
- On request from author(s)
- Not available
- Not applicable

EDITOR:

Leslie Swartz

KEYWORDS:

own group bias, own race bias, face recognition, encoding

FUNDING:

Programme for the Enhancement of Research Capacity (PERC), University of Cape Town; South African Centre for Digital Language Resources (SADILAR), Department of Science and Technology of South Africa as part of the South African Research Infrastructure Roadmap (SARIR); Oppenheimer Memorial Trust



Face recognition is biased in favour of in-group identity, particularly strongly for race or ethnicity but to some extent also for sex and age. This 'own group bias' (OGB) can have profound implications in practical settings, with incorrect identification of black suspects by white witnesses constituting 40% of criminal exonerations investigated by the Innocence Project. Although authors have offered several explanations for the OGB in face recognition, there is little consensus, apart from the acknowledgement that the bias must reflect perceptual learning history. One matter that is not currently clear is whether the bias occurs at encoding, or at retrieval from memory. We report an experiment designed to tease out bias at encoding, versus bias at retrieval. Black and white South African participants encoded 16 target faces of both the same and other race and gender, and attempted immediately afterward to match the target faces to members of photograph arrays that either contained or did not contain the targets. After a further delay, they attempted to identify the faces they had encoded from memory. Results showed a strong crossover OGB in the delayed matching task, but an asymmetrical OGB at retrieval (only white participants showed the OGB). Further investigation of recognition performance, considering only images correctly matched in the delayed matching task, showed a narrowly non-significant OGB at retrieval, but the investigation was likely not sufficiently powered to discover the effect, if it exists.

Significance:

- We demonstrate the presence of a crossover OGB in face recognition in a sample of black and white South Africans in a delayed matching task (a measure of encoding).
- Our findings show that the OGB may change rapidly. In the present study, the OGB took a crossover form at retrieval immediately after encoding, but was asymmetrical when assessed shortly afterwards.
- We used a novel approach for disentangling effects at encoding and at retrieval, but do not provide clear evidence to distinguish whether the OGB is a failure of encoding or of memory retrieval.

Introduction

The own group bias (OGB)¹ refers to a recognition advantage for faces of members of one's own group. The most common formulation posits an unequal advantage: groups that co-occupy a common environment will show an advantage for members of their own group over members of another group. However, the recognition advantage is typically asymmetrical in countries or regions where demographic or economic representation is unequal (e.g. the United Kingdom²). Members of the dominant group will usually show an OGB, but members of the subordinate groups often do not show it.^{2,3} The practical implications of the OGB are profound. The Innocence Project in the USA⁴ has shown, for instance, that over 40% of exonerations of people falsely imprisoned based on eyewitness identifications were convicted in part as a result of cross-race eyewitness identifications – a figure far out of proportion to the demography of the USA.

Explanations of the OGB in the literature are plentiful, but it is probably fair to say that no single account has won out empirically or theoretically. All theories appear to accept the notion that the OGB must be a consequence to some degree of differential exposure to one's own and other group faces.⁵ Our interest is in what might be called the micro-chronology of the OGB. Whereas the OGB is often referred to in the literature as a memory bias – for instance Hugenberg and colleagues declaring their intent to "...explain the proliferation of own group biases in face memory"^{6(p.1392)}, and Yaros and co-authors⁷ arguing that the OGB may emerge due to "tuned" memory mechanisms – researchers have pointed out for a number of years that the bias may stem from the differential ability to encode faces of other groups, rather than a reduced ability to recognise or retrieve them from memory⁸. Recent claims are that the OGB is likely due to the less efficient encoding of out-group faces in visual working memory^{9,10}, or to preferential distribution of attention during encoding¹¹.

Megreya and co-authors¹² have published research that questions the basis of the OGB in memory processes, arguing instead that the OGB may be entirely due to encoding difficulties. Participants in Megreya et al.'s study had difficulty in matching the face of a 'target' in a film still to digital photographs of the same person. This finding was more pronounced when participants attempted to match faces from a group different from their own. This outcome has important implications, especially for theoretical explanations of the OGB bias: many theoretical accounts have focused on memory retrieval operations or are at least unclear on the relative role of encoding and retrieval in the provenance of the OGB, and this should be addressed if the phenomenon to be explained is entirely about encoding rather than retrieval. But this may be somewhat premature: at this stage in the development of thinking about the OGB, it is not clear what the relative contribution of each process is. Indeed, several authors have offered evidence that either questions whether the deficit is in any way encoding based, or that posits that it is in part retrieval based. Papesch and Goldinger¹³ were able to show that disruptions during retention intervals modulated the OGB for out-group faces, but not for in-group faces. Stelter et al.¹⁴ showed that eye movement activity differed across in-group and out-group targets during a recognition, but not an encoding phase. They also showed that the OGB was a function of poor performance in recognition of new faces specifically.

© 2023. The Author(s). Published under a Creative Commons Attribution Licence.

Our aim was to explore the relative contributions of encoding and retrieval in the manifestation of the OGB in face recognition. It is important at the outset to acknowledge that it is difficult to disentangle encoding from recognition, and in this respect the difficulty appears to lie primarily in establishing that a stimulus has been encoded. We offer one approach, which is to test recognition with a delayed matching task, and then after a longer period with a recognition memory task. In a delayed matching task, participants view a stimulus, and then immediately attempt to match the stimulus in a test array once it has been removed from view. If participants are accurate in the delayed matching task, this suggests successful encoding. Note that Megreya et al.¹² did not use a delayed matching task, but an in-view matching task. In other words, they attempted to match a target image to one embedded in an array of images while both the target and the array were in full view. However, we believe that in-view matching tasks do not demonstrate encoding as clearly as delayed matching tasks because matching targets in-view might be accomplished without encoding faces at all. Participants may be able to use non-face properties of stimuli, such as colour temperature, or other image artefacts to effect the match, or might be able to do so with minimal encoding of the target image, e.g. by relying on specific, transient and localised characteristics of faces, such as skin blemishes, and hairstyle oddities. This distinction between face identity encoding and face image encoding was powerfully made over 40 years ago.¹⁵ For this reason we departed from the procedure used by Megreya et al.¹² There were some further departures, too, that we thought important. In the first instance, we wished to see whether the results they obtained replicate across different viewpoints and with different stimulus materials. They used only frontal views of participants, in both matching and target views, and they also appeared to have made their matching arrays very difficult (the faces in the array were the most similar in a moderately large set). Repeating their experiment with different viewpoints of the faces (e.g. in profile vs. frontal views) would test whether the memory trace of the face is well enough encoded to match and recognise it over different transformations (i.e. different face poses). Furthermore, repeating their experiment with line-up arrays that are constructed around faces that are not specifically selected for high similarity seems important to us, especially as one of the most significant potential applications of knowledge about the OGB is in the eyewitness domain, where line-ups are unlikely to be constructed in this way (see Wells et al.¹⁶ for a discussion of why this is not a good idea).

However, the most important goal in our experiment was to extend Megreya et al.'s¹² original experiment and to attempt to separate out encoding and retrieval processes, and to estimate the relative importance of each in the manifestation of OGB bias. Thus, in our experiment, participants were tested for delayed face matching ability, and, after a suitable delay, for their ability to recognise the face to which they were exposed in the delayed matching task. We used a line-up task, designed to permit the calculation of the signal detection theory (SDT) measures d' (discrimination) and c (criterion). Through this, and by calculating successful recognition of faces that were incorrectly matched in the first part of the procedure, we attempted to assess the relative contributions of encoding and retrieval deficits in the OGB effect.

Method

Target stimuli

Stimuli were digital images taken from a large database of black South African and white South African faces (containing multiple views of over 1000 individuals) and curated by the first author. Sixteen target faces were used in this experiment: four male and four female black South African faces, and the same again for white South African faces. The target faces were randomly sampled by the authors from the total collection, and target ratings were obtained to ascertain whether there were any idiosyncrasies present in the target images that may have resulted in them being more memorable than the other target faces. A total of 22 (15 female) ($M_{\text{age}} = 20.52$ years; $SD_{\text{age}} = 1.21$ years) participants rated the target faces on a number of dimensions that are known to affect face recognition: typicality, distinctiveness, attractiveness, perceived criminality, age, wealth, memorability, and familiarity. Each

face was rated on a scale from 0 (not at all) to 8 (extremely), whereas age was estimated numerically. Two images of the same target were presented side-by-side during the rating task: one photograph was of the target in a neutral three-quarter side pose, the other image was of the target in a frontal casual/smiling pose. Each image pair was presented one at a time, along with a randomised order of rating dimensions. The size of each face image was approximately 7.94 cm in width and 10.35 cm in height, with a resolution of 300 x 391 pixels (8-bit). The image background and the target clothing were edited to be standard and consistent across all the target images.

Line-up construction

Target present (TP) and target absent (TA) photographic line-ups were constructed for the target images. TP line-ups contained the target and five foils. In the TA line-ups, the target was replaced with a foil that was randomly selected by the authors, resulting in a six-foil line-up (i.e. there was no designated suspect). No foils were repeated or appeared in any of the other line-ups. The line-up members were selected to be subjectively moderately similar in appearance to the target (i.e. following the principle set down in Wells et al.¹⁶ that line-up members should be matched to the target on some but not all characteristics). Three of the present authors independently selected ten possible foils for each target face from a database of hundreds of black South African and white South African faces. The most frequently chosen foils were used as the final images in the line-ups.

Corresponding to the 16 target faces, each participant saw 16 line-ups. The line-ups appeared in one of two orders: eight of the line-ups were TP line-ups, and the other eight were TA line-ups. Two line-ups were created for each target face: a frontal pose line-up, and three-quarter pose line-up. The frontal neutral pose and the three-quarter pose were used to control for picture recognition and to ensure that identification was made on memory for the target and not the target photograph. The line-up photographs were in colour and standardised. The backgrounds were edited to remain consistent across all the images. All clothing, jewellery, and distinctive markings were digitally removed from the images. The size of each face image was approximately 6 cm x 7.8 cm, with a resolution of 227 x 296 pixels (8-bit). The image background and the target clothing were edited to be consistent across the target images.

Participants

A total of 64 (53 female) participants from the University of Cape Town participated in the study in exchange for course credit; 32 (50%) of the participants identified themselves as white South Africans, and 32 (50%) of the participants identified themselves as black South Africans. All participants reported normal or corrected to normal vision.

Procedure

Encoding and delayed matching phase

The study was conducted at the University of Cape Town, in a quiet computer laboratory. The experiment was presented on computer using E-Prime 2.0, at a resolution of 1024 x 768 pixels.

After providing demographic information, participants were informed that they would be presented with a series of target faces, one at a time, for five seconds. One of the target faces appeared on the screen, in a frontal casual/smiling pose. The size of the target image was 12 cm x 15.7 cm, at a resolution of 456 x 594 pixels.

After five seconds, the face disappeared, and a six-member simultaneous TP or TA line-up was immediately displayed, in a delayed matching task. Participants either saw a three-quarter line-up or a frontal line-up. Participants were informed that the target they had just seen may or may not be present in the line-up. They were cautioned that the clothing and background of the target image may be different from the image they had studied. If participants recognised the target, they had to indicate the corresponding number above the line-up member on the keypad; if they thought the target was not present, they had to indicate '0' on the keypad. This delayed matching procedure was repeated for the

remaining 15 target faces; the order of the target faces and matching line-up pairs was randomised across participants.

When the line-up delayed matching phase was finished, participants completed a filler task for 10 minutes (a Sudoku game).

Recognition phase

Following the distractor task, participants were told that they would view a series of six-member line-ups and that their task was the same as before – to select the target face that they had studied earlier. They were informed that the line-ups may be different from the line-ups they had seen in the delayed matching session. Participants were again cautioned that the target face may or may not be present in the line-up. It was emphasised that the line-ups would appear in a random order and would not necessarily correspond with the order in which the participants had studied the target faces.

Participants attempted to identify the targets from the same line-ups used in the delayed matching procedure. For example, if they saw a TA three-quarter pose line-up at delayed matching, they saw the same line-up at recognition. The position of the line-up members was rearranged between delayed matching and recognition to guard against commitment effects – if participants recognised the line-up from the delayed matching session, they may have been tempted to select the same line-up position as before.

If participants recognised the target, they indicated the corresponding number above the line-up member on the keypad; if they thought the target was not present, they selected the appropriate option on the keypad. This recognition procedure was repeated for the remaining 15 line-ups; the order of arrays was randomised across participants.

When all recognition trials were completed, participants who responded '0' (not present) to a TP line-up – i.e. who had incorrectly rejected the line-up – were given another chance to respond to the line-up. They were shown the same line-ups to which they had responded 'not present' and were then required to make a selection (i.e. a forced choice; however, data from the forced choices were not included in the present analyses). The order of these line-ups was randomised.

Once participants had completed their forced-choice line-up tasks, they were thanked, debriefed, and dismissed from the study. Please note that

further details regarding preparation of stimuli, as well as descriptive statistics of task performance, appear in Figure 1.

Results

Hit and false alarm scores were computed for each participant and were used to derive discrimination (d') and criterion (c) signal detection measures. Because we did not use designated suspects in our TA line-ups, false alarms were computed as the total number of identifications of foils in TA line-ups. (Although eyewitness researchers often divide the total number of foil identifications to estimate the false alarm rate, this constrains the false alarm rate to a maximum of $1/\text{number_of_line-up members}$. This may be appropriate for a perfectly fair line-up, but very few line-ups are perfectly fair.¹⁷ We have not divided by the number of foils. The consequence is that false alarms are more probable than hits, and this can be expected to decrease d' , but as long as one recognises that it is comparisons across d' and c that are important, not absolute levels, it should make little difference. Judging absolute levels might lead one to believe that participants in this study had poor discrimination in some conditions (indeed, $d' < 0$ in some conditions), but that is not a real reflection of ability).

Our analyses are generally based on these SDT measures, rather than on raw hit, false Alarm, or other untransformed accuracy measures. Figure 2 shows raw measures for the information of interested readers.

We conducted a four-way linear mixed model analysis of the experimental data, that is stimulus group (black vs. white) X line-up face view (frontal vs. three-quarter) X task condition (delayed matching vs. recognition) X participant group (black vs. white), using the package LME4¹⁸, within R¹⁹. 'Stimulus group' and 'task condition' were within-participant factors. The analysis of variance revealed several significant effects, as shown in Table 1.

At the level of main effects, we observed differences in d' across 'line-up face view', and for 'task condition' ($p < 0.05$). We found two-way interactions for both 'task condition and stimulus group', and for 'stimulus group and participant group'. Finally, we found three-way interactions for 'participant group, stimulus group, and task condition', as well as for 'stimulus group, task condition and line-up face view'.

The most important findings from the analysis for our concerns here are the classic two-way interaction of 'stimulus group and participant group' (the OGB effect), and the three-way interaction of 'stimulus group,

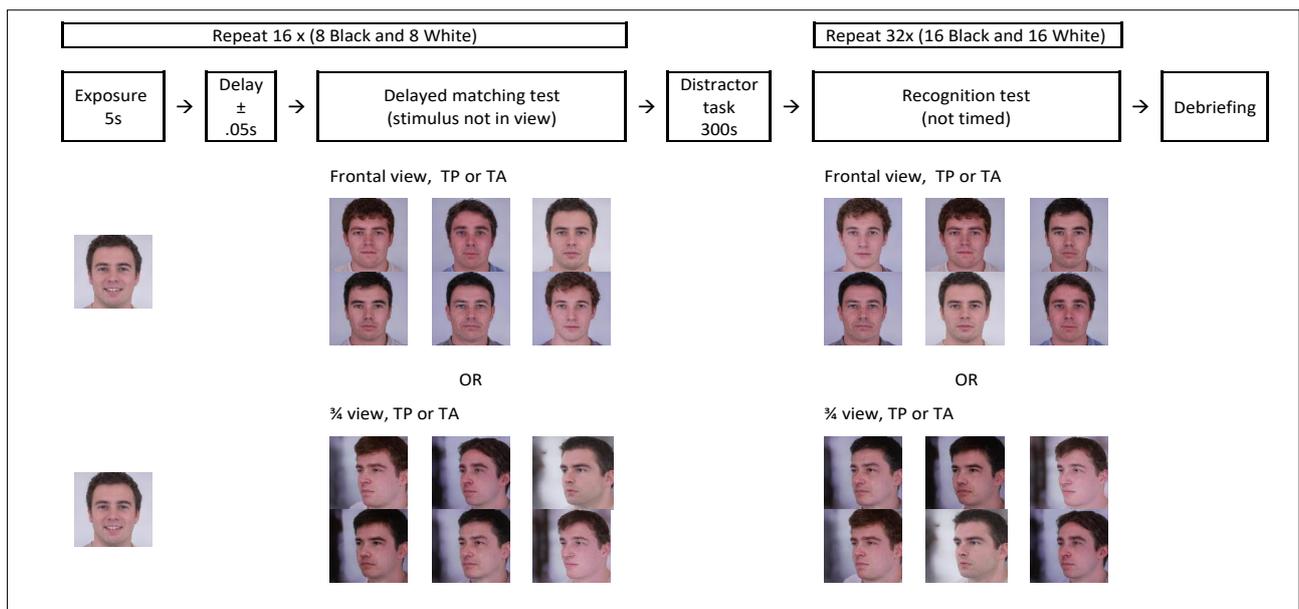


Figure 1: Experimental procedure, showing timeline and other details. For display purposes, faces are synthetic, i.e. not of real people. Assignment to target present (TP) or target absent (TA), and to frontal or three-quarter view, was random. Faces at exposure were casual or smiling, and neutral at delayed matching and/or recognition.

participant group, and task condition'. The latter effect is particularly important to our enquiry – it tells us that the OGB effect in the present sample differs between the two types of task condition (delayed matching and recognition).

Table 1: Mixed linear model ANOVA on discrimination (d'), and criterion (c), by participant group, stimulus group, task condition (delayed match vs recognition), and view (frontal vs three-quarter). Only fixed effects are shown; for random effects ICC = 0.49 for d' , and 0.48 for c . β coefficients are standardised. All $df_1 = 1$, and $df_2 = 60$.

Effect	Discrimination (d')			Criterion (c)		
	F	p -value	$ \beta $	F	p -value	$ \beta $
Participant group	0.03	0.869	0.87	3.40	0.070	0.53
Stimulus group	0.01	0.905	0.55	0.50	0.484	0.54
Task condition	34.37	0.001	0.93	12.58	0.001	0.00
View	11.50	0.001	0.89	2.73	0.104	0.10
Participant : Stimulus	26.29	0.001	1.08	5.70	0.020	1.25
Participant : Task	0.20	0.653	0.29	0.01	0.912	0.81
Stimulus : Task	5.05	0.028	0.30	0.00	0.958	0.72
Participant : View	2.38	0.128	0.49	0.37	0.546	0.81
Stimulus : View	0.24	0.627	0.53	0.61	0.439	0.40
Task : View	9.42	0.003	0.28	0.03	0.871	0.51
Participant : Stimulus : Task	4.10	0.047	0.41	3.91	0.053	1.19
Participant : Stimulus : View	0.12	0.727	0.36	0.89	0.350	0.87
Participant : Task : View	0.89	0.349	0.08	0.51	0.476	0.86
Stimulus : Task : View	2.74	0.103	0.76	0.45	0.506	0.74
Participant : Stimulus : Task : View	0.53	0.467	0.47	2.09	0.154	1.01

A classic bias in face recognition ability is for own group members to recognise faces belonging to their own group with greater facility than members of other groups. Megreya et al.¹² have argued though that this facility is also present when participants are asked to match faces to arrays: own-group faces are matched more accurately than out-group faces. This raises the question of whether there is a retrieval component at all in the own-group bias in face recognition – the effect could be entirely due to poorer encoding of out-group faces. In the present experiment, we observed the classic crossover interaction of stimulus and participant group, on d' , as shown in Figure 3. White participants were better at delayed matching and recognising white faces (mean $d'=0.52$, $SD=0.95$) than they were at delayed matching and recognising black faces (mean $d'=0.02$, $SD=0.89$), and black participants were better at delayed matching and recognising black faces (mean $d'=0.56$, $SD=1.22$) than they were at delayed matching and recognising white faces (mean $d'=0.04$, $SD=1.12$). A set of linear contrasts showed that white South Africans performed better, overall, with white South African faces than with black South African faces ($t(60)=3.54$, $SE=0.14$, $p<0.001$, Cohen's $d=0.52$). Cohen's d was estimated by dividing the mean difference by the square root of the sum of the variances of the random, and residual effects²⁰), and that black South Africans performed better with black South African faces than with white South African faces ($t(60)=3.71$, $SE=0.14$, $p<0.001$, Cohen's $d=0.54$). As we were interested in testing whether viewing faces in frontal vs three-quarter view would impact the OGB, it is worth noting that the 'view' factor was not implicated in any interaction involving either participant group or stimulus group, but it was involved in a significant interaction with task condition, showing that participants found it difficult to match or recognise faces across frontal and three-quarter views, but this effect was greater for the delayed matching condition than for the recognition condition ($\Delta d'=0.89$ vs 0.34).

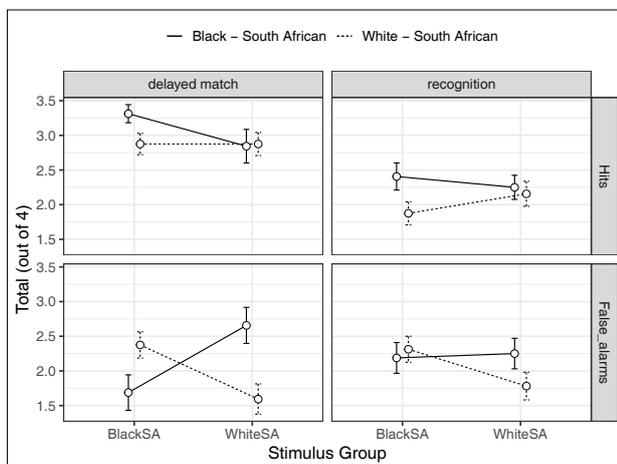


Figure 2: Hits and false alarms as a function of stimulus group, race of respondent, and task type. Error bars are standard errors of the mean.

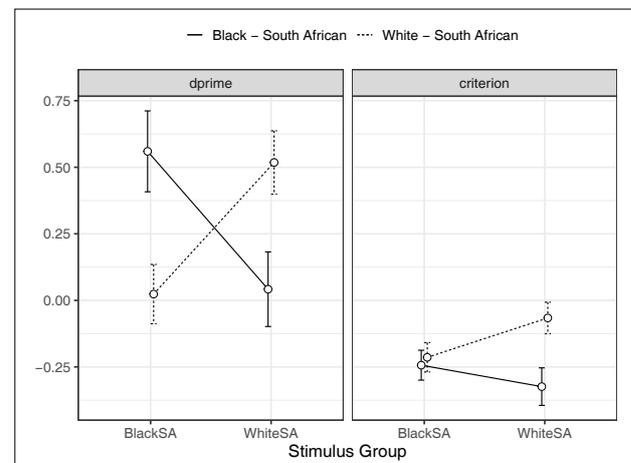


Figure 3: Own group bias in discrimination (d') and criterion (c), collapsed over delayed matching and recognition tasks. Error bars are standard errors of the mean. Both interactions shown here are statistically significant.

Most pertinent to our concerns was the three-way interaction of participant group, stimulus group, and task condition – or in other words, the two-way interaction of participant group and stimulus group (the OGB effect) considered as a function of which task was being completed. As Table 1 shows, this effect was significant, and Figure 4 shows the effect graphically. Follow-up contrasts showed that, in the delayed matching task, both groups of participants were better at matching their own-group faces: black South Africans were better at matching black South African faces than white South African faces ($M_{black-black}=0.98$ vs $M_{black-white}=0.1$, $SD_{black-black}=1.02$ vs $SD_{black-white}=1.35$, $t(118)=4.78$, $p<0.001$, Cohen's $d=0.92$), and white South Africans were better at matching white South African faces than black South African faces ($M_{white-white}=0.79$ vs $M_{white-black}=0.31$, $SD_{black-black}=0.93$ vs $SD_{black-white}=0.88$, $t(118)=2.57$, $p<0.012$, Cohen's $d=0.50$).

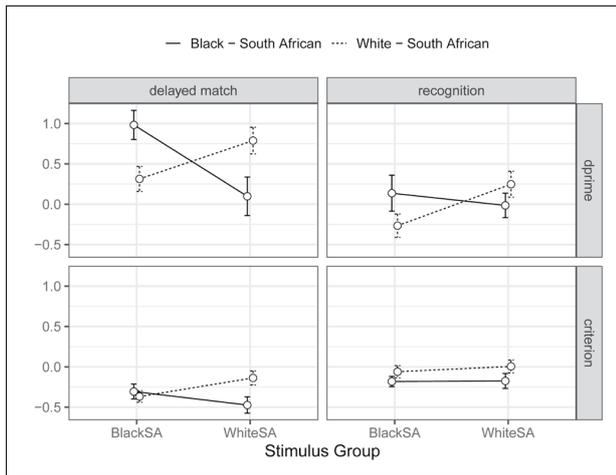


Figure 4: Own group bias in discrimination (d') and criterion (c), across delayed matching and recognition tasks. Error bars are standard errors of the mean. At matching, both black and white participants show the OGB in d' , but only the white participants show the OGB at recognition. The interactions between stimulus group and participant group are not significant.

In the recognition task, however, white South Africans showed the OGB ($M_{\text{white-white}}=0.25$ vs $M_{\text{white-black}}=-0.27$, $SD_{\text{white-white}}=0.91$ vs $SD_{\text{white-black}}=0.82$, $t(118)=2.8$, $p<0.007$, Cohen's $d=0.54$), whereas black South Africans did not ($M_{\text{black-black}}=0.14$ vs $M_{\text{black-white}}=-0.01$, $SD_{\text{black-black}}=1.26$ vs $SD_{\text{black-white}}=0.87$, $t(118)=0.82$, $p<0.416$, Cohen's $d=0.16$).

In addition to d' , we computed criterion for each participant in the experiment. The analysis of criterion as the dependent variable for the full four-way design produced a significant two-way interaction between participant group and stimulus group ($F(1, 60)=5.70$, $p<0.02$). This interaction shows that white and black participants were both more conservative when considering faces of their own group than when considering faces of the other group ($M_{\text{white-white}}=-0.07$ vs $M_{\text{white-black}}=-0.21$, $SD_{\text{white-white}}=0.47$ vs $SD_{\text{white-black}}=0.44$, and $M_{\text{black-black}}=-0.24$ vs $M_{\text{black-white}}=-0.32$, $SD_{\text{black-black}}=0.45$ vs $SD_{\text{black-white}}=0.56$). The three-way interaction between participant group, stimulus group, and task condition was narrowly not significant ($F(1, 60)=3.91$, $p<0.054$). We show the marginally non-significant three-way interaction in Figure 4, where it seems that (1) there is a typical OGB for criterion present, and that (2) the OGB may be apparent in the delayed matching task, but not in the recognition task.

Thus far the results have shown a clear crossover OGB in the delayed matching task, as well as an asymmetrical OGB in the recognition task. It is not clear that the OGB in the recognition task can be said to be independent of that in the delayed matching task – that is, it seems evident that faces that are not matched or recognised accurately in the delayed matching task will not be recognised accurately in the later task either. Failure in the matching task may also be an indication of encoding failure. In order to test the point that Megreya et al. make about the possible entire dependency of the OGB on encoding processes, we reduced our data set to only those faces that had been accurately matched in the delayed matching phase and re-analysed the recognition data for just those faces. The idea here was to choose faces that we were confident had been encoded successfully.

One implication of selecting only faces that were correctly matched in the delayed matching task, is that the OGB shown in the delayed matching task was effectively controlled for – as all faces were correctly matched, there could be no bias. This reduction left us with 78% of our original participants (selecting those who made at least one correct decision at delayed matching), but as some participants performed better than others in the delayed matching task, this resulted in an uneven distribution of stimuli across conditions. Thus, to take the extreme conditions, whereas 86% of the original stimuli were taken into account for black participants viewing white South African faces in TP conditions, 46% of stimuli were taken into account for white participants viewing black South African

faces in TA conditions. This 'attrition' undoubtedly affected the potential power of analyses of recognition of stimuli that had been successfully matched in the delayed matching task.

A linear mixed model testing discrimination (d') for faces correctly matched in the delayed matching task showed a non-significant effect for the key two-way interaction of interest (between participant group and stimulus group), that was significant in the earlier model ($F(1, 49.35)=2.80$, $p<0.104$). Although the interaction effect was not significant, it seemed appropriate to us to follow up with focused contrasts directly exploring a potential OGB, and these showed that, in the recognition task, testing recognition of only faces successfully matched in the delayed matching task, neither white nor black South Africans were better at their own-group faces, although this was narrowly not the case for white South Africans (for black South Africans: $M_{\text{black-black}}=0.38$ vs $M_{\text{black-white}}=0.29$, $SD_{\text{black-black}}=0.59$ vs $SD_{\text{black-white}}=0.46$, $t(64.8)=0.18$, $p<0.86$, Cohen's $d=0.18$, and for white South Africans: $M_{\text{white-white}}=0.33$ vs $M_{\text{white-black}}=0.09$, $SD_{\text{white-white}}=0.43$ vs $SD_{\text{white-black}}=0.47$, $t(43.5)=1.94$, $p<0.059$, Cohen's $d=0.48$). While the asymmetrical OGB seen in the recognition test using all stimuli was not technically significant in the recognition test using only faces successfully encoded (matched in the delayed matching task), it was very close to being so. Of course, this analysis is based of necessity on a smaller sample, and has less statistical power, so it is not strong evidence that controlling for the OGB in the matched delay task (our operationalisation of encoding) eliminates the OGB at recognition. We also conducted a mixed linear model analysis on criterion scores, but did not find any significant results, or any suggestion of an effect (all $p>0.48$).

Discussion

The OGB in face recognition is a well-established phenomenon, with serious consequences in applied contexts, especially when recognition is treated as person identification, as one finds in law enforcement procedures. Despite often being the subject of empirical and theoretical investigation, not much is known about the cognitive processes underlying the phenomenon. One important question concerns whether the OGB is an encoding or a recognition phenomenon, and we have brought results from an empirical study to bear on this question, or at least on the micro-chronology of the OGB.

Black and white South African participants in our study were asked to match black or white target faces to corresponding images in arrays of same group faces immediately after viewing them. They were then asked, after an intervening period in which they completed a distractor task, to recognise the target faces from memory, in the same arrays. Our results show a strong OGB in discrimination accuracy for both black and white participants at the delayed matching phase of the study, but only white participants showed the bias at the recognition memory phase of the study. We found similar results for the measure of response criterion that we computed, although the result for the important interaction of participant group, stimulus group, and task condition was narrowly not significant. Our finding of an asymmetry in the delayed recognition task for discrimination, rather than in the delayed matching task, is partly novel, and partly in line with extant research. Most studies that have been conducted on the OGB (for race) in South Africa have reported an asymmetric OGB at recognition. Usually, white participants show a strong OGB, whereas black participants rarely show an OGB at all (but see Wittwer et al.'s Study 6²¹ for an example of a crossover), and sometimes show an inverse OGB, recognising white faces with more accuracy than they do black faces²²). Some authors, including ourselves, have ascribed this to the way in which the social and political context have structured intergroup contact in South Africa (and other countries with similar intergroup histories), and therefore perceptual contact. It is not easy, though, to explain the finding of a crossover OGB among black participants at encoding (the delayed matching task) that fails to materialise in the recognition test. One possibility is that performance on the recognition task bottomed out, approaching basement level, but this does not seem to have occurred differentially for black vs white participants, and is not a convincing explanation. It is possible that it could reflect a real difference in memory consolidation of out-group faces by black participants. What does seem clear to us is that the argument that

the typical asymmetric OGB between majority and minority groups is due to socio-political factors – e.g. that social and political inequality might impact perceptual learning of out-group faces – is made more difficult to sustain by our results. If we had considered only data from the recognition task we might well have reported another instance of an asymmetric OGB, but in looking at the encoding/delayed matching task, we noted a crossover OGB. A source of evidence that is sometimes marshalled in an attempt to understand the OGB as it manifests in particular contexts is the perceptual contact history of participants. Methods of assessing past perceptual history generally rely on self-reporting, and in the meta-analysis reported in Meissner and Brigham¹ were only poorly correlated with the OGB ($r=0.13$). We did not collect self-reported data on perceptual contact history, but do not believe it would have shed any light on the nature of the OGB in our study, given its poor predictive record reported in Meissner and Brigham¹. We thus do not have a good sense of why we observed an OGB at encoding but not at recognition for black participants. An anonymous reviewer has suggested that it could be a task-dependent difference, possibly interacting with social-motivational factors, which we think is possible. On the more general point of why there is often an asymmetric pattern for the OGB, with disadvantaged groups showing a weaker or non-existent OGB when compared to advantaged groups, we think that Malpass²³ notion of a ‘social interaction utility’ is useful: members of the disadvantaged group have a positive utility associated with interaction and person recognition (often being economically dependent on the advantaged group, it is important to interact with, and recognise members of that group), whereas members of the advantaged group have a negative utility (there is no clear advantage to interaction and recognition).

An important question in research on the OGB concerns the degree to which the OGB is a function of encoding, or of recognition. The use of delayed matching and recognition tasks allowed this to be addressed in our study. We conducted a second analysis on images correctly matched (therefore showing no OGB), and although we did not find an OGB for these images at recognition using a statistical significance test as the criterion, the size of the effect we observed in this task ($d=0.48$) was very similar to that observed in the recognition task ($d=0.52$). Our results likely have equivocal bearing in the end on whether the OGB is an encoding phenomenon, rather than a recognition memory phenomenon. What is most interesting about our findings, though, and which has nothing to do with the final task we used, is that we observed a strong crossover OGB at delayed matching, which is not at all usual in South African studies, and after a brief delay of 5 minutes, the OGB manifested in a recognition memory task as an asymmetrical effect, which is typically the form other studies in South Africa have previously reported.

The claim that the OGB is an encoding phenomenon makes sense from a perceptual learning viewpoint, especially in line with Valentine’s multidimensional face space model²⁴, and his explanation of the OGB in terms of the model²⁵: the dimensions available to represent out-group faces are fewer, and less well developed than for in-group faces, and encoding of out-group faces will be less well differentiated from other exemplars of that group. Of course, our method for separating encoding and recognition processes is admittedly ad hoc: we tested recognition of faces that had been correctly matched, thus attempting to control encoding processes. We did not directly show that own-group faces were better encoded than out-group faces, although in demonstrating an OGB at encoding we think that differences were implicitly demonstrated. Face encodings are complex patterns of electrical activation across multiple brain regions, and not directly accessible, although there are some clues to the neural underpinnings of the OGB and its connections to socio-affective processes.²⁶ Some researchers have shown differential event-related potentials to own- and other-group faces²⁷, with a potentially special role for the P200 visual component of the event-related potentials²⁸. These investigations suggest that we are making progress toward a better understanding of the brain mechanisms underlying the OGB.

Apart from the obvious limitation of not being able to assess encoding proficiency directly, there are other limits to what we are able to conclude from our study. An important methodological limitation may be our method of checking that a stimulus had been encoded, and the knock-on effect of this for the accuracy of our recognition test. As a reminder, we

assessed encoding with a delayed matching task: participants attempted a match immediately after viewing a target stimulus. After a further delay of 5 minutes, we assessed recognition memory by asking participants to choose the target image from a line-up, which either contained or did not contain the target. Our recognition test could thus be said to have exposed participants to the target stimulus twice, and this taints the comparison between encoding and recognition in which we are interested. In other words, although the delayed matching task tests memory for a stimulus seen once, the recognition task tests memory for a stimulus seen once on its own, and once within an array. It is important to note that participants were not told the position of the target in the array in the delayed matching task, and that the photographs of the target changed from the original presentation (frontal casual/smiling view) to that in the line-up arrays used in the delayed matching and recognition tasks (three-quarter profile, or neutral/passport style view). It is still possible that there was some strengthening of the memory assessed in the recognition memory task, but our results show that memory performance decreased considerably between the delayed matching task and the recognition memory task (see Figures 3 and 4), so it does not seem to have counteracted that decrease, if at all. It would also have been unlikely to affect our white and black participants differentially, and because our key effect of interest was the nature of the OGB at the two time points, it does not appear to us to be confounded with the two- or three-way interactions in which we were interested.

Whereas we studied matching and recognition of two-dimensional images of faces, face recognition in real contexts is of three-dimensional surfaces that change locally, and globally, over time, in interaction with perceivers. We also used a short delay period between encoding (matching), and recognition, unlike many face recognition tasks ‘in the wild’. However, it is likely that the OGB would be higher in more naturalistic contexts, because there are known additional biases across groups for emotion²⁹ and age³⁰, among other biases, and we do not see a reason to believe that this would differentially affect encoding and recognition deficits already present. There is an extensive and convincing argument in support of the view that one expects real witnesses to perform far worse than participants in laboratories.³¹

In conclusion, although we have not reported clear evidence in favour of the view that the OGB in face recognition is likely a consequence of poor encoding of other group faces, we have identified an interesting, rapid change in the manifestation of the OGB. Whilst it took a strong crossover form at encoding (delayed matching), it reverted to an asymmetric form after a brief delay. There are some applied implications of this result, although the evidence is at this stage too slight to base strong recommendations on. If the OGB is a failure of encoding then there may be little justification for devising methods that focus on improving recall or recognition of out-group faces, and it may be better to develop training programmes that focus on encoding processes. There is some evidence that in-group members focus on different face regions when encoding out-group faces than out-group members do³², although it is not yet clear that reshaping cross-group encoding will work³³. On the other hand, if the OGB is a failure of retrieval, some methods such as the Cognitive Interview³⁴ and the Person Description Interview³⁵ may be useful interventions when recovering information from memory about face appearance and identity – although it should be borne in mind that these methods are typically good for improving recall memory, but not recognition memory³⁶. It could also be that the OGB is a function of both encoding and retrieval processes, and we are presently considering ways of adapting our procedure to accommodate this possibility.

Ethical considerations

Ethical clearance to conduct the study was granted by the Department of Psychology Ethics Committee of the University of Cape Town. All participation was voluntary and informed consent was obtained before study commencement.

Competing interests

We have no competing interests to declare.



Authors' contributions

C.T.: Conceptualisation, method, data analysis, data curation, writing, project leadership, funding acquisition, student supervision.
 A.M.M.: Conceptualisation, method, writing, project leadership. A.N.: Conceptualisation, method, data collection, data analysis, writing – revision.
 K.K.: Conceptualisation, method, data collection, writing – revision.

References

- Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces. *Psychol Public Policy Law*. 2001;7(1):3–35. <https://doi.org/10.1037/1076-8971.7.1.3>
- Walker PM, Hewstone M. A perceptual discrimination investigation of the own-race effect and intergroup experience. *Appl Cogn Psychol*. 2006;20(4):461–475. <https://doi.org/10.1002/acp.1191>
- Malpass RS, Kravitz J. Recognition for faces of own and other race. *J Pers Soc Psychol*. 1969;13(4):330–334. <https://doi.org/10.1037/h0028434>
- Innocence Project. How eyewitness misidentification can send innocent people to prison [webpage on the Internet]. c2020 [cited 2023 Feb 03]. Available from: <https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/>
- Rhodes G. Face recognition. In: Reisberg D, editor. *The Oxford handbook of cognitive psychology*. New York: Oxford University Press; 2013. p. 46–68.
- Hugenberg K, Wilson JP, See PE, Young SG. Towards a synthetic model of own group biases in face memory. *Vis Cogn*. 2013;21(9–10):1392–1417. <https://doi.org/10.1080/13506285.2013.821429>
- Yaros JL, Salama DA, Delisle D, Larson MS, Miranda BA, Yassa MA. A memory computational basis for the other-race effect. *Sci Rep*. 2019;9(1):19399. <https://doi.org/10.1038/s41598-019-55350-0>
- Bornstein BH, Laub CE, Meissner CA, Susa KJ. The cross-race effect: Resistant to instructions. *J Criminol*. 2013;2013:745836. <https://doi.org/10.1155/2013/745836>
- Zhou X, Mondloch CJ, Emrich SM. Encoding differences affect the number and precision of own-race versus other-race faces stored in visual working memory. *Atten Percept Psychophys*. 2018;80(3):702–712. <https://doi.org/10.3758/s13414-017-1467-6>
- Stelter M, Degner J. Investigating the other-race effect in working memory. *Br J Psychol*. 2018;109(4):777–798. <https://doi.org/10.1111/bjop.12304>
- Brown TI, Uncapher MR, Chow TE, Eberhardt JL, Wagner AD. Cognitive control, attention, and the other race effect in memory. *PLoS ONE*. 2017;12(3), e0173579. <https://doi.org/10.1371/journal.pone.0173579>
- Megreya AM, White D, Burton AM. The other-race effect does not rely on memory: Evidence from a matching task. *Q J Exp Psychol (Hove)*. 2011;64(8):1473–1483. <https://doi.org/10.1080/17470218.2011.575228>
- Papesh MH, Goldinger SD. Deficits in other-race face recognition: No evidence for encoding-based effects. *Can J Exp Psychol*. 2009;63(4):253–262. <https://doi.org/10.1037/a0015802>
- Stelter M, Rommel M, Degner J. (Eye-) tracking the other-race effect: Comparison of eye movements during encoding and recognition of ingroup faces with proximal and distant outgroup faces. *Soc Cogn*. 2021;39(3):366–395. <https://doi.org/10.1521/soco.2021.39.3.366>
- Bruce V. Changing faces: Visual and non-visual coding processes in face recognition. *Br J Psychol*. 1982;73(1):105–116. <https://doi.org/10.1111/j.2044-8295.1982.tb01795.x>
- Wells GL, Rydell SM, Seelau EP. The selection of distractors for eyewitness lineups. *J Appl Psychol*. 1993;78(5):835–844. <https://doi.org/10.1037/0021-9010.78.5.835>
- Fitzgerald R, Tredoux CG, Juncu S. Estimation of eyewitness identification error rates in fair and biased lineups. Paper presented at: American Psychology and Law Society Annual Meeting; 2022 Mar 16–20; Denver, CO, USA.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
- R Core Team. R: A language and environment for statistical computing [software on the Internet]. Vienna: R project; 2018. Available from: <https://www.R-project.org/>
- Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J Exp Psychol Gen*. 2014;143(5):2020–2045. <https://doi.org/10.1037/xge0000014>
- Wittwer T. The own group bias in face processing: The effect of training on recognition performance [dissertation]. Cape Town: University of Cape Town; 2020.
- Wright DB, Boyd CE, Tredoux CG. Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Appl Cogn Psychol*. 2003;17(3):365–373. <https://doi.org/10.1002/acp.898>
- Malpass RS. An excursion into utilitarian analysis. *Behav Sci Res*. 1990;24(1–4):1–15. <https://doi.org/10.1177/106939719002400101>
- Valentine T. Face-space models of face recognition. In: Wenger MJ, Townsend JT, editors. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. New York: Lawrence Erlbaum Associates Publishers; 2001. p. 83–113.
- Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q J Exp Psychol A*. 1991;43(2):161–204. <https://doi.org/10.1080/14640749108400966>
- Natu V, O'Toole AJ. Neural perspectives on the other-race effect. *Vis Cogn*. 2013;21(9–10):1081–1095. <https://doi.org/10.1080/13506285.2013.811455>
- Herzmann G, Willenbockel V, Tanaka JW, Curran T. The neural correlates of memory encoding and recognition for own-race and other-race faces. *Neuropsychologia*. 2011;49(11):3103–3115. <https://doi.org/10.1016/j.neuropsychologia.2011.07.019>
- Stahl J, Wiese H, Schweinberger SR. Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia*. 2010;48(7):2027–2040. <https://doi.org/10.1016/j.neuropsychologia.2010.03.024>
- Halberstadt AG, Cooke AN, Garner PW, Hughes SA, Oertwig D, Neupert SD. Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion*. 2022;22(3):403–417. <https://doi.org/10.1037/emo0000756>
- Dehon H, Brédart S. An 'other-race' effect in age estimation from faces. *Perception*. 2001;30(9):1107–1113. <https://doi.org/10.1068/p3122>
- Davis D, Loftus EF. Eyewitness science in the 21st century. In: Wixted J, editor. *Stevens' handbook of experimental psychology and cognitive neuroscience*. 4th ed. Hoboken, NJ: John Wiley & Sons; 2018. p. 1–38. <https://doi.org/10.1002/9781119170174.epcn116>
- Hills PJ, Lewis MB. Reducing the own-race bias in face recognition by shifting attention. *Q J Exp Psychol (Hove)*. 2006;59(6):996–1002. <https://doi.org/10.1080/17470210600654750>
- Wittwer T, Tredoux CG, Py J, Paubel PV. Training participants to focus on critical facial features does not decrease own-group bias. *Front Psychol*. 2019;10:2081. <https://doi.org/10.3389/fpsyg.2019.02081>
- Fisher RP, Geiselman RE. *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C Thomas Publisher; 1992.
- Demarchi S, Py J. A method to enhance person description: A field study. In: Bull R, Valentine T, Williamson T, editors. *Handbook of psychology of investigative interviewing: Current developments and future directions*. Chichester: Wiley-Blackwell; 2009. p. 241–256. <https://doi.org/10.1002/9780470747599.ch14>
- Memon A, Meissner CA, Fraser J. The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychol Public Policy Law*. 2010;16(4):340–372. <https://doi.org/10.1037/a0020518>