Self-reported face recognition abilities for own- and other-race faces

Alejandro J. Estudillo[1,2]

[1]Bournemouth University

[2]University of Nottingham Malaysia

RUNNING HEAD: Insights for own- and other-race face recognition

KEYWORDS: face identification, face recognition, other-race effect, eyewitness identification, PI20, self-reported face recognition

This is an author-created, un-copyedited version of an article accepted for publication in the Journal of Criminal Psychology

Abstract

Purpose

The other-race effect shows that people are better recognizing faces from their own-race compared to other-race faces. This effect can have dramatic consequences in applied scenarios whereby face identification is paramount, such as eyewitness identification. This study investigates observers' insights into their ability to recognize own- and other-race faces.

Design/methodology/approach

Chinese ethnic observers performed objective measures of own- and other-race face recognition —the Cambridge Face Memory Test Chinese and the Cambridge Face Memory Test original—, the PI20 —a 20-items self-reported measured of general face recognition abilities—, and the ORE20 —a new developed 20-items self-reported measure of other-race face recognition—.

Findings

Recognition of own-race faces was better compared to other-race faces. This effect was also evident at a phenomenological level, as observers reported to be worse recognizing other-race faces compared to own-race faces. Additionally, although a moderate correlation was found between own-race face recognition abilities and the PI20, individual differences in the recognition of other-race faces was only poorly associated with observers' scores in the ORE20.

Implications

These results suggest that observers' insights to recognize faces are more consistent and reliable for own-race faces.

Practical implications

Self-reported measures of other-race recognition could produce misleading results. Thus, when evaluating eyewitness' accuracy identifying other-race faces, objective measures should be employed.

Introduction

People are generally more efficient and accurate recognizing faces of their own-race compared to other-race faces. This effect, which is known as the other-race effect (ORE, Malpass & Kravitz, 1969), is evident across different cultures and countries (Meissner & Brigham, 2001) and has been found with different paradigms, including perceptual matching (Kokje, Bindemann, & Megreya, 2018; Megreya, White, & Burton, 2011), face recognition (Chiroro & Valentine, 1995; Estudillo, Lee, Mennie, & Burns, 2020; Malpass & Kravitz, 1969; Wong, Stephen, & Keeble, 2020) and line-up identification tasks (Evans, Marcon, & Meissner, 2009). A meta-analysis study comprising nearly 5000 participants across 39 different studies showed that people are 2.23 times more likely to recognize own-race faces compared to other-race faces (Meissner & Brigham, 2001). This evidence points that the ORE is a very robust effect.

Rather than being homogeneous, the size of the ORE presents substantial individual differences across observers (Wan et al., 2017). For example, a recent study with Caucasian and Asians observers showed the standard ORE at a group level, that is recognition performance was better for own compared to other race faces (Wan et al., 2017). However, an individual differences analysis showed that around 8% of these observers performed so extremely poor at recognizing other-race faces that they could be considered to suffer a specific type of face-blindness for other race faces. The authors concluded that the lack of contact with other-race faces is the main cause of the ORE (Wan et al., 2017; see also Estudillo et al., 2020).

Regardless of its origin, it has become clear that the ORE can have catastrophic consequences in applied scenarios whereby face identification is of paramount importance, such as in eyewitness identification parades and id-verification settings (e.g., passport control officers). A paradigmatic case illustrating this issue was Ronald Cotton's wrongful conviction case (see http://www.theinnocenceproject.org). Mr. Cotton is an African-American citizen who was accused of sexual assault in 1984. The rape victim, a Caucasian lady, misidentified Mr. Cotton as the rapist. As consequence, Mr. Cotton spent more than 10 years in prison for a crime he did not commit until he was exonerated by DNA evidence in 1995. Thus, if an eyewitness and the perpetrator are from different races, it is crucial to determine how reliable an eyewitness is recognizing faces from the perpetrator's race. Therefore, developing tools to assess eyewitnesses' ability to recognize other-race faces is an important endeavor for forensic scientists.

Recognition confidence (i.e., "how confident are you that you have done a correct recognition decision?") is a potential maker that can provide some hints about an eyewitness's accuracy identifying a perpetrator. Recognition confidence has been widely used in forensic research and even police and lawyers consider that this marker can provide a reliable measure of identification skills (Potter & Brewer, 1999; Sauer, Palmer, & Brewer, 2019). These conclusions are somehow supported by some research. For example, it has been found that recognition confidence is positively associated with face identification accuracy (Wixted & Wells, 2017), but this association is modulated by participants' face recognition abilities. This is such that poor face recognizers are much more likely to make high confidence identification errors compared to good face recognizers (Grabman, Dobolyi, Berelovich, & Dodson, 2019). However, other authors have shown that high recognition confidence can be misleading at individual level (Sauer et al., 2019). This can be particularly important in other-race face identification, as it has been shown that people are generally more overconfident when identifying other-race faces compared to own-race faces (Dodson & Dobolyi, 2016).

In addition, findings about the reliability of recognition confidence for other-race faces are mixed. For example, although some studies have recently found that high recognition confidence is equally associated with the identification accuracy of own and other race faces (Dodson & Dobolyi, 2016; Grabman et al., 2019), other studies reported that observers not only tend to be less accurate in judging whether they will identify an other-race face compared to an own-race face (Hourihan, Benjamin, & Liu, 2012; Smith, Lindsay, Pryke, & Dysart, 2001), but also they tend to be have a less clear memory for other-race faces (Brigham, Bennett, Meissner, & Mitchell, 2007; Smith, Stinson, & Prosser, 2004). One feature of recognition confidence that might explain this disparity in the results is its situation-specificity. In recognition confidence procedures, observers are simply asked about their confidence about having seen a particular face. However, this procedure is specific for the seen face and does not gather information about how good the observer is recognizing either own- and other-race faces.

One potential alternative to recognition confidence could be self-reported measures of face recognition (Bate et al., 2018; Bobak, Mileva, & Hancock, 2019; Livingston & Shah, 2018; Shah, Gaule, Sowden, Bird, & Cook, 2015; Ventura, Livingston, & Shah, 2018). These measures describe different daily life situations involving face recognition abilities, so, in contrast to recognition confidence, self-reported measures of face recognition are not situation-specific. For example, Sha and colleagues (Shah, Gaule, et al., 2015) introduced the PI20 questionnaire, a twenty-statements self-reported measure of face identification. Agreement with these statements is scaled on a five-point Likert-scale. The authors reported strong negative correlations between the score in the PI20 and objective measures of face identification, such as famous face recognition, the Cambridge Face Memory Test (CFMT-Original, Duchaine & Nakayama, 2006), and the Glasgow Face Matching Test (Shah, Sowden, Gaule, Catmur, & Bird, 2015) —a perceptual measure of face identification that simulates the id-verification scenarios and that it has been used in several applied studies with police (Robertson et al., 2016) and passport control officers (White, Kemp, Jenkins, Matheson, & Burton, 2014). Interestingly, the PI20 has been translated to different languages, including Portuguese (Ventura et al., 2018), Japanese (Nakashima et al., 2020) and Mandarin (Estudillo & Wong, 2020). In general, studies using the PI20 have concluded that adults have moderate to strong insights into their face recognition abilities (Livingston & Shah, 2017; Ventura et al., 2018 but see Bobak et al., 2018; Estudillo & Wong, 2020).

A question that arises is whether these insights can be generalized to other-race face recognition. At a theoretical level this has important consequence for models of metacognition. For example, according to the well-known Dunning–Kruger effect, unskilled people (i.e., non-experts in a specific field) present very poor insights about their actual performance (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Fakcharoenphol, Morphew, & Mestre, 2015). This effect which, has been found in different domains, including humour, logic, grammar knowledge and face perception (Dunning et al., 2003; Pennycook, Ross, Koehler, & Fugelsang, 2017; Zhou & Jenkins, 2020), makes a clear prediction about our study: as people are generally experts recognizing own-race faces, but not other-race faces (Estudillo et al., 2020; Tanaka, Heptonstall, & Hagen, 2013), people would have more accurate insights into their recognition abilities for own-race faces compared to other-race faces. Thus, this study also allows us to test the Dunning-Kruger effect from a different domain: the recognition of own- and other-races faces. From a more applied perspective, having insights to recognize other-race faces would be important in forensic scenarios, such as during the identification of other-race perpetrators or during id verification processes, as it could provide information about how reliable an observer is identifying other-race faces. However, to date, self-reported measures of face recognition abilities have not been adapted to other-race faces.

The present study aims to explore observers' insight to recognize own- and other-race faces. To achieve this, we used the PI20 to evaluate observers' own-race face recognition and a new developed self-reported measure for evaluating other-race face recognition (the ORE20). Given the high reliability and construct validity of the PI20 (Shah, Gaule, et al., 2015), the ORE20 is a partial adaptation of the PI20. Participants performed objective measures for the recognition of own- and other-race faces, the PI20 and the ORE20. We expect participants to be better recognizing own- than other-race faces. Following previous research with the PI20, we also expect that participants will have moderate to strong insights into their face recognition abilities for own-race faces. This would be evident by a negative association between the objective measure of own-race face recognition and the PI20. Finally, if participants have insights into their recognition for other-race faces we would expect that the ORE20 would be associated with their performance in the objective measure of the recognition of other-race faces.

Methods

Participants

Eighty-five Chinese ethnic participants (58 females) from the University of Nottingham Malaysia took part in this study for course credits. Observers mean age was of 21 yeas (SD = 3). All observers reported having normal or corrected-to-normal vision. Participants gave their informed consent and were debriefed at the end of the study. This study was approved by the ethics committee of the University of Nottingham Malaysia.

Materials, apparatus and procedure

This study involves an objective face recognition stage and a self-reported face recognition abilities stage. The order of these stages was counterbalanced across participants. Testable platform was used to present stimuli and to record observers' responses. As English is the teaching language at the University of Nottingham Malaysia, all the instructions and questionnaires in this study were in English.

Objective face recognition stage

In this stage, observers performed both the CFMT-Original (Duchaine & Nakayama, 2006) and the CFMT-Chinese (McKone et al., 2012) in a counter-balanced order. Both tests follow an identical format, but they use Caucasian and Chinese faces as stimuli, respectively. These tests are valid measures of face recognition as they require the recognition of faces across different images, and no simple pictorial recognition (Bruce, 1982; Estudillo, 2012; Estudillo & Bindemann, 2014; Longmore, Liu, & Young, 2008). Participants are required to learn and recognize different unfamiliar faces in three different stages. In the *same image* stage, observers are asked to learn a target identity, which is presented in three different orientations (i.e., mid-profile left, frontal and mid-profile right). Observers are then presented with a three-alternative forced-choice task, whereby they are required to select the studied target identity among two foils, with one trial per orientation. This stage is repeated for each of six target identities, giving a total of 18 trials. In the *novel images* stage, observers are presented with the same six target identities all in once for 20 seconds. Then, they have to identify a novel instance of the target face among two distractors. This stage contains a total of 30 trials. The last stage is identical to the previous stage, but faces at test are presented with visual noise in order to make the task more challenging. This stage contains a total of 24 trials, thus the maximum score an observer can obtain is 72. Internal reliability analysis showed alpha values of 0.86 for the CFMT-Original and 0.90 for the CFMT-Chinese. These values are in agreement with previous research (e.g., Bowles et al., 2009; Estudillo et al., 2019).

Self-reported face recognition abilities stage

In this stage, observers filled in both the PI20 and the ORE20 questionnaires in a counter-balanced order. The PI20 (Shah, Gaule, et al., 2015) is a self-reported measure of face recognition. It contains 20 items describing the experience of face recognition. In the instructions of this test, we emphasized that these items refer to the recognition of Chinese peers (i.e., own-race faces for our sample).

The ORE20 questionnaire (see Table 1) was developed to conduct this study. This questionnaire is a self-reported measure of face recognition abilities for other-race faces. The questionnaire comprises 20 items reporting different situations describing the experience of recognizing Caucasian faces (other-race faces for our sample). Most of the items of the ORE20 were adapted from the PI20 (e.g., Item 2 ORE20: to recognise Caucasian people, I rely on non-facial cues, such as voice, Item 10 PI20: Without hearing people's voices, I struggle to recognize them; Item 10 ORE20: My face recognition ability is similar for Caucasians and Chinese faces, Item 2 PI20: My face recognition ability is worse than most people). However, this adaptation is not possible for some of the PI20 items (e.g., PI20 item 11: Anxiety about face recognition has led me to avoid certain social or professional situations; PI20 Item 13: I am very confident in my ability to recognize myself in photographs; Item 15: My friends and family think I have bad face recognition or bad face memory). Thus, other items of the ORE20 were created with the aim of capturing different aspects involved in the processing of other races faces, such as learning (e.g., Item 20 ORE20: I have to try harder to memorise Caucasian faces than Chinese faces), recognition (e.g., Item 13 ORE20: I struggled to recognise my Caucasian classmates or colleagues), discrimination (e.g., Item 14 ORE20: I find that most Caucasian faces look alike) and race categorization (e.g., Item 9 ORE20: I mistake familiar Caucasian people as Chinese people when they change their hairstyle).

Observers were asked to rate their agreement with each statement of the PI20 and the ORE20 on a five-point Likert-scale (1 = strongly agree, 5 = strongly disagree). In both questionnaires, higher total scores reflect lower recognition abilities. However, for the sake of simplicity, the scores of the positive items were reversed (PI20: items 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 14, 15, 16, 18, 20; ORE20: items 2, 3, 4, 5, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18, 20), so that higher scores in both questionnaire indicate better self-perceived face recognition abilities. Internal reliability analysis revealed alpha values of 0.84 and 0.89, for the PI20 and the ORE20 respectively.

Results

In a first part of our analysis, we compared observers' recognition performance for own and other-race faces. Observers were better recognizing own- compared to other-race faces [t(84) = 7.77, $p < .001$, d = .68, CI = .48 - .88] (Figure 1a). Interestingly, the same pattern was also obtained at a phenomenological level as observers reported to be worse recognizing other-race faces —as reflected by their scores in the ORE20— compared to own-race faces —as reflected by their scores in the PI20— [t(84) = 6.30, $p < .001$, d = .65, CI = .42 - .88] (Figure 1b).

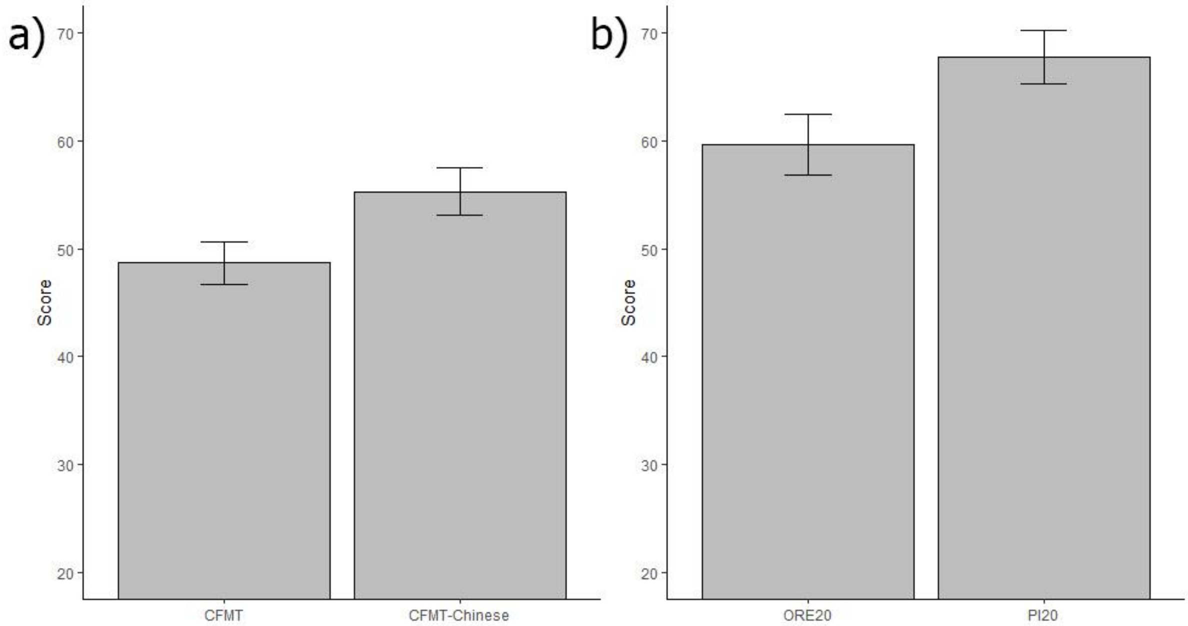

Figure 1. (a) Mean scores in the CFMT-Chinese and in the CFMT-Original. (b) Mean scores in the ORE20 and PI20. Higher scores reflect better self-reported recognition. Error bars represent 95% CI.

In a second part of our analysis we explored observers' insight to recognize own- and other-race faces. Scores in the PI20 were moderately associated with the scores in the CFMT-Chinese [r = .45, $p < .001$, CI = .26 - .60] (Figure 2a). On the other hand, scores in the ORE20 were only poorly associated with the scores in the CFMT-Original [r = .21, $p = .04$, CI = .00 - .41] (Figure 2b). The differences between both correlations were explored using

Silver and colleagues' test of non-overlapping dependent correlations (Diedenhofen & Musch, 2015; Silver, Hittner, & May, 2004). This test showed that the magnitude of the association between own-race face recognition and the PI20 was stronger that the magnitude of the association between other-race face recognition and the ORE20 (z = 2.08, $p < .05$). The PI20 was also associated with observers' performances in the CFMT-Original [r = .37, $p < .001$, CI = .17 - .54] (Figure 2c). However, Hittner and colleagues' test of overlapping dependent correlations (Diedenhofen & Musch, 2015; Hittner, May, & Silver, 2003) showed no differences between the magnitudes of this association and the association between the ORE20 and the CFMT-Original [z = 1.58, $p = .11$].

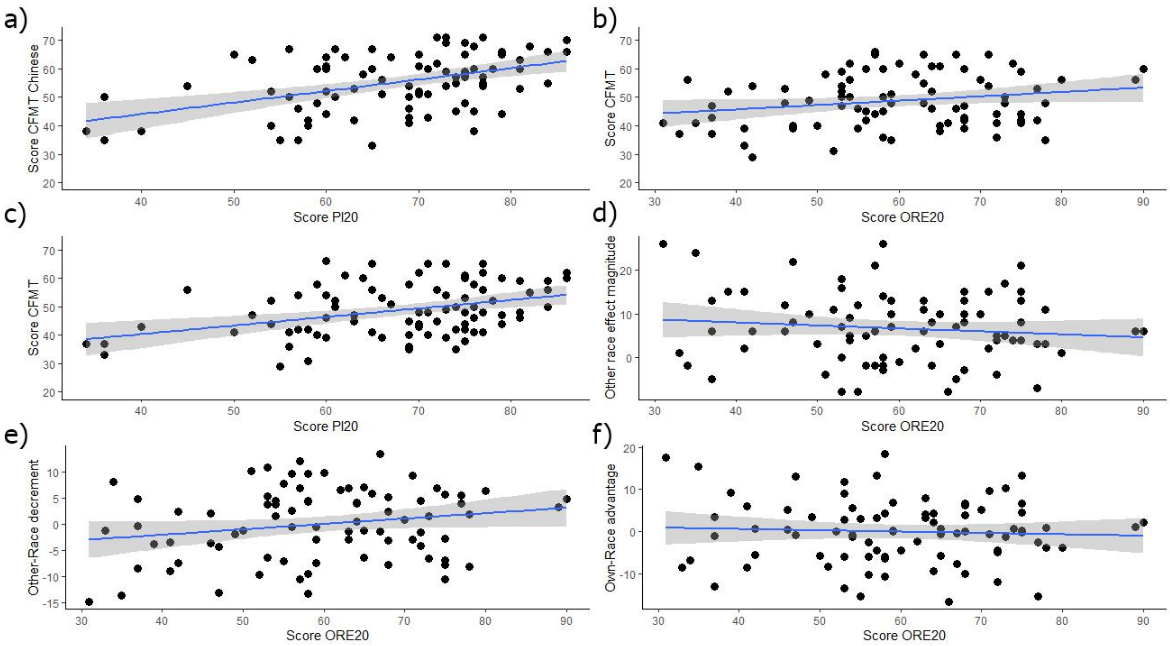

Figure 2. (a) Simple correlation between the own-race face recognition measure (CFMT-Chinese) and the PI20. (b) Simple correlation between the other-race face recognition measure (CFMT-Original) and the ORE20. (c) Simple correlation between the other-race

face recognition measure (CFMT-Original) and the PI20. (d) Simple correlation between the other-race face effect magnitude (CFMT-Chinese – CFMT-Original) and the ORE20. (e) Simple correlation between the other-race decrement (residuals from regressing own- from other-race performance) and the ORE20. (f) Simple correlation between the own-race decrement (residuals from regressing other- from own-race performance) and the ORE20.

We also explored whether scores in the ORE20 questionnaire predict the other-race effect magnitude. We first considered the magnitude of the other-race effect as the difference between accuracy performance in the CFMT-Chinese and the CFMT-Original (i.e., subtraction method). In this sense, it would be expected that higher scores in the ORE20 would be negatively associated with the other-race effect magnitude (see Figure 2d). We found that the score in the ORE20 was not associated with the other-race effect magnitude [r = -.11, p = *.26,* CI = -.31 - .10]. However, it has been argued that the subtraction method for calculating the other-race effect might hinder the contribution of the two components of the other-race effect, namely the own-race advantage and other-race decrement (DeGutis, Mercado, et al., 2013). On the contrary, calculating the other-race effect with a regression method allows the isolation of these two components by regressing other- from own-race performance to produce own-race advantage residuals, and own- from other-race to produce other-race decrement residuals (DeGutis, Mercado, et al., 2013; DeGutis, Wilmer, Mercado, & Cohan, 2013). Thus, once the residuals for both components were calculated, they were correlated with the scores in the ORE20 (see Figure 2e and 2f). We found that scores in the ORE20 were not associated either with the other-race decrement [r = .19, p = .07, CI = -.15 - .39] nor with the own-race advantage [r = -.05, p = .61, CI = -.26 - .16].

In conclusion, although our results show that observers have moderate insights into their recognition abilities for own-race faces, their insights to recognize other-race faces are very limited.

Discussion

In recent years, there has been a growing interest in phenomenological measures of face recognition. The PI20 questionnaire was created for this aim and has been used with normal and prosopagnosic population (Shah, Gaule, et al., 2015). Using this instrument, several studies have shown that human observers have moderate to good insights into their face recognition abilities (Livingston & Shah, 2018; Shah, Sowden, et al., 2015; Ventura et al., 2018). The present study aimed to explore whether observers also have insights into their recognition of other-race faces. For this aim, we created the ORE20, a 20 items self-reported measure of the recognition of other-race faces (Caucasian faces in our case).

Our results show that our observers were better recognizing own-race faces compared other-race faces. This pattern of results replicates the ORE observed in other studies (e.g., Chiroro & Valentine, 1995; Estudillo et al., 2020; Malpass & Kravitz, 1969; Meissner & Brigham, 2001; Wong et al., 2020). Interestingly, the ORE observed in our study was also evident at a phenomenological level as observers reported to be worse recognizing other-race faces compared to own-race faces. We also found a moderate association between our objective measure of face recognition (i.e., the CFMT-Chinese in our case) and the PI20, replicating previous research with a Chinese ethnic population. This association was stronger in magnitude compared to the weak association found between our objective measure of other-race face recognition (i.e., the CFMT-Original) and the ORE20, demonstrating that the self-reported face recognition abilities are stronger for own-races compared to other-race faces. We also found no association between the other-race effect magnitude (calculated by both subtractions and regression methods) and the ORE20.

Our results are in agreement with previous studies showing that observers tend to report that their memory is less clear when they have to recognize other-race faces compared to own-race faces (Smith et al., 2004). In fact, observers' metamnemonic accuracy is also worse for other-race faces (Hourihan et al., 2012). These results, in conjunction with our own results suggests that observers do not only have remarkable problems to recognize other-race faces, but also that their insights for the recognition of other-race faces can be highly misleading.

The weak association found between the recognition of other-race faces (CFMT-Original) and the ORE20 cannot be explained by the psychometric properties of these tests, as our reliability analysis showed that both tests have a strong internal consistency. However, it is important to note that the limited insight to recognize other-race faces contrasts with the moderated-to-good insights observers have into their general face recognition abilities, as shown by others (Livingston & Shah, 2018; Shah, Sowden, et al., 2015; Ventura et al., 2018) and the results of the present study. These general insights into face recognition abilities can also explain the association found between the PI20 and the recognition of other-race faces. Despite the large differences found in terms of accuracy, research has shown that the recognition of own- and other-race faces share around 45% of variance (see e.g., Wan et al., 2017). Thus, the association between the PI20 and the CFMT-Original seem to reflect these general face recognition processes, and not the specific processes of recognizing other-race faces. From these results it could be argued that, compared to the ORE20, the PI20 is a better proxy of other-race face recognition and, therefore, could be potentially used as an index of the recognition of other-race faces. However, one should be cautious about this conclusion, as the magnitude of the association found between other-race face recognition and the PI20 was not different to that found between other-race face recognition and the ORE20.

Our results support theories of metacognition that claims that insights about performance are modulated by experience (Dunning et al., 2003). Specifically, as human observers tend to have extensive experience with own-race faces, they have considerable amount of opportunities to evaluate their face recognition skills for own-race face. The lack of experience with other-race faces, thus, limits their evaluation of face recognition skills with other-race faces. This lack of experience also explains why our observers are still aware that, in comparison to the recognition of own-race faces, they are worse recognizing other-race faces. Lack of experience with a particular racial group produces problems to discriminate faces from that group. This is such that people without experience with a racial group tend to report that faces from that group tend to look alike (Meissner & Brigham, 2001). Thus, observers would know that they will not be able to recognize other-race faces. For this reason, they will rate worse their recognition abilities for other-race faces compared to own-race faces. Although given the differences between the PI20 and the ORE20 questionnaires, we are cautious about this conclusion, converging evidence from expertise research seems to support this hypothesis, as experts have greater metacognitive awareness of their expertise ability compared to novices (Dunning et al., 2003; Fakcharoenphol et al., 2015; Persky & Robinson, 2017). To determine whether and how experience with other-race faces affects the insights into other-race face recognition, future research should include measures of contact with other races (Zhao, Hayward, & Bülthoff, 2014).

In conclusion, our results show that although observers have moderate insights into their recognition abilities for own-race faces, their insights for other-race face recognition abilities are limited. These results have important consequences in applied face recognition settings. For example, in forensic settings, such as line-up identification parades, it might be necessary to assess an eyewitness ability to recognize faces. In these scenarios, self-reported measures of other-race recognition could produce misleading results. For this reason, when evaluating eyewitness' accuracy identifying other-race faces, objective measures should be employed. Although, self-reported measures of own-race faces could provide some indication of observers' actual face recognition abilities, at the best, these measures should be used as a complement of objective face recognition measures.

Funding

The author received no specific funding for this work.

Conflict of Interest

The author declares no conflict of interest.

Ethical Approval

This study has been run following the Code of Ethics and Conduct authorized by American Psychological Association and it was approved by the ethics committee of the University of Nottingham Malaysia

Informed Consent

Participants gave their informed consent and were debriefed at the end of the study.

References

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., … Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(1), 22. https://doi.org/10.1186/s41235-018-0116-5

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, *72*, 872–881. https://doi.org/10.1177/1747021818776145

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The Influence of Race on Eyewitness Memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 257–281). Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116.

Chiroro, P., & Valentine, T. (1995). An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

DeGutis, J., Mercado, R. J., Wilmer, J., Rosenblatt, A., Meissner, C., Brigham, J., … Curran, T. (2013). Individual Differences in Holistic Processing Predict the Own-Race Advantage in Recognition Memory. *PLoS ONE*, *8*(4), e58253. https://doi.org/10.1371/journal.pone.0058253

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and Eyewitness Identifications: The Cross-Race Effect, Decision Time and Accuracy. *Applied Cognitive Psychology*, *30*, 113–125. https://doi.org/10.1002/acp.3178

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.07.001

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Estudillo, A. J. (2012). Facial Memory: The Role of the Pre-Existing Knowledge in Face Processing and Recognition. *Europe's Journal of Psychology*, *8*(2), 231–244. https://doi.org/10.5964/ejop.v8i2.455

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *I-Perception*, *5*(7), 589–601. https://doi.org/10.1068/i0669

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*, 270–276. https://doi.org/10.1002/acp.3609

Evans, J. R., Marcon, J. L., & Meissner, C. A. (2009). Cross-racial lineup identification: assessing the potential benefits of context reinstatement. *Psychology, Crime & Law*, *15*(1), 19–28. https://doi.org/10.1080/10683160802047030

Fakcharoenphol, W., Morphew, J. W., & Mestre, J. P. (2015). Judgments of physics problem difficulty among experts and novices. *Physical Review Special Topics - Physics Education Research*, *11*(2), 020128. https://doi.org/10.1103/PhysRevSTPER.11.020128

Grabman, J. H., Dobolyi, D. G., Berelovich, N. L., & Dodson, C. S. (2019). Predicting High Confidence Errors in Eyewitness Memory: The Role of Face Recognition Ability, Decision-Time, and Justifications. *Journal of Applied Research in Memory and Cognition*, *8*(2), 233–243. https://doi.org/10.1016/J.JARMAC.2019.02.002

Hittner, J. B., May, K., & Silver, N. C. (2003). A monte carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, *1*(3), 158–162. https://doi.org/10.1016/J.JARMAC.2012.06.004

Kokje, E., Bindemann, M., & Megreya, A. M. (2018). Cross-race correlations in the abilities to match unfamiliar faces. *Acta Psychologica*, *185*, 13–21. https://doi.org/10.1016/J.ACTPSY.2018.01.006

Livingston, L. A., & Shah, P. (2018). Article Commentary: People with and without Prosopagnosia Have Insight into Their Face Recognition Ability. *Quarterly Journal of Experimental Psychology*, *71*, 1260–1262. https://doi.org/10.1080/17470218.2017.1310911

Longmore, C. a, Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 77–100. https://doi.org/10.1037/0096-1523.34.1.77

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

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., … Caldara, R. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLoS ONE*, *7*(10), e47956. https://doi.org/10.1371/journal.pone.0047956

Megreya, A. M., White, D., & Burton, A. M. (2011). The Other-Race Effect does not Rely on Memory: Evidence from a Matching Task. *Quarterly Journal of Experimental Psychology*, *64*(8), 1473–1483. https://doi.org/10.1080/17470218.2011.575228

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. https://doi.org/10.1037/1076-8971.7.1.3

Nakashima, S. F., Ukezono, M., Sudo, R., Nunoi, M., Kitagami, S., Okubo, M., … Takano, Y. (2020). Development of a Japanese version of the 20-item prosopagnosia index (PI20-J) and examination of its reliability and validity. *The Japanese Journal of Psychology*, *90*(6), 603–613. https://doi.org/10.4992/jjpsy.90.18235

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin and Review*, *24*(6), 1774–1784. https://doi.org/10.3758/s13423-017-1242-7

Persky, A. M., & Robinson, J. D. (2017). Moving from Novice to Expertise and Its Implications for Instruction. *American Journal of Pharmaceutical Education*, *81*(9), 6065. https://doi.org/10.5688/ajpe6065

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour-accuracy relationships held by police, lawyers and mock-jurors. *Psychiatry, Psychology and Law*, *6*(1), 97–103. https://doi.org/10.1080/13218719909524952

Robertson, D. J., Noyes, E., Dowsett, A., Jenkins, R., Burton, A. M., & Burton, M. (2016). Face recognition by Metropolitan Police Super-recognisers Correspondence to. *PLoS ONE*, *11*, e0150036. https://doi.org/10.1371/journal.pone.0150036

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*(3), 147–165. https://doi.org/10.1037/law0000203

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 1–11. https://doi.org/10.1098/rsos.140343

Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20 item prosopagnosia index (PI20): Relationship with the Glasgow face-matching test. *Royal Society Open Science*, *2*(11), 0–5. https://doi.org/10.1098/rsos.150305

Silver, C., Hittner, J., & May, K. (2004). Testing Dependent Correlations with Nonoverlapping Variables: A Monte Carlo Simulation. *The Journal of Experimental Education*, *73*(1), 53–69.

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of Eyewitness Errors: Can False Identifications Be Diagnosed in the Cross-Race Situation? *Psychology, Public Policy, and Law*, *7*(1), 153–169. https://doi.org/10.1037/1076-8971.7.1.153

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do They All Look Alike? An Exploration of Decision-Making Strategies in Cross-Race Facial Identifications. *Article in Canadian Journal of Behavioural Science*, *36*(2), 146–154. https://doi.org/10.1037/h0087225

Tanaka, J. W., Heptonstall, B., & Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Visual Cognition*, *21*(9–10), 1183–1201. https://doi.org/10.1080/13506285.2013.826315

Ventura, P., Livingston, L. A., & Shah, P. (2018). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item Prosopagnosia Index in a Portuguese sample. *Quarterly Journal of Experimental Psychology*, *71*, 2677–2679. https://doi.org/10.1177/1747021818765652

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., Mckone, E., … Mckone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*(1), 102–122. https://doi.org/10.1037/xge0000249

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport Officers' Errors in Face Matching. *PLoS ONE*, *9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

Wixted, J. T., & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. https://doi.org/10.1177/1529100616686966

Wong, H. K., Stephen, I. D., & Keeble, D. R. T. (2020). The Own-Race Bias for Face Recognition in a Multiracial Society. *Frontiers in Psychology*, *11*, 208. https://doi.org/10.3389/fpsyg.2020.00208

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61–69. https://doi.org/10.1016/j.visres.2014.09.006

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition*, *203*, 104345. https://doi.org/10.1016/j.cognition.2020.104345