An Experimental Manipulation of Contact and Motivation in the Cross-Race Effect

Tristan Rayner

Supervisor: Colin Tredoux

Department of Psychology

University of Cape Town

Word Count: Body: 9994

Abstract: 224

PLAGIARISM DECLARATION

PLAGIARISM

This means that you present substantial portions or elements of another's work, ideas or data as your own, even if the original author is cited occasionally. A signed photocopy or other copy of the Declaration below must accompany every piece of work that you hand in.

DECLARATION

1. I know that Plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the American Psychological Association formatting for citation and referencing. Each significant contribution to, and quotation in, this essay/report/project from the work or works, of other people has been attributed, cited and referenced.

3. This essay/report/project is my own work.

4. I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

NAME:	**Tristan Rayner**
SIGNATURE:	**T.R.**
STUDENT NUMBER:	**RYNTRI001**
SUPERVISOR:	**Colin Tredoux**

Abstract

The cross-race effect (CRE) refers to a bias in recognising, and differentiating between, own-race faces as opposed to cross-race faces. The amount of contact one has with a particular group has long been thought to affect the CRE. However, the methods used to measure the amount of contact with cross-race faces are problematic, and contact has rarely been experimentally manipulated in studies. Recent studies have identified motivation as a key factor in the CRE. Consequently, an experimental manipulation of contact and motivation was conducted on Black, South African students (*N* = 20) from the University of Cape Town. The study aimed to determine if (1) amount of contact affected the CRE, (2) if motivation to identify faces affected the CRE, and (3) if the interaction between contact and motivation affected the CRE. Participants attended two sessions a week, for four weeks, which used practice tasks to increase their contact with Egyptian faces (a group with which they had little contact) and facial recognition tests to assess their performance. Participants were measured on their discrimination accuracy (*d'*), response criterion (c), and reaction times. Three mixed design ANOVAs were run on these measures. The CRE was observed in the study, however it showed no significant change as the amount of contact increased. Similarly, neither motivation, nor the interaction between motivation and contact, significantly changed the CRE.

Keywords: cross-race effect; faces; facial recognition; motivation; contact; facial recognition training

Table of Contents

An Experimental Manipulation of Contact and Motivation in the Cross-Race Effect

The cross-race effect (CRE) is one of the most robust findings within the face recognition literature (Meissner & Brigham, 2001). The CRE refers to a bias in recognising, and differentiating between, own-race faces as opposed to cross-race faces (Appendix A; Malpass & Kravitz, 1969; Meissner & Brigham, 2001; Sporer, 2001). This finding has been replicated across participant populations (Meissner, Brigham, & Butz, 2005; Sporer, 2001; Wan, Crookes, Reynolds, Irons, & Mckone, 2015; Wright, Boyd, & Tredoux, 2003) and research paradigms (Meissner & Brigham, 2001). However, there is no clear consensus on a primary underlying mechanism for the CRE (Young, Hugenberg, Bernstein, & Sacco, 2012). Identifying this mechanism remains one of the main challenges for researchers. It is especially important, considering that the CRE can have serious legal consequences (Brigham, Brooke-Bennett, Meissner, & Mitchell, 2007).

For example, in the U.S., over 75% of wrongful criminal convictions involve eyewitness misidentification, with many involving cross-race identifications (The Innocence Project, n.d.). Consequently, there is a pressing need for a unifying theoretical model of the CRE that is able to clarify the kinds of contexts in which the CRE is most pronounced. This would be greatly beneficial in understanding eyewitness misidentifications and minimising wrongful convictions. Numerous models of the CRE have been proposed but no single model has been agreed upon (Young et al., 2012). These models are broadly divided into three groups - perceptual expertise, social cognitive and hybrid models - based on their theoretical perspective (Young et al., 2012).

Perceptual Expertise Models

These models propose that increasing contact with a particular face type results in greater perceptual expertise in recognising, and distinguishing, between faces of that type (Malpass & Kravitz, 1969). A lack of contact with cross-race faces results in diminished perceptual expertise for those faces which, in turn, causes deficient cross-race face encoding and recognition - resulting in the CRE (Young et al., 2012).

The differential experience hypothesis. This hypothesis proposes that the amount of contact with a face type determines how proficiently a face is processed. In general, someone will have high levels of contact with own-race faces (due to family or community settings) and comparatively low levels of contact with cross-race faces. This is proposed to be the cause of the CRE; for example, Walker and Tanaka (2003) demonstrated that participants exhibited a bias in discriminating own-race faces over cross-race faces in a same/different matching task. Contact is theorised to improve mechanisms that aid in face processing, such as generating appropriate recognition and encoding strategies (Meissner & Brigham, 2001). Such mechanisms may fail to develop with low levels of contact, as is often the case with cross-race faces. The differing amounts of contact can arise from structural or implicit racial segregation that occurs within, and across, many countries. For example, Rhodes et al. (2009) found that time living in a Western country negatively predicted the size of the CRE in Asian participants viewing White faces. Therefore, the CRE arises from the differential exposure to own-race faces, compared to cross-race faces (Malpass & Kravitz, 1969).

The differential processing hypothesis. A second hypothesis proposes that the manner in which a face is processed is determined by the expertise in that face type; leading to differential processing of own-race and cross-race faces (Young et al., 2012). Rhodes, Brake, Taylor, and Tan (1989) showed that inverted faces impair participants' reaction times and recognition rates in own-race faces, but not cross-race faces. Inversion inhibits the ability to process faces in terms of their configural information, which is the process used when there is expertise in a face type (Logothetis & Sheinberg, 1996). Therefore, faces for which we have expertise (i.e. own-race faces) are likely processed in a configural manner, whereas face for which we do not have expertise (i.e. cross-race faces) are likely processed in a feature-based manner (Rhodes et al., 1989; Michel, Rossion, Han, Chung, & Caldara, 2006). This is supported by the fact that misaligning the top-and-bottom halves of faces impair recognition for own-race faces more than cross-race faces (Appendix B; Michel et al., 2006). The configural processing of own-race faces is argued to be a more efficient encoding strategy than the feature-based processing of cross-race faces, resulting in the CRE (Hugenberg, Young, Bernstein, & Sacco, 2010).

Limitations of perceptual expertise models. A meta-analysis of the CRE found that self-rated interracial contact only accounts for 2% of the variability across the literature (Meissner & Brigham, 2001). Contact is clearly important in face recognition, as learning a new face type requires many examples of faces within that type. However, the research indicates that more contact with cross-race faces, in isolation, is unlikely to eliminate the CRE (Young et al., 2012). In addition, perceptual expertise models are of questionable validity as they cannot account for recent findings on specific aspects of the CRE; such as the effects of motivation on face recognition, or non-race based face recognition biases (Hugenberg et al., 2010; Young et al., 2012).

Social Cognitive Models

These models emphasise the *kinds* of contact, or the contexts in which faces are processed, rather than the amount (Levin, 2000). That is, contact needs to be *meaningful*, in that it has to encourage an individual to actively attend to, and learn, faces with which they are in contact. These models hypothesise that faces are initially coded on social categories, which determines how a face is processed (Hugenberg et al., 2010).

The feature-selection model. This model posits that face recognition involves a gating process, where faces are automatically coded for race, before other factors that specify identity, like gender or age (Levin, 2000). Faces determined to be cross-race are searched for features that are common to all members of that race; these common features are called category-specifying features (Young et al., 2012). Alternatively, own-race faces are searched for features that can be used to distinguish individuals within a race; these distinguishing features are called identity-specifying features (Levin, 2000). In line with this hypothesis, Ge et al. (2009) showed that cross-race faces are recognised slower, and less accurately, than own-race faces, but are more rapidly categorised by race. In addition, solely encoding the category-specifying features of a face results in difficulty distinguishing between faces of that type (Young et al., 2012). As the CRE arises in cases where individuals are required to be distinguished from one another, cross-race faces are less likely to be correctly identified.

Limitations of social cognitive models. Similar to the perceptual expertise models, social cognitive models cannot account for the variety of findings on the CRE (Young et al., 2012). Specifically, the feature-selection model struggles to account for biases in recognition beyond race-based ones, such as the well-established ‘cross-age’ effect (Rodin, 1987; Young et al., 2012).

Hybrid Models

These models attempt to combine the social cognitive and perceptual expertise aspects of the CRE. They are highly integrative in nature, and seek to incorporate the theories outlined above.

The in-group/out-group model (IOM). The IOM argues that the CRE is not race-based, but rather based on perceived in-groups and out-groups (Sporer, 2001). Like the feature-selection model, faces are automatically scanned for category-specifying information. This information is then used to judge whether the face belongs to an in-group, or an out-group. In-group faces are argued to elicit configural processing, whereas out-group faces elicit categorical processing (Ge et al., 2009; Sporer, 2001). Categorical processing may simultaneously cue cognitive disregard, the tendency to dismiss ‘irrelevant’ stimuli, resulting in shallow encoding (Rodin, 1987; Sporer, 2001). Hehman, Mania, and Gaertner (2010) demonstrated how a superficial in-group/out-group distinction, university affiliation, can generate a comparable effect to the CRE. The CRE, therefore, results from categorising a cross-race face as belonging to an out-group.

The Categorisation-Individuation Model (CIM). The CIM argues that motivation is key in face processing and that the CRE arises from three interacting factors: category activation, perceiver motivation, and perceptual expertise (Hugenberg et al., 2010). Category activation is argued to occur automatically for cross-race faces; however, motivation can arrest categorisation, in favour of individuation. Therefore, the CRE arises from the categorical processing of cross-race faces, but can be diminished if there is motivation to process that face. In line with this prediction, Hugenberg, Miller, and Claypool (2007) showed that simply informing participants of the CRE, and asking them to try hard to learn cross-race faces resulted in a lowered CRE (Appendix C). Additionally, this is supported by the in-group/out-group recognition biases mentioned above (Hehman et al., 2010).

Limitations of hybrid models. There is a dearth of studies that have extensively tested the hypotheses generated by these models. In addition, there is almost no evidence examining the cross-cultural applicability of their models. Wan et al. (2015) found that informing Asian participants of the CRE, in an Australian context, did not have any significant effect on the CRE. This contradicts Hugenberg et al.'s (2010) findings and may indicate that CIM is only applicable to certain contexts.

Measures of Contact

As contact is central to the CRE, methods used to assess the amount of contact with own-race and cross-race faces will invariably affect any results obtained. Therefore, a measure of contact should be as accurate as possible; this is not the case in the CRE literature. There are two methods of measuring contact: first, it can be inferred from demographic data of participants. For example, Walker and Tanaka (2003) used White students from Ohio, U.S. and compared them to Asian students from Canada. There was an implicit inference that the White students had different amounts of contact with White and Asian faces, than the Asian students had. The second method of measuring of contact uses individuals' self-reported amounts of contact; for example, some studies have used self-report questionnaires to assess contact (Hancock & Rhodes, 2008; Wan et al., 2015). These questionnaires include around fifteen Likert-type items (such as, 'I live in an area where I interact with Caucasian people') from which they generate an average score of contact (Wan et al., 2015). This method attempts to quantitatively measure contact; however, the validity and reliability of these questionnaires have never been tested. Therefore, it is possible that these measures are not giving accurate estimations of contact, or give inconsistent estimations of contact.

Consequently, in a meta-analysis of 39 research articles, it was suggested that the "seemingly weak relationship between self-rated contact and the [CRE] may be due to limitations in the range of variability present in [measures of contact]" (Meissner & Brigham, 2001, p. 17). A preliminary observation of the studies cited thus far shows that 78% of studies used the demographic inference method while 22% used the self-reported contact method. Both of these approaches are problematic, as the first method uses inference to measure contact and does not necessarily reflect the actual amount of contact individuals have had. On the other hand, the second method measures contact through self-report questionnaires, which may give inaccurate, or inconsistent measures of contact.

Alternative measures. Some studies have used novel measures of contact, such as Li, Dunning, and Malpass (1998 as cited in Meissner & Brigham, 2001), who showed that, within a White population, 'basketball fans' (a sport in which the majority of players, in the U.S., are Black) exhibited a lower CRE for identifying Black faces, than 'basketball novices' did. This method is flawed in that it assumes that one group has more contact with cross-race faces, but does not quantitatively measure this contact. Seutloali (2014) used population data on the towns in which participants have spent most of their lives to infer contact. However, this method assumes that there is a direct relationship between the proportion of racial groups within a town, and the amount an individual has interacted with these groups. This is not always the case, as individuals, through structural or implicit segregation, will differentially interact with different groups. This method cannot account for this differential interaction and, therefore, generates an inaccurate measure of contact.

Facial recognition training. A number of early articles studied whether it was possible to increase a person's ability to recognise faces (Goldstein & Chance, 1985; Malpass, Lavigueur, & Weldon, 1973; Woodhead, Baddeley, & Simmonds, 1979). These studies generated mixed results: some found that training had no effect on facial recognition abilities (Woodhead et al., 1979); while others found that training could significantly improve them (Goldstein & Chance, 1985; Malpass et al., 1973). Malpass (1981) argued that training can bring about short-term improvements for cross-race faces, but no improvements for own-race faces.

Conclusion

In summary, the three main models of the CRE have been presented and shown to use two methods of assessing contact: via inference from demographic details, or through self-report measures. Weaknesses of all three models have been discussed and may be partly attributable to the lack of a robust measure of contact. Previous research has shown that facial recognition training has had mixed results; however, this has never been tested in the CRE paradigm. It is, therefore, imperative to design a study that can accurately measure contact; this could be achieved with an experimental manipulation of contact. Such a manipulation would result in a stronger experimental design, and more informative results.

Research Aim and Questions

The present study manipulated the amount of contact with cross-race faces, in order to discern whether this affected the CRE, if at all. The amount of contact that Black, South Africans had with male Egyptian faces (two populations that had little contact with one another) was manipulated over time. As such this study posed three questions:

1. Did the amount of contact with cross-race faces affect the CRE?
2. Did motivation to individuate cross-race faces affect the CRE?
3. Did the interaction between contact and motivation affect the CRE?

Method

Design and Setting

This study used a 2 x 8 mixed design. The between-subject variable was whether participants were motivated to individuate faces while the within-subject variables were the amount of contact participants had with the faces, and the race of the faces:

		Pre-Test	**Test 1**	**Test 2**	**Test 3**	**Test 4**	**Test 5**	**Test 6**	**Test 7**
Motivation	Black	B_{MP}	B_{M1}	B_{M2}	B_{M3}	B_{M4}	B_{M5}	B_{M6}	B_{M7}
	Egyptian	E_{MP}	E_{M1}	E_{M2}	E_{M3}	E_{M4}	E_{M5}	E_{M6}	E_{M7}
Control	Black	B_{CP}	B_{C1}	B_{C2}	B_{C3}	B_{C4}	B_{C5}	B_{C6}	B_{C7}
	Egyptian	E_{CP}	E_{C1}	E_{C2}	E_{C3}	E_{C4}	E_{C5}	E_{C6}	E_{C7}

Time (Amount of Contact)

Participants were scored for each trial answered correctly in the practice tasks, which were calculated into a total score for each session. These scores were displayed in the form of ‘leaderboards’, which ranked participants’ total scores from highest to lowest (Appendix D). At the beginning of each session, excluding the first session, the researcher gave these leaderboards to the participants in the motivation condition. Leaderboards are commonly used in games to elicit motivation and are especially effective when participants are competitive (Zichermann & Cunningham, 2011). Therefore, the leaderboards motivated participants to try increase their scores, as well as surpass other participants’ scores. The control condition was not assigned points, nor shown a leaderboard.

Contact was operationalised through seven training sessions over four weeks. Each session contained three practice tasks, which, in total, lasted approximately 45 minutes. These practice tasks were designed to improve participants’ ability to recognise, and distinguish between Egyptian faces. Contact steadily increased as participants viewed and were required to identify, or distinguish between, approximately 160 Egyptian faces each session.

In addition, each session contained a facial recognition test, which assessed participants’ facial recognition abilities for Black and Egyptian faces. The first session’s test served as a pre-test (when participants had little-to-no contact with Egyptian faces). The rest of the tests assessed participants’ performance after each training session.

The study took place in the ACSENT Lab, in the Department of Psychology at the University of Cape Town which provided a quiet, controlled environment and minimised distractions from the tasks.

Participants

Sample size calculations. An a priori sample size was estimated for a mixed design ANOVA analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). Assuming *a* =

.05, an effect size of *f* = .22, a target power of .80, two groups, eight measurements and a correlation between the same measure over two different times of *r* = .40, twenty-four participants, in total, were needed for the study. The effect size was obtained from Meissner and Brigham's (2001) meta-analysis of 39 research articles on the CRE. A correlation of the same measure over two different times was estimated from Slone, Brigham and Meissner (2000).

Recruitment. Participants were recruited via a campus-wide email and word of mouth (Appendix E). Half of the participants were randomly assigned to the motivation condition, with the other half were assigned to the control condition. Students were rewarded by being entered into a raffle to win a monetary prize.

Inclusion criteria. All participants were required to be South African and members of the Black racial group. Although cross-gender recognition biases have been shown to exist, these effects are relatively minor (Levin, 2000). Therefore, the study included both male (*n* = 9) and female participants (*n* = 11). Due to issues surrounding the definition of race (see Posel, 2001), participants were those who self-reported as belonging to the Black racial group. One participant was excluded from the analysis, as they self-reported as belonging to the Indian racial group. Finally, participants had to have little-to-no experience with Egyptian populations. This was assessed using a short questionnaire (Appendix H), adapted from Wan et al. (2015).

Apparatus

Participants were tested on computers which have E-prime installed. E-prime has a high degree of timing precision, can present stimuli, and can record participants' responses. Each computer also had three practice tasks installed: FaceGame (Daniels, Mackier, & van Helsdingen, 2013), a face attribute task, and a memory matching task.

FaceGame. The first practice task is a computer game in which the participant was required to move around a virtual character on the screen and 'talk' to different non-player characters (NPCs) in the game. These NPCs required the participant find other, distinct NPCs, which presented them with faces to remember; participants were then required to pick the correct face out of a line-up of six other faces. There were two forms of this trial, the first of which required the participant to identify a face that had not been previously learned, out of new faces. The second form required the player to pick a learned face out of new faces. Participants were required to complete 7 trials, which took approximately 15 minutes.

Face attribute task. The second task involved participants generating creative descriptions of faces. Participants were required to create semantic retrieval cues for faces that were presented to them (e.g. ‘He looks like a fast runner’). Participants were encouraged to be as creative as possible in their generated descriptions. Participants were then required to create descriptions for fifteen faces. Thirty faces (fifteen old and fifteen new) were subsequently presented and participants were required to answer if they had seen the face before.

Memory matching task. The third task contained ten pairs of matching faces, with each face on a separate card, which were arranged in a five-by-four matrix, face-down. The aim of the task was to match every pair of faces in the cards, but only two cards could be face-up at any time. Once two faces were matched, they remained face-up; the goal of the task was to get all the cards face-up. Participants were required to complete as many of the matrices as possible in 15 minutes.

For all three practice tasks, faces were randomised and separate from those used in the facial recognition test. This ensured that participants were learning to distinguish within a new population, rather than just learning the stimuli of the tests. A pilot study was conducted, using White participants (viewing White and Egyptian faces), in order to assess how long each practice task took, as well to test the functioning of the practice tasks and facial recognition tests. Several minor corrections to the procedure (outlined below) were made as a result of this pilot study.

Stimuli

The stimuli for this study consisted of 735 images: 160 South African, Black faces and 575 Egyptian faces. All faces were of adult males. Due to the vast number of faces needed, the faces were computer generated using the ID program. ID can generate faces from a population using an eigenface-based mathematical model (Tredoux, Nunez, Oxtoby, & Prag, 2007). All images were front-facing, contained only the head, and were in colour. Images were standardised for size and position.

Procedure

Each experimental session contained a maximum of seven participants and lasted approximately 60 minutes. In the first session, participants were given three forms to complete: an informed consent form (Appendix F); a demographic form (Appendix G); and a form ensuring that they have minimal contact with Egyptian populations (Appendix H).

Participants, in the motivation condition, were each shown their leaderboard as they arrived. This indicated how well they performed in the previous session, compared to all the other participants. The leaderboard was personalised to indicate a given participant’s score, while replacing all other participants’ names with a dash. This ensured anonymity of participants' scores, and enabled them to share scores only if they wished to. Leaderboards were given in every session except the first session, as participants had no previous scores on which they could be ranked.

Participants were then presented with an instruction slide that indicated what was required of them in the test and which keys to press to continue the task. Three practice trials commenced before the test in order to ensure that they understood the task. The test began with an encoding phase, which consisted of presenting 20 faces: 10 Black and 10 Egyptian. In this phase, participants were presented with the faces, one at a time, in a randomised order. Each face was presented for 3 seconds, with 1.5 seconds between each face. After seeing all 20 faces, participants were given a word-search task to complete, as a distractor.

Next, in the recognition stage, participants were presented with 40 faces, 20 of which are those they already studied ('old' faces) and 20 of which have not been studied ('new' faces). The faces were in a randomised order for each participant. Participants were required to press '1' on the keyboard if the face is old and '0' if the face is new. The next face was presented only once the participant answered. Once all 40 faces were viewed, the participants began the practice tasks.

Subsequent sessions repeated the above procedure; however, in the recognition stage, an additional 10 'previous faces' (5 Black, and 5 Egyptian) were added. These were faces randomly selected out of those which the participant had encoded in previous sessions. This allowed assessment of how participants learned new faces as well as how well they remembered faces learned previously.

After the test had finished, participants were given each of the three practice tasks: FaceGame, the face attribute task, and the memory matching task. Each of these tasks took approximately 15 minutes, with participants switching to the next task once they had finished the task, or the allotted time for the task was over. The tasks were randomised using a Latin square design. When participants had completed each task, they were free to leave. This ensured that each participant had an hour's worth of contact (15 minutes for the facial recognition test and 45 minutes of practice tasks) with the Egyptian faces.

Participants were required to attend two sessions a week, for four weeks. The first session's test scores (pre-test) served as the control for both motivation and amount of contact. After seven sessions, the final (eighth) session only consisted of the facial recognition test and no subsequent practice tasks. The experiment terminated after the final session, and data was collected and analysed.

Statistical Analysis

Data was analysed using the SPSS statistical software package. Signal Detection Theory (Macmillan & Creelman, 2005) was used to obtain measures of discrimination accuracy (*d*') and response criterion (*c*). Mixed design ANOVAs were conducted on three measures commonly found to be affected by the CRE: discrimination accuracy, response criterion, and reaction time (Meissner & Brigham, 2001). Descriptive statistics were computed alongside the analyses, along with an examination of the assumptions of ANOVA.

Ethical Considerations

This study received ethical approval from the Ethics Committee of the Department of Psychology at the University of Cape Town (Appendix I).

Consent. A signed, informed consent form (Appendix F) which contained information about the study was collected from all participants. Any questions participants had, were answered by the researcher. The researcher's email address was provided, if the participants had any questions outside of the sessions.

Voluntary Participation. Participants were informed that their participation in the study was entirely voluntary, and they were able to withdraw from the study at any time. Participants who did not complete all eight sessions were not entered into the raffle draw.

Confidentiality. All details of participants, as well as their scores, were kept confidential. The participants were kept anonymous in all analyses and reporting of the data.

Benefit. All participants were entered for a raffle, with the chance to win either R1500, R750, R500 or R250. They also learned about the CRE, and were given any relevant literature, if they wished. Finally, the study contributed to the literature on the CRE by examining the potential of an experimental manipulation of contact.

Harm. There was a minor issue of harm in that participants may have considered that low scores on the practice tasks and/or facial recognition test indicated that they may hold racial biases. This seemed unlikely, and was planned to be dealt with on a case-by-case basis.

Debriefing. At the end of the experiment, all participants were thanked and given an opportunity to ask the researcher any final questions, make comments or give suggestions. After all participants had completed their sessions, each participant was sent an email that thanked them again for their participation and informed them of the results of the raffle.

Results

Hit and False Alarm Rates. Hit rates were obtained by dividing the number of faces a participant correctly identified as 'Old' by the total amount of 'Old' faces seen. Conversely, false alarm rates were obtained by dividing the number of faces a participant incorrectly identified as old, by the total amount of 'New' faces seen. Figure 1 shows the different patterns of hit and false alarm rates for Black faces and Egyptian faces. There is a fairly substantial difference between the hit and false alarm rates for Black faces, with both remaining relatively stable across sessions. On the other hand, there is much less of a difference between the hit and false alarms rates for the Egyptian faces. There is also a fair amount of variation between sessions. Hit rates for Egyptian and Black faces were quite similar, with Egyptian faces being slightly higher. However, this comes at the cost of Egyptian faces having a much higher false alarm rate than the Black faces.

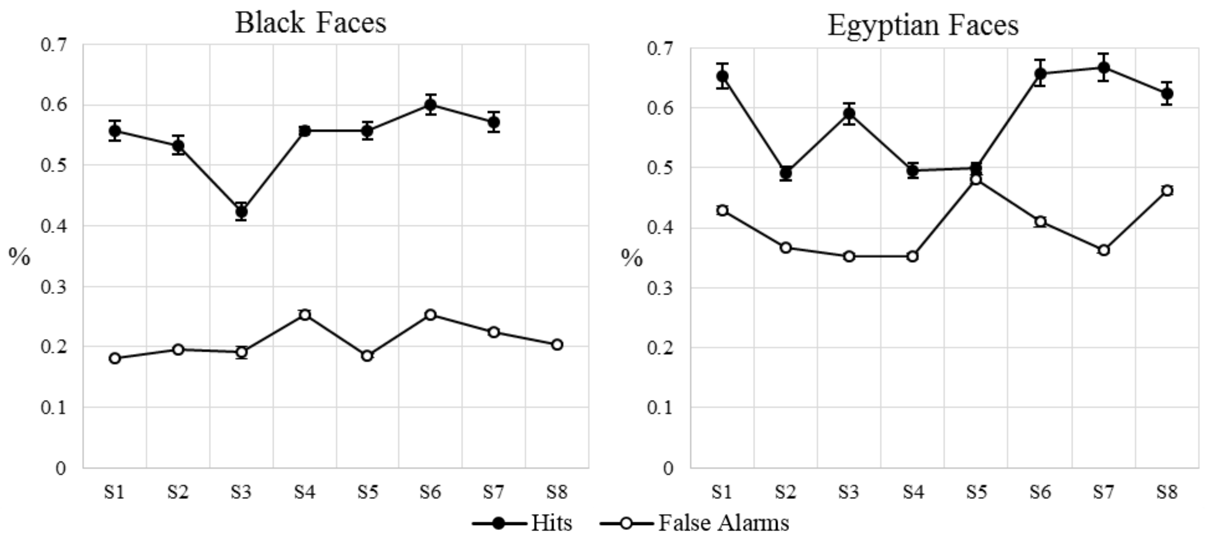


Figure 1. Mean hit and false alarm rates for Black and Egyptian faces, with 95% confidence intervals.

d' Analysis. Discrimination accuracy was calculated in three steps, using Macmillan & Creelman's formula for *d*' (2005). First, the hit rates and the false alarm rates were calculated for each participant, in each session. Next, the hit and false alarm rates were converted into *z*-scores. However, some scores could not be converted as their rates where either 0 or 1, which cannot be computed in the normal distribution. Consequently, Stanislaw & Todorov's (1999) suggested method, to replace scores of 0 (the minimum) with 0.5 ÷ *n* and scores of 10 (the maximum) with (*n* – 0.5) ÷ *n*, was used to calculate these scores. This method allows for scores of 0 or 10 to be computed into *z*-scores. Finally, *d'* was calculated by subtracting the false alarm rate z-score from the hit rate *z*-score.

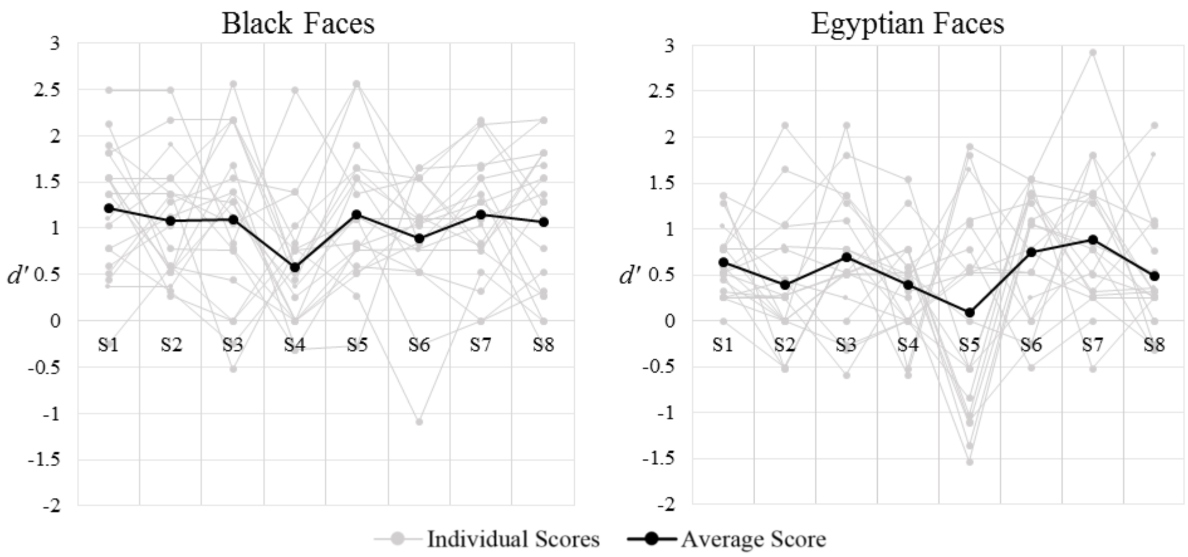


Figure 2. Average, and individuals', discrimination accuracies over sessions.

Figure 2 shows that there was a large amount of variation in individual participants' scores over the eight sessions. Participants, on average, were better at discriminating Black faces than Egyptian faces. In session 5, there was a sharp drop in some participants' scores on Egyptian faces, while many other remain more consistent with the rest of their scores. This will be further investigated later in the analysis.

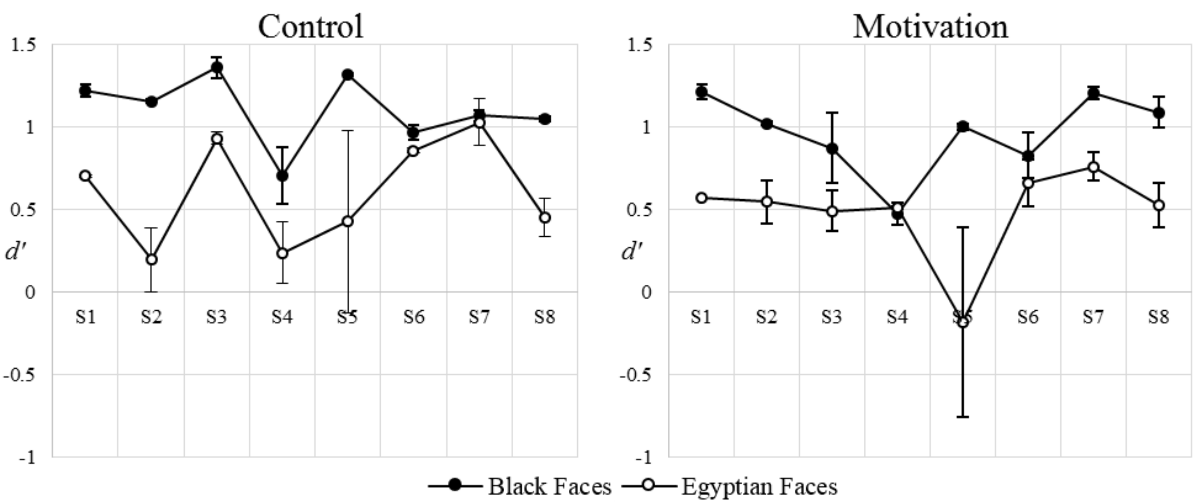


Figure 3. Mean discrimination accuracy across sessions for control and motivation groups, with 95% confidence intervals.

Figure 3 indicates that almost all of the *d'* scores for Black faces were higher than those for Egyptian faces. There was a fair amount of variation in both groups' performance across different sessions. Egyptian faces in session 5 showed a large amount of individual variation for both the control and motivation groups, which is likely due to the large drop in some individuals' scores that was identified above. Table 1 shows that, in each session, there was a moderate amount of individual variation in performance on Black faces and a large amount of individual variation for Egyptian faces, in comparison to their respective means. It also shows that the participants' average performance was better in every session for Black faces than it was for Egyptian faces. In session 5, performance on Egyptian faces had a smaller mean (*M* = .09) and much larger standard deviation (*SD* = 1.09) than any other session. This may indicate some kind of stimulus effect, or other factor that caused performance to fluctuate to the extent that it did. A more fine-grained analysis reveals that, in session 5, form 1 of the test had a significantly lower average performance (*d'* = -0.48) than both form 2 (*d'* = 0.60), and the rest of the sessions. The poor performance on this form of the test cannot be due to an environmental factor, as not all participants took this form of the test on the same day. Therefore, even though the faces and forms of test were randomised, this form happened to be significantly more difficult than the rest of the tests. Consequently, session 5 may be an outlier in participants' performance. This will be further explored later.

Table 1.
Discrimination accuracy over sessions and face race

	Black Faces		Egyptian Faces	
	M	*SD*	*M*	*SD*
Session 1	1.22	.68	.63	.39
Session 2	1.08	.64	.39	.69
Session 3	1.09	.86	.69	.69
Session 4	.58	.67	.39	.53
Session 5	1.14	.72	.09	1.09
Session 6	.89	.66	.75	.68
Session 7	1.14	.65	.88	.79
Session 8	1.07	.77	.49	.65

Boxplots and P-p plots were computed for participants' average discrimination accuracy across the eight sessions, in order to assess whether the assumption of normality was upheld. Participants' performance on Black faces was slightly positively skewed in sessions 1 and 8, while they were slightly negatively skewed in sessions 2 and 5. Session 6 was moderately negatively skewed, and showed the biggest deviation from the normal distribution. There were 2 outliers in the data, in sessions 4 and 6. The P-p plot reveals that most of the points lie close to the normal line, but there are some exceptions. Participants' performance on Black faces had minor deviations in normality; however, ANOVA is fairly robust to such minor deviations, and therefore the analysis will continue.

Participants' performance on Egyptian faces also contained 2 outliers, in sessions 2 and 8. In addition, sessions 1, 2, 4 and 7 were slightly positively skewed, while session 5 was moderately negatively skewed. The P-p plot showed that the many of the points deviated from the normality line. Many of the points lay below the normality line, which indicated that their overall performance was negatively skewed. This suggests that participants' average performance on Egyptian faces may violate the assumption of normality. As previously mentioned, ANOVA is fairly robust to such violations; however, it is important to keep this in mind as this violation may affect the accuracy of the results.

Mauchly's test indicated that both performance across Sessions, $W(7, 27) = .15$, $p = .339$, and the interaction of Sessions and Face Race, $W(7, 27) = .23$, $p = .707$, did not violate the assumption of sphericity.

A mixed design ANOVA indicated that there was a significant main effect for Sessions, $F(7, 27) = 2.66$, $p = .013$, $\eta^2 = .129$, and for Face Race, $F(1, 2) = 42.81$, $p < .001$, $\eta^2 = .704$. On the other hand, the main effect of Group was not significant, $F(1, 2) = .68$, $p = .420$, $\eta^2 = .036$. As there were only 2 different races, the descriptive statistics reveal that participants' performance on Black faces ($M = 1.03$, $SE = .10$) was significantly better than their performance on Egyptian faces ($M = .55$, $SE = .08$). A post-hoc analysis, using Tukey's LSD test, was conducted on Sessions (Table 2), in order to explore where the significance lies within the sessions. This test revealed that sessions 1, 3, 6, and 7 had significantly higher scores than session 4, and that session 7 had significantly higher scores than session 5. Therefore, the majority of significant results were present because of the low discrimination accuracy in session 4, as compared to the rest of the sessions.

Table 2.
Significant differences between sessions.

Sessions		ΔM (a-b)	p	95% Confidence Interval	
a	b			Lower	Upper
1	4	.444	.005	.151	.737
3	4	.432	.024	.064	.800
4	6	-.346	.047	-.686	-.005
4	7	-.533	.014	-.946	-.120
5	7	-.374	.025	-.677	-.051

In terms of interaction effects, Sessions*Group, $F(7, 27) = 1.11$, $p = .364$, $\eta^2 = .058$; Face Race*Group, $F(1, 2) = .03$, $p = .873$, $\eta^2 = .001$; and Sessions*Face Race*Group, $F(7, 27) = .70$, $p = .672$, $\eta^2 = .037$ were all non-significant. However, Sessions*Face Race, $F(7, 27) = 2.32$, $p = .029$, $\eta^2 = .114$, was significant. This interaction will be explored below.

Figure 4 shows that the pattern of performance on Black and Egyptian faces more or less mimic each other, with the exception of Session 5, where there is an ordinal interaction. However, session 5 was identified as problematic earlier, due to its large variation. Consequently, the analysis was rerun, excluding session 5 for both Black and Egyptian faces. In the new analysis, Face Race, $F(1, 2) = 25.92$, $p < .001$, $\eta^2 = .590$; and Sessions, $F(7, 27) = 3.13$, $p = .007$, $\eta^2 = .148$, remained significant. However, the interaction between Face Race and Sessions, $F(7, 27) = 1.54$, $p = .171$, $\eta^2 = .079$, was no longer significant. Consequently, the significant interaction between Face Race and Sessions found in the first analysis is likely just a product of the problematic data in session 5.

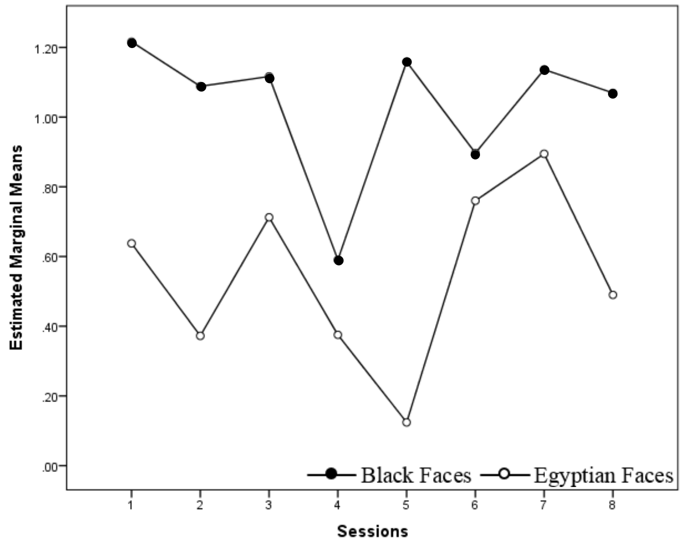


Figure 4. Interaction of Sessions and Face Race.

A second mixed design ANOVA was run, using just Egyptian faces. This was done due to the fact that participants were not expected to improve their facial recognition for own-race (Black) faces. Consequently, this analysis was conducted in order to reveal whether, in the absence of Black faces from the analysis, participants improved their scores on Egyptian faces over the sessions. Mauchly's test indicated that performance across Sessions, $W(7, 27) = .15$, $p = .347$, did not violate the assumption of sphericity.

A mixed design ANOVA indicated that there was a significant main effect for Sessions, $F(7, 27) = 2.81$, $p = .010$, $\eta^2 = .135$. Like the previous analysis, the main effect of Group was still not significant, $F(1, 2) = .59$, $p = .452$, $\eta^2 = .042$. A post-hoc analysis, using Tukey's LSDs test, was conducted on Sessions (Table 3), in order to explore where the significance lies within the sessions. This test revealed that session 7 had significantly higher discrimination accuracy than sessions 2, 4, and 5; session 3 was significantly higher than session 5; and session 6 was significantly higher than session 4.

Table 3.
Significant differences between sessions.

Sessions		ΔM (a-b)	p	95% Confidence Interval	
a	b			Lower	Upper
2	7	-.522	.020	-.950	-.094
3	5	.588	.033	.052	1.125
4	6	-.385	.042	-.754	-.016
4	7	-.519	.039	-1.009	-.030
5	7	-.771	.013	-1.355	-.186

In terms of the interaction effect, Sessions*Group, $F(7, 27) = 1.24$, $p = .287$, $\eta^2 = .064$, was not significant. The fact that two later sessions were significantly higher than earlier sessions, provides weak evidence that discrimination accuracy may have improved. On the other hand, session 3 was significantly higher than session 5, which indicates that this may not be the case.

***c* Analysis.** Response criterion was calculated by adding the *z*-scores of the hit rate to the *z*-score of the false alarm rate, and then multiplying this number by -.05. This measure indicates how conservative/liberal participants were in their responses. That is, scores of 0 indicated that participants were unbiased in their decisions; positive scores indicated that participants were more likely to say that they had not seen the face before (i.e. conservative); and negative scores indicated that they were more likely to say that had seen the face before (i.e. liberal).

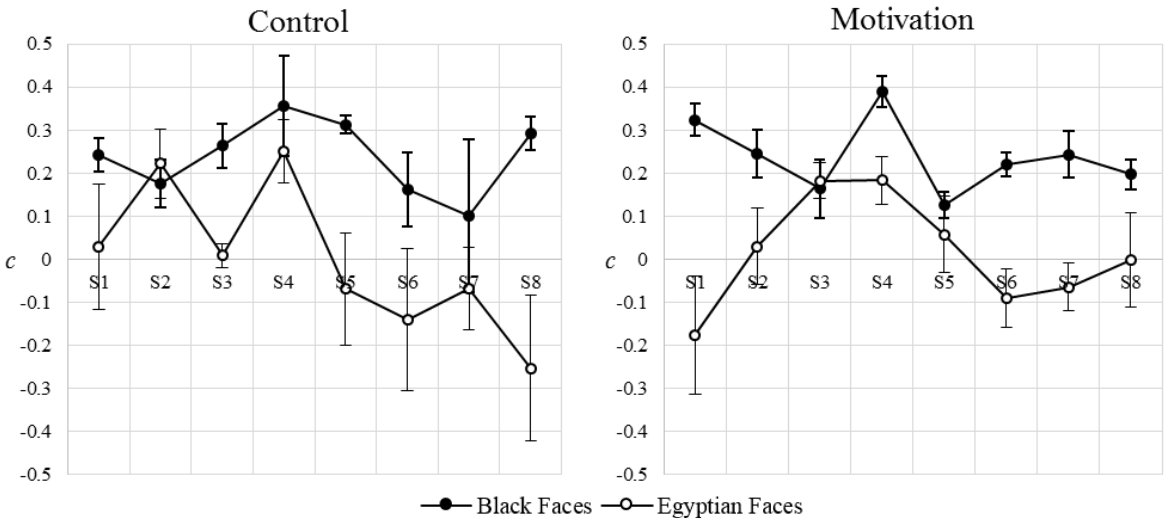


Figure 5. Mean response criterion across sessions for control and motivation groups, with 95% confidence intervals.

Figure 5 shows that participants were conservative in identifying Black faces across all of the sessions. On the other hand, for Egyptian faces, participants' response criterion varied between liberal and conservative. There was a large amount of variation across sessions for Egyptian faces, while there was much less variation for Black faces. In almost all of the sessions, there was a change in the response criterion (*criterion shift*) between Black and Egyptian faces. The criterion shift indicated that participants were more conservative in judging whether they had previously seen a Black face than an Egyptian face. Table 4 indicates that there was a large amount of variation within each session for both Black and Egyptian faces, as compared to the means of each session.

Table 4.
Response criterion over sessions and face race

	Black Faces		Egyptian Faces	
	M	*SD*	*M*	*SD*
Session 1	.29	.37	-.08	.50
Session 2	.21	.39	.12	.44
Session 3	.21	.40	.10	.30
Session 4	.38	.48	.21	.41
Session 5	.21	.31	.01	.45
Session 6	.19	.37	-.11	.41
Session 7	.18	.50	-.07	.35
Session 8	.24	.34	-.11	.48

Boxplots and P-p plots were computed for participants' average response criterion across the eight sessions, in order to assess whether the assumption of normality was upheld. Participants' response criterions in sessions 2 and 6 were slightly negatively skewed, while slightly positively skewed in session 8. In addition, session 3 was moderately positively skewed. The P-p plot showed that many of the points lay above the normal line; which indicated that the faces, in general, were slightly positively skewed. That being said, most of the sessions had little skewness, or were roughly normally distributed, and there were no outliers. ANOVA is fairly robust to these small deviations and, therefore, the analysis will continue.

Participants' response criterion for Egyptian faces contained a number of outliers across sessions 1, 2, 3, 5, and 6. Sessions 1 and 8 were slightly positively skewed, while session 5 was slightly negatively skewed. The P-p plot showed that the majority of points lay on the normality line, which indicated that the overall performance was normally distributed. Therefore, participants' response criterions for Egyptian faces were roughly normally distributed, but should be treated with caution, due to the number of outliers in the sessions. ANOVA is fairly robust to outliers, but it is important to keep them in mind, as it may affect the accuracy of the results. Therefore, the analysis will continue.

Mauchly's test indicated that performance across sessions, $W(7, 27) = .10$, $p = .136$, upheld the assumption of sphericity. However, the interaction of Sessions and Face Race, $W(7, 27) = .07$, $p = .047$, was significant, and therefore violated the assumption of sphericity. Consequently, the Greenhouse-Geisser correction will be used in the analysis of the interaction between Sessions and Face Race.

A mixed design ANOVA indicated that there was a significant main effect of Sessions, $F(7, 27) = 2.17$, $p = .042$, $\eta^2 = .107$, and Face Race, $F(1, 2) = 17.793$, $p = .001$, $\eta^2 = .497$. On the other hand, the main effect of Group was not significant, $F(1, 2) = .009$, $p =$

.925, η^2 = .001. As there were only 2 different races, the descriptive statistics reveal that there was a significant criterion shift between face race, where participants were more conservative for Black faces (*M* = .24, *SE* = .05) than Egyptian faces (*M* = .01, *SE* = .06). A post-hoc analysis, using Tukey's LSD test, was conducted on Sessions (Table 5) in order to explore where the significance lies within the sessions. Table 5 indicates that participants had significantly higher response criterions in session 4 than sessions 1, 5, 6, 7, and 8. This indicates that the low discrimination accuracy previously identified in session 4 was likely due to an increased response criterion in participants.

Table 5.
Significant differences between sessions.

Sessions		ΔM (a-b)	*p*	95% Confidence Interval	
a	b			Lower	Upper
1	4	-.190	.020	.033	.348
4	5	.189	.041	.009	.368
4	6	.257	< .001	-.137	.377
4	7	.243	.005	-.081	.404
4	8	.236	.011	-.062	.411

In terms of interaction effects, Sessions*Group, $F(7, 27) = .259$, $p = .968$, $\eta^2 = .014$; Face Race*Group, $F(1, 2) = .024$, $p = .879$, $\eta^2 = .001$; Sessions*Face Race, $F(4.42, 27) = .885$, $p = .485$, $\eta^2 = .047$; and Sessions*Face Race*Group, $F(7, 27) = 1.28$, $p = .267$, $\eta^2 = .066$ were all non-significant.

Reaction Time Analysis. Reaction time (RT), or the time between onset of a face and a participant's response to that face, was recorded in milliseconds. An average RT was calculated for participants' scores in each session, for each face race. The scores were then converted into seconds. Table 6 shows that, for both Black and Egyptian faces, session 1 had the longest *RT*s as well as the largest variation within a session. The times, in general, decrease in length and variation over the sessions, with session 8 having a lower mean and standard deviation then session 1, for both Black and Egyptian faces.

Table 6.
Reaction time (in seconds) over sessions and face race

	Black Faces		Egyptian Faces	
	M	*SD*	*M*	*SD*
Session 1	2.42	1.36	2.35	1.54
Session 2	1.91	.70	2.20	1.03
Session 3	1.92	.57	1.92	.58
Session 4	1.66	.51	1.85	.66
Session 5	1.81	.52	1.74	.56
Session 6	1.62	.42	1.52	.63
Session 7	1.79	.87	1.58	.36
Session 8	1.67	.45	1.67	.65

Figure 6 indicates that for both control and motivation groups, there was a reduction in the time it took participants to respond over the eight sessions. The highest average response times were in session 1 and gradually decreased and the stabilised around 1.7 seconds. The control group showed a slight difference between the patterns of their *RT*s on Black faces as opposed to Egyptian faces. On the other hand, the motivation group had very similar patterns of *RT*s for both Black and Egyptian faces.

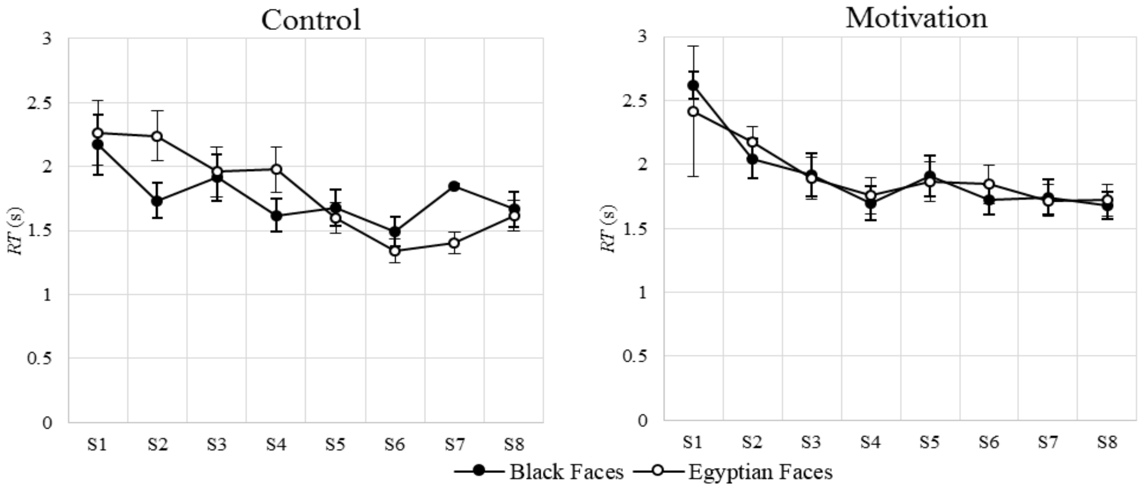


Figure 6. Reaction time across sessions for the control and motivation groups, with 95% confidence intervals.

Boxplots and P-p plots were computed for participants' average *RTs* across the eight sessions. A visual inspection of the boxplot for Black faces showed that nearly all of the sessions were largely positively skewed, with the exception of session 6, which was slightly negatively skewed, and session 5, which was roughly normally distributed. There were also several outliers in the data, particularly for session 8. The P-p plot also revealed that, overall, the participants' scores did not follow the normal line. This indicates that participants' performance on Black faces likely violates the assumption of normality.

Participants' *RT*s for Egyptian faces also contained a number of outliers across sessions. Similar to the Black faces, the majority of sessions were positively skewed, with the exception of sessions 3 and 4, which were roughly normally distributed. The P-p plot also showed that many of the points deviated from the normal line. Therefore, both *RT*s on Black and Egyptian faces likely violate the assumption of normality. Consequently, log transformations were conducted in order to make the sessions more normally distributed.

Mauchly's test indicated that both performance across sessions, $W(7, 27) = .033$, $p = .003$, and Sessions*Face Race, $W(7, 27) = .067$, $p = .040$, violated the assumption of sphericity. Consequently, the Greenhouse-Geisser correction will be used in the analysis of participants' performance across the sessions and for the interaction between sessions and face race.

A mixed design ANOVA indicated that there was a significant main effect of Sessions, $F(3.3, 27) = 5.29$, $p = .002$, $\eta^2 = .227$. Neither the main effects of Face Race, $F(1, 2) = .001$, $p = .970$, $\eta^2 < .001$, nor Group, $F(1, 2) = .50$, $p = .487$, $\eta^2 = .027$, were significant. A post-hoc analysis, using Tukey's LSD test, was conducted on Sessions (Table 7) in order to explore where the significance lies within the sessions. Table 7 indicates that session 1 had significantly longer *RT*s than sessions 4, 5, 6, 7, and 8; sessions 2 and 3 had significantly longer *RT*s than sessions 6, 7, and 8; and session 5 had significantly longer *RT*s than session 6. These results form a pattern that suggest that participants' *RT*s decreased over the course of the sessions.

Table 7.
Significant differences between sessions.

Sessions		ΔM (a-b)	*p*	95% Confidence Interval	
a	b			Lower	Upper
1	4	.106	.024	.016	.197
1	5	.105	.015	.023	.187
1	6	.145	< .001	.073	.217
1	7	.128	.007	.039	.218
1	8	.126	.014	.028	.224
2	6	.007	.007	.030	.169
2	7	.024	.024	.012	.154
2	8	.028	.028	.010	.151
3	6	.082	.004	.030	.134
3	7	.066	.010	.017	.114
3	8	.063	.015	.014	.112
5	6	.040	.046	.001	.079

In terms of interaction effects, Sessions*Group, $F(3.3, 27) = .86$, $p = .474$, $\eta^2 = .046$; Face Race*Group, $F(1, 2) = .07$, $p = .794$, $\eta^2 = .004$; Sessions*Face Race, $F(4.1, 27) = 1.45$, $p = .227$, $\eta^2 = .074$; and Sessions*Face Race*Group, $F(4.1, 27) = 1.21$, $p = .314$, $\eta^2 = .063$ were all non-significant.

Previous Faces. As all previous faces had, by definition, been seen before, participants could only score 'hits' or 'misses' on these faces. Consequently, *d*' and *c* cannot be calculated for these faces. Therefore, previous faces were only analysed in terms of their hit rates. Figure 7 shows the average hit rates of participants for Black and Egyptian faces between the two groups. In the control group, participants had higher hit rates on Egyptian faces than Black faces. In addition, there was an upward trend across the sessions. In the motivation group, participants' hit rates on faces were varied more substantially and did not consistently show better performance on Egyptian faces than Black faces, or vice versa. In addition, the motivation group's average hit rates showed a slight downward trend across sessions. Further statistical analyses were not conducted on the data on previous faces, as the hit rate cannot be meaningfully interpreted without the false alarm rate.

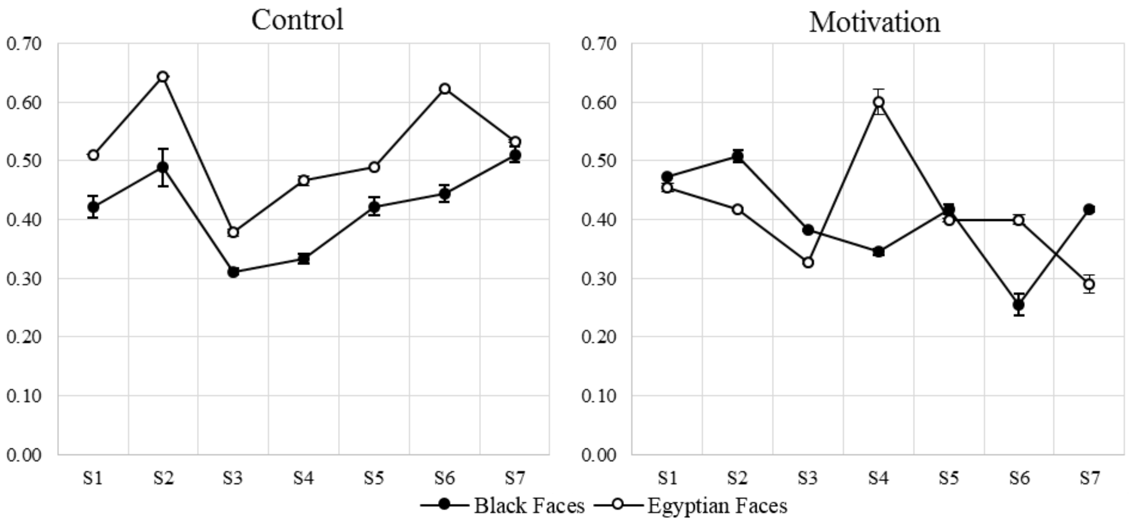


Figure 7. Hit rates across sessions for the control and motivation groups, with 95% confidence intervals.

Discussion

The current study replicated the results of the majority of studies, in that the participants were significantly better at recognising own-race faces than cross-race faces (Malpass & Kravitz, 1969; Meissner & Brigham, 2001; Sporer, 2001). This was seen in both their increased discrimination accuracy (ΔM = .48) and higher response criterions (ΔM = .23) for own-race faces. It is important to note that ANOVAs which include repeated measures tend to inflate η^2 and the effect sizes found for discrimination accuracy (η^2 = .704) and response criterion (η^2 = .497) are unlikely to be accurate (Dunlap, Cortina, Vaslow, & Burke, 1996). No significant differences were found between participants' *RT*s for Black and Egyptian faces. This finding is contrary to other previous studies, where participants took significantly longer to identify cross-race faces than own-race faces (Chance & Goldstein, 1987; Valentine, 1991).

The first research question posed in this study was whether increasing contact would affect the CRE in participants' performance. In terms of discrimination accuracy, there is weak evidence that participants may have improved. The ideal pattern to support this finding would be that every subsequent session had significantly higher accuracy than the previous session. The results showed that participants were significantly better at identifying Egyptian faces in two of the later sessions (six and seven), than a number of the earlier sessions (two, four, and five). However, they were also significantly better in one of the earlier sessions (three) than a later session (five). In addition, there is no clear, increasing trend, across sessions in any of the visual representations of the data. Consequently, the results are far from the ideal pattern. This contradicts the predictions of the differential experience hypothesis, which proposes that the CRE should decrease as contact increases (Malpass & Kravitz, 1969). The weak pattern that is present may be due to the large amounts of variation inherent in faces. That is, faces have a plethora of attributes on which they can vary, for example: features, distinctiveness, and likability. So, although the faces in this studied were randomised, none of the attributes of these faces were accounted for. These attributes could result in differential performance in participants' discrimination accuracy between sessions. Attributes of faces could be randomised, and included in the study's design; however, to do so would require a vast sample of faces that have been coded for their attributes. Assembling this sample would take a significant amount of time and resources, and was judged to be too impractical for the purposes of this study.

In terms of the response criterion, the results show that the participants were consistently conservative in identifying own-race faces, while they were much more inconsistent in identifying cross-race faces. Although, response criterion was significant across sessions, it showed no clear trend. The majority of the significance was present because of the sharp increase in response criterion for session four. This spike may indicate that the faces in that session had attributes which prompted participants to be much more conservative in their answers. Therefore, similar to participants' discrimination accuracy, increasing the amount of contact participants had with Egyptian faces had no effect on participants' response criterion.

The previous faces were included in the analysis as a rough measure to discern if there was any difference in participants' performance on faces with long intervals between presentation and recognition. The control group had a higher hit rate for Egyptian faces, than Black faces. However, as mentioned above, this may be due to participants being more liberal in their recognition of Egyptian faces. The motivation group showed no clear pattern in their performance. These results suggest that contact did not affect participants' performance.

Finally, there was a clear pattern of participants identifying faces faster, as contact increased. Earlier sessions (one, two, and three) had significantly longer *RTs* than later sessions (four, five, six, seven, and eight). This can also be seen in the visual representations of the data, where there is a decreasing trend across subsequent sessions. This decrease was observed for both Black and Egyptian faces. It is important to note that, as contact increased, participants got significantly faster at identifying the faces, however they did not get any more accurate in doing so. Due to the fact that this occurred for both Black and Egyptian faces, this is likely an artefact of repeating the facial recognition test every session (i.e. a training effect). That is, participants take less time to identify the faces as they become more familiar with the format and requirements of the test.

Consequently, to answer the first research question, the results suggest that experimentally increasing contact does not affect the CRE. Malpass (1981) proposed three explanations of why this may be the case; first, the amount of contact with cross-race faces in experiments like this is miniscule, when compared to the thousands of hours of contact with faces that participants have had in their life. It may be possible to learn to recognise cross-race faces as accurately as own-race faces. However, it would either take a much longer time that anything that is feasible in an experimental setting, or only result in marginal increases in facial recognition abilities (Woodhead et al., 1979). Second, participants' strategies for encoding and recognising faces are developed early in life, during critical developmental periods, and are unlikely to change after these periods (Malpass, 1981). This suggests that it is highly unlikely that cross-race faces will ever be identified as accurately as own race faces, because participants did not have contact with those faces during these critical periods. Finally, it may be that our understanding of the processes that underlie facial recognition are not adequate in constructing an effective training programme (Malpass, 1981). Subsequently, it is possible to eliminate the CRE, but we do not have the appropriate knowledge to do so.

The second research question was whether motivating participants would cause the CRE to change. The results show that participants' discrimination accuracy, response criterion, and *RT*s were not significantly different between the motivation and control groups. This finding contradicts the results of some studies which found that motivation can reliably eliminate the CRE (Hugenberg et al., 2010; Young et al., 2012). However, this supports more recent research which found that motivation instructions have no effect on the CRE (Bornstein, Laub, Meissner, & Susa, 2013). Alternatively, in line with Wan et al.'s (2015) proposal, it may be that motivation has different effects across different research contexts. Therefore, the results may suggest that motivation is not relevant in a South African context. However, it is also possible that this null effect may be due to the operationalisation of motivation (i.e. leaderboards) in the present study. That is, the leaderboards used may be inadequate tools for inducing motivation in participants. Further research is necessary to determine whether motivation has any effects on the CRE in a South African context.

Finally, the third research question was whether the interaction between motivation and contact would affect the CRE. The results show that there were no interaction effects present between motivation and contact in participants' discrimination accuracy, response criterion, or *RT*s. This finding is in stark contrast with the CIM model, which predicted that contact and motivation should have had a substantial effect on the CRE, as they are two of the core factors in facial recognition processing (Hugenberg et al., 2010).

Limitations. The present study was conducted on a small group of participants, and may not have had adequate statistical power to detect the effects of contact and motivation on the CRE. The appropriate number of participants that was suggested by the power analysis (i.e. $N = 24$) was not met, due to the high dropout rate of participants ($N = 10$). A possible consequence of the small sample in this study was that not all of the data upheld the assumptions of mixed designs ANOVAs. These violations may have biased the results discussed above. Future research should aim to recruit larger samples of participants, which would ensure more accurate results. In addition, this research should include a cross-over design. That is, Egyptian participants should be recruited to balance the race of the participants across both conditions. This would reduce the chance that results are due to confounding variables, such as stimulus effects. The participants used in this study were also drawn from a relatively small population (university students) and the results may not be generalisable to broader populations.

Another limitation was that the computer generated faces, created in the ID program, used in this study were not adequate substitutes for real-life faces. That is, the disembodied heads used in this study bear little resemblance to the more wholesome experience of faces people would view in their lives. In addition, a number of these faces had visual distortions around certain facial features, which were artefacts of the process that is used to create the faces. Future research should either manually edit each of the faces used to ensure that there are no graphical distortions, or use a database of real faces. Although significantly harder to gather, the real faces would reinforce the external validity of the study. Another problem was that, as mentioned above, the faces were not standardised, and had large variation in their attributes. This could be solved by constructing the test with a set of faces that are judged to have similar attributes. Such a test could be repeatedly tested, and normed, to ensure that the attributes of the faces have little variation, between and across sessions. This would be the gold standard for the test, but may not be feasible, given the vast amount of time and resources that would be required.

Similar to the faces, the practice tasks used in this study were limited, in that they are not accurate portrayals of the contact that people experience in real-life. The artificial presentation of faces had no effect on the CRE, whereas actual contact may have had an effect. The practice tasks themselves were also fairly rudimentary, and could be developed to increase participants' engagement and investment in increasing their facial recognition ability. The relatively small number of sessions, or time spent increasing contact, may also have limited the study. In line, with Malpass' (1981) suggestion, the amount of sessions that the participants spent training, needs to be significantly increased, in order to detect any kind of effect on the CRE.

Directions for future research. The present study was the first of its kind in that, within the CRE paradigm, no other studies have attempted to experimentally manipulate contact. More research is needed to expand on, and strengthen, the results found in this study. Future research could incorporate many of the suggestions listed above, to avoid the limitations of the present study. Information on how contact and motivation can vary in different contexts is sorely needed; research using different races of participants, and faces, will be vital in identifying the conditions in which the CRE is most vulnerable to change.

Furthermore, the predictions and past results of the various models of the CRE, highlighted in the introduction, need to be tested using experimental manipulations. For example, one could recreate Michel et al.'s (2006) experiment of misaligned top-and-bottom halves of faces, to assess whether experimentally increasing contact will cause participants perform more similarly on cross-race faces and own-race faces. This would provide useful insight into the plausibility of the feature-selection model.

Of particular note, this study found that increased contact caused participants to get faster at identifying faces, but no more accurate. This finding may be related to how confident participants feel in their ability to recognise the faces correctly, as contact increases. Therefore, a study that included participant confidence levels within the research design could further explore why participants are demonstrating this effect.

Conclusion

The CRE literature has been divided for many years, with multiple theories proposing that they have found the elusive 'underlying mechanism'. However, evidence has yet to be found that proves any one of these theories conclusive. Perhaps it is time for a change in tack, where well-established findings can be re-examined using novel measures, and methods. Experimentally manipulating contact is just one of a multitude of possible methods that can shed new light on widely accepted assumptions about the CRE. Such a renaissance may provide the key to assimilating the current theories, and finally creating a single, unifying theory of the CRE. Such a theory could reveal how the CRE operates, as well as in what contexts is it most likely to be present. This would be greatly beneficial in many justice systems around the world, as it would allow for much finer analysis of cross-racial identification.

References

Bornstein, B. H., Laub, C. E., Meissner, C. A., Susa, K. J. (2013). The cross-race effect: Resistant to instructions. *Journal of Criminology*, vol. 2013, 1-6. doi:10.1155/2013/745836

Brigham, J. C., Brooke-Bennett, L., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (pp. 257-281). Mahwah, NJ: Lawrence Erlbaum & Associates.

Chance, J. E., & Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. *Bulletin of the Psychonomic Society*, *25*(6), 415-418.

Daniels, L., Mackier, D., & van Helsdingen, G. (2013). FaceGame. (Unpublished honour's thesis). University of Cape Town, South Africa.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170-177. doi:10.1037/1082-989X.1.2.170

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. -G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41, 1149-1160. doi:10.3758/BRM.41.4.1149

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., Slater, A., Tian, J., & Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*(8), 1199-1210. doi:10.1068/p6136

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211-214. doi:10.3758/BF03329829

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45-56. doi:10.1348/000712607X199981

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

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

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*(4), 559-574. doi:10.1037/0096-3445.129.4.559

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.*, *19*, 577-621. doi:10.1146/annurev.ne.19.030196.003045

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Mahwah, New Jersey: Lawrence Erlbaum & Associates.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330-334. doi:10.1037/h0028434

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, *14*(2), 285-292. doi: 10.3758/BF03212392

Malpass, R. S. (1981). Training in face recognition. In G.M. Davies, H.D. Ellis, & J.W. Shepherd (Eds.), *Perceiving and Remembering Faces* (pp. 271-285). London: Academic Press.

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*(1), 3-35. doi:10.1037//1076-8971.7.1.3

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology, 19*, 545-567. doi:10.1002/acp.1097

Michel, C., Rossion, B., Han, J., Chung, C., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*(7), 608-615. doi:10.1111/j.1467-9280.2006.01752.x

Posel, D. (2001). Race as common sense: Racial classification in twentieth-century South Africa. *African Studies Review, 44*(2), 87-113. doi:10.2307/525576

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*, 313-331. doi:10.1111/j.2044-8295.1989.tb02323.x

Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology*, *100*, 717-728. doi:10.1348/000712608X396503

Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition*, *5*(2), 144-165. doi:10.1521/soco.1987.5.2.144

Seutloali, G. M. (2014). *Own-race bias in facial recognition amongst black, coloured and white participants* (Unpublished honour's thesis). University of Cape Town, South Africa.

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*(2), 71-84. doi:10.1207/S15324834BASP2202_1

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36-97. doi:10.1037/1076-8971.7.1.36

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149. doi:10.3758/BF03207704

The Innocence Project (n.d.). Eyewitness misidentification. Retrieved from http://www.innocenceproject.org/causes-wrongful-conviction/eyewitness-misidentification

Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2007). An evaluation of ID: An eigenface based construction system. *South African Computer Journal*, *37*, 90-97.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *43*(2), 161-204. doi:10.1080/14640749108400966

Walker, P. M., & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, *32*, 1117-1125. doi:10.1068/p5098

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & Mckone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition*, *144*, 91-115. doi:10.1016/j.cognition.2015.07.011

Woodhead, M. M., Baddeley, A. D., & Simmonds, D. C. V. (1979). On training people to recognise faces. *Ergonomics*, *22*(3), 333-343. doi:10.1080/00140137908924617

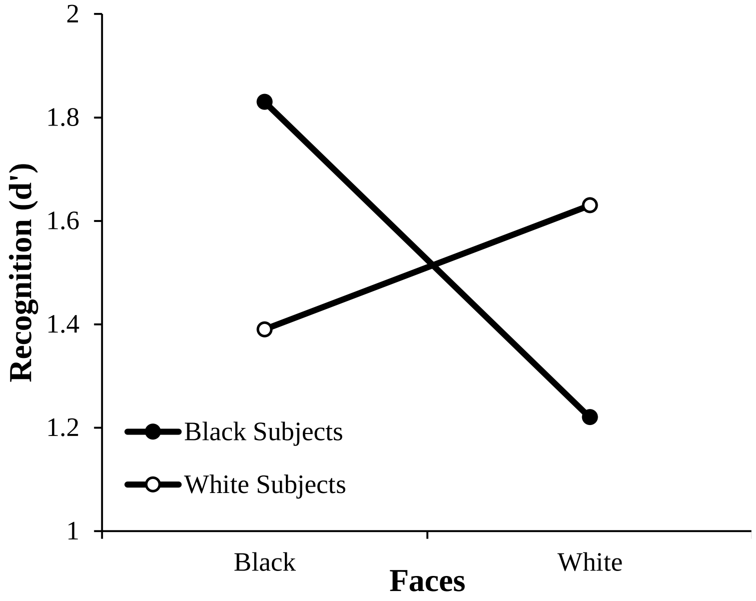
Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17*, 365-373. doi:10.1002/acp.898

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*(2), 116-142. doi:10.1177/1088868311418987

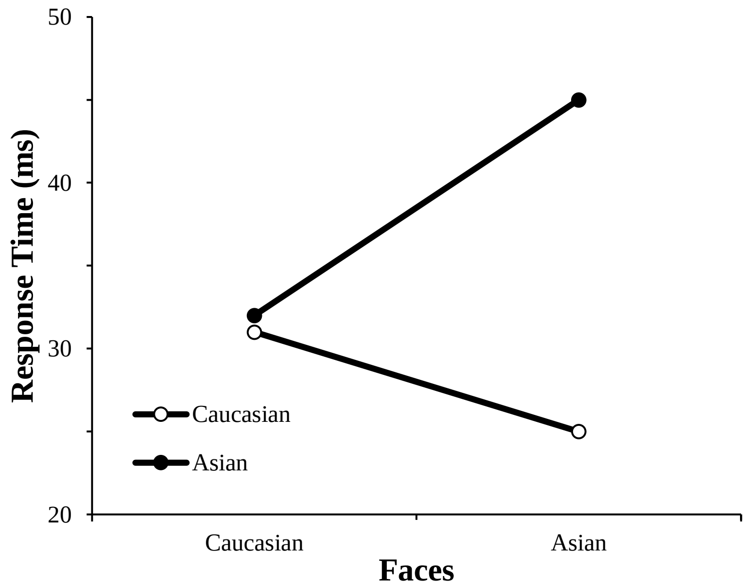
Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. Canada: O'Reilly Media Inc.

Appendices

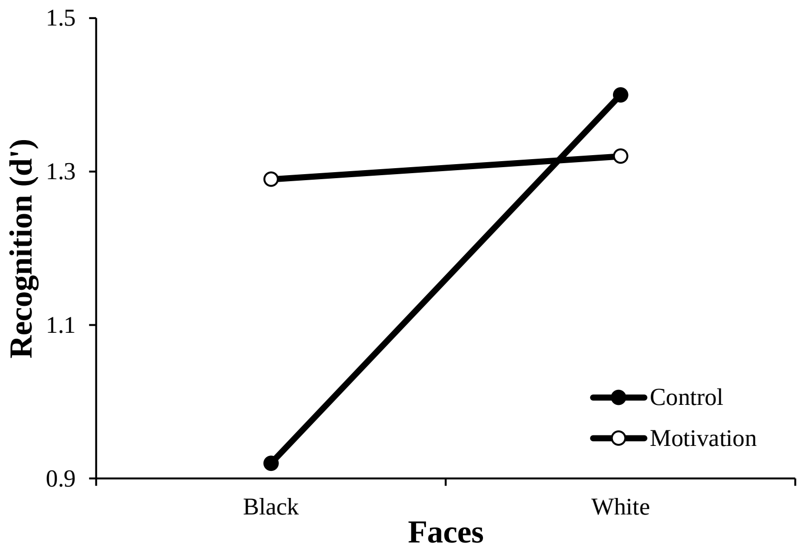
Appendix A. The classic, disordinal interaction commonly found between participant group and ethnicity of faces; recognition performance measured by d' (Data from Malpass, Lavigueur, & Weldon, 1973).



Appendix B. Interaction between participant group and ethnicity of faces on response time when identifying misaligned top-and-bottom faces (Data from Michel et al., 2006).



Appendix C. Disordinal interaction between motivation/control groups and ethnicity of faces; recognition performance measured by d’ (Data from Hugenberg et al., 2007).



Appendix D. Example of the leaderboard for the motivation condition.

Leaderboard

Rank	Name	Total Score
1	-	45
2	-	42
3	-	39
4	Michael	36
5	-	35
6	-	33
7	-	32
8	-	30
9	-	28
10	-	26

Appendix E. Announcement to be released for recruiting participants.

Opportunity for students to participate in psychological research.

Dear Students,

An upcoming study in the Department of Psychology is looking for **Black** participants to recruit. The study involves playing games to try and learn a new population of faces. The study aims to recruit 28 participants.

You will be required to attend **2 sessions** of 75 minutes a week, for **4 weeks**.

Participants who complete the study will stand a chance to win one of four prizes: either R1500, R750, R500, or R250.

If you are interested in participating, or have any questions, please contact me via email at Tristan.rayner.ishb@gmail.com

Appendix F. Consent form.

Informed Consent Form

University of Cape Town

Department of Psychology

<u>An Experimental Manipulation of Contact and Motivation in the Cross-Race Effect</u>

Dear Participant,

1. **Invitation and Purpose**

   You are invited to take part in this study which investigates amount of contact and motivation in facial recognition. I am a student researcher in the Department of Psychology at the University of Cape Town.

2. **Procedures**

   - If you decide to take part in this study, you will be required to attend two experimental sessions a week, for four weeks. In these sessions, you will be tested on your ability to recognize different types of faces and then play practice games to get better at this recognition.

- Each session should take approximately 60 minutes, while the eighth (last) session should last approximately 15 minutes.
- Participating in this study is entirely voluntary. You are free to leave the study at any time with no penalty.

3. **Risks, Discomfort and Inconveniences**
   - This study poses little risk of harm to you.
   - If, at any point, you feel discomfort about the tests, or practice tasks, you are free to notify the researcher.
   - You may be inconvenienced by having to give up approximately 7 hours and 15 minutes of your time.
4. **Benefits**

   This study gives you an opportunity to learn about an area of research in Psychology. In addition, it teaches you to distinguish between a new population of faces.
5. **Privacy and Confidentiality**
   - Any information gathered from you for the study will be strictly confidential. This includes your name, demographic details, and scores obtained on any tests or tasks. You have the right to request that any information you have shared be removed from the study.
   - No one except myself, and my university supervisors, will be allowed to view, or have access to, any data obtained.
   - The findings of this research will be written up in the form of an Honour's thesis and may be published in an academic journal.
6. **Compensation**

   If you attend, and complete all eight sessions of the study, you will be entered into a lottery where you will stand a chance to win one of four prizes: R1500, R750, R500, or R250.
7. **Contact Details**

   If you have any questions, suggestions, or complaints about the study please contact Tristan Rayner at tristan.rayner.ishb@gmail.com, or Dr. Colin Tredoux at colin.tredoux@uct.ac.za.
8. **Signatures**

   (Subjects name) _________________________ has been informed of the nature and purposes of the procedures described above, including any risks involved in its performance. He/she has been given time to ask any questions and these questions have been answered to the best of the investigator's ability. A signed copy of this consent form will be made available to the subject.

______________________ ______________________

Investigator's Signature **Date**

I, have been informed about this research study and understand its purpose, possible benefits, risks, and discomforts. I agree to take part in this research as a participant. I know that I am free to withdraw this consent and quit this study at any time, and that doing so will not cause me any penalty or loss of benefits that I would otherwise be entitled to enjoy. I am aware that the research will be written up in the form of an honours research project and may be published in an academic journal.

______________________ ________________

Participant's Signature **Date**

Appendix G. Demographic information form.

Demographic Information

Please complete the following:

1. Name: ______________________________

2. Student Number: ______________________________

2. Race: ______________________________

3. Gender: ______________________________

4. Age: ______________________________

Appendix H. Questionnaire assessing amount of previous contact with Egyptian faces.

Please read the questions carefully and circle the appropriate answer:

1. Have you ever lived in Egypt?	Yes	No
2. Have you ever visited Egypt?	Yes	No
3. Do you have any Egyptian friends?	Yes	No
4. Do you know any Egyptian people?	Yes	No

Appendix I. Ethical Approval

Lauren Wild
to me

Jun 27 (2 days ago)

Dear Tristan

Thank you; that looks good. You now have ethical approval for your study.

Kind regards

Lauren

Lauren Wild (PhD)
Associate Professor
Department of Psychology
University of Cape Town
Rondebosch 7701
South Africa
Tel. (021) 650 4607