

# The importance of internal and external features in recognizing faces that vary in familiarity and race

**Perception**

2022, Vol. 51(11) 820–840

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/03010066221122299

[journals.sagepub.com/home/pec](https://journals.sagepub.com/home/pec)**Menahal Latif  and Margaret C. Moulson**

Department of Psychology, Toronto Metropolitan University (Formerly Ryerson University), Canada

## Abstract

Familiar and unfamiliar faces are recognized in fundamentally different ways. One way in which recognition differs is in terms of the features that facilitate recognition: previous studies have shown that familiar face recognition depends more on internal facial features (i.e., eyes, nose and mouth), whereas unfamiliar face recognition depends more on external facial features (i.e., hair, ears and contour). However, very few studies have examined the recognition of faces that vary in both familiarity and race, and the reliance on different facial features, whilst also using faces that incorporate natural within-person variability. In the current study, we used an online version of the card sorting task to assess adults' ( $n = 258$ ) recognition of faces that varied in familiarity and race when presented with either the whole face, internal features only, or external features only. Adults better recognized familiar faces than unfamiliar faces in both the whole face and the internal features only conditions, but not in the external features only condition. Reasons why adults did not show an own-race advantage in recognition are discussed.

## Keywords

face familiarity, card sorting, internal features, external features, other-race effect

Date Received: 16 March 2022; accepted: 1 August 2022

## Introduction

Faces provide a wealth of information about other people, including their identity, gender, approximate age, ethnic group, emotional state, and focus of attention. The ability to perceive this information quickly and accurately is a remarkable accomplishment of the adult human brain. Adults are face experts—we

---

### Corresponding author:

Menahal Latif, Department of Psychology, Toronto Metropolitan University (Formerly Ryerson University), Toronto, Ontario M5B 2K3, Canada.

Email: [menahal.latif@ryerson.ca](mailto:menahal.latif@ryerson.ca)

remember thousands of individual faces (Jenkins et al., 2011), we use specialized processing mechanisms that apply to faces but not other visual stimuli (e.g., holistic processing; Rossion, 2013), and we have neural systems that are preferentially tuned to faces compared to other objects (Grill-Spector et al., 2017).

Adult face expertise is not, however, applied equally across different faces that we encounter. Both children and adults can more accurately recognize familiar faces when compared to newly encountered unfamiliar faces (Baker et al., 2017; Ellis et al., 1979). When recognizing familiar faces, people can tolerate high amounts of variation in appearance and setting (Johnston & Edmonds, 2009). In contrast, unfamiliar face recognition is disrupted by variations in appearance, even when images contain shared characteristics (Baker et al., 2017; Hancock et al., 2000; Sandford & Ritchie, 2021). This difference in tolerance for within-person variation is observed using the card sorting task (Jenkins et al., 2011; Kramer et al., 2018; Laurence et al., 2016). In this task, participants are given a set of 40 images of two identities and are asked to sort these images into piles representing each identity. However, participants are unaware of the number of identities present within the set. The multiple images of a single individual presented in the set test tolerance for within-person variability (i.e., recognition), whereas the multiple identities in the set test sensitivity to between-person variability (i.e., discrimination). Jenkins et al. (2011) found that participants created around 7.5 piles in the card sorting task for unfamiliar faces, indicating that they perceived significantly more distinct identities than the actual number present in the set (i.e., two identities). Although recognition across variability is difficult, this difficulty is minimized when presented with familiar faces when compared to unfamiliar faces. Zhou and Mondloch (2016) had participants complete card sorting tasks with either familiar or unfamiliar faces. When the faces were familiar celebrities, participants formed two piles (i.e., accurately perceived two identities), whereas when the faces were unfamiliar celebrities, participants formed about six piles (i.e., perceived six identities).

Moreover, familiar face recognition consists of increased reliance on internal features (eyes, nose, and mouth; Di Oleggio Castello et al., 2017; Ellis et al., 1979), whereas unfamiliar face recognition consists of increased reliance on external features (Bonner et al., 2003; Liu et al., 2013; Want et al., 2003). Past literature suggests that as familiarity increases there is more reliance on internal features and less reliance on external features, leading to an internal feature advantage for familiar face recognition (Bonner & Burton, 2004; Bonner et al., 2003; Campbell et al., 1995; Ellis et al., 1979; Megreya & Bindemann, 2009). Ellis et al. (1979) presented adult participants with celebrity faces and asked them to write names of the faces they recognized. Participants showed an internal feature advantage such that when faces were modified so that only internal features were available, they recognized 50% of the target faces, but when only external features were available, they recognized 30% of the target faces.

However, not all studies have found differential reliance on internal features for familiar versus unfamiliar faces. Using a card sorting task, Kramer et al. (2018) found impaired performance (i.e., an increase in number of identities perceived) when faces were manipulated such that only the external features were visible compared to when only the internal features were visible, suggesting an internal feature advantage. However, this pattern was found for both familiar and unfamiliar faces, which suggests that internal and external features may be equally informative/uninformative for both familiar and unfamiliar face recognition. In this study, however, the external feature condition contained images with both the external features and the background (i.e., varied settings), whereas the internal features condition contained images with only the internal features and no background. This difference in the background may have introduced more variability in the external feature condition leading to poorer performance. Thus, there is mixed evidence about the differential reliance on internal versus external features for recognition of faces that differ in familiarity.

Decades of research suggest that familiarity with an individual face can greatly influence recognition accuracy. Similarly, the category to which a face belongs influences its likelihood of being recognized. Adults show poorer recognition of faces that belong to other racial groups when

compared to faces of their own racial groups (for a review, see Meissner & Brigham, 2001), a perceptual phenomenon known as the other-race effect, cross-race effect, own-race advantage, or own-race bias. Individuals from primarily monoracial societies exhibit a robust other-race effect as most people frequently encounter individuals of their own race, whereas encounters with individuals of other races are relatively less frequent. Hancock and Rhodes (2008) found that a higher level of inter-group contact is associated with fewer recognition errors for other-race faces. Recent work suggests that the other-race effect also results from increased difficulty with recognition across variability (Laurence et al., 2016; Tüttenberg & Wiese, 2019; Zhou & Mondloch, 2016). Using the card sorting task, Laurence et al. (2016) found that both East Asian and White participants created, on average, five piles for own-race unfamiliar faces and twice the number of piles for other-race unfamiliar faces, demonstrating that recognition across variability is more difficult for other-race faces.

Researchers have also examined the use of internal and external features for recognizing own-versus other-race faces. Megreya and Bindemann (2009) used a face matching task with British and Egyptian participants. Participants were required to match faces with images containing either the external features only or the internal features only. They found that Egyptian participants matched both Egyptian and British unfamiliar faces more accurately using internal features. In direct contrast, British participants matched both Egyptian and British unfamiliar faces more accurately using external features. This finding supports previous literature that suggests a reliance on external features for unfamiliar faces in the British participants (Campbell et al., 1995; Ellis et al., 1979). Contrary to previous findings, Egyptian adults did not show a reliance on external features for unfamiliar face matching, likely because they have extensive experience recognizing faces wearing headscarves, which leads to an internal feature advantage that allows them to recognize even unfamiliar faces using internal features alone (i.e., the ‘headscarf effect’) (Megreya & Bindemann, 2009; Suhkre et al., 2015). Thus, it is important to assess the differential reliance on features for faces that vary in race, in a population without an overall internal feature advantage. Harvard (2021) tested White and East Asian participants on a face-matching task and found significantly better matching performance for other-race faces when external features were presented along with internal features. Similarly, Wong et al. (2020) tested Malaysian–Malay, Malaysian–Chinese, Malaysian–Indian, and White participants and found significantly better recognition performance for other-race faces when external features were presented along with internal features, and better recognition performance for own-race faces when internal features were presented alone. Thus, these findings suggest that different mechanisms may be involved in processing own- and other-race faces such that we rely more on internal features for own-race faces, whereas we rely on both internal and external features for other-race faces. Although Wong et al. (2020) investigated differential recognition performance with the presence and absence of external features for own- and other-race faces, they did not assess recognition performance with the presence and absence of internal features. Thus, it is important to consider the extent to which the absence of internal features can affect other-race face recognition.

Despite decades of research investigating familiar and unfamiliar face recognition, as well as the other-race effect, very few studies have investigated their interaction using images containing within-person variability (Tüttenberg & Wiese, 2019). Zhou and Mondloch (2016) used a card sorting task to assess the recognition of faces that varied in familiarity and race across natural variability. East Asian participants showed a typical other-race effect for unfamiliar faces: they created twice as many piles for other-race faces when compared to own-race faces. However, the other-race effect was eliminated during familiar face recognition: participants created an equal number of piles for familiar other-race and familiar own-race faces, most often arriving at the correct solution (i.e., two piles; Zhou & Mondloch, 2016). Although this study examined the differential recognition of faces that varied in familiarity and race and identified the absence of the other-race effect during

familiar face recognition, it did not investigate the underlying reasons why recognition differs for faces that vary along these two dimensions. One such reason may be the differential reliance on internal and external features.

In the current study, we investigated the differential use of internal and external features in adult face recognition and examined whether this effect differs for faces that vary in familiarity and race. In this study, adults completed an online card sorting task with own-race familiar, own-race unfamiliar, other-race familiar, or other-race unfamiliar faces. These faces were modified to contain only internal features, only external features, or the whole face. We hypothesized that face familiarity would influence the differential reliance on particular features for face recognition: specifically, we predicted there would be greater reliance on internal features for familiar faces and greater reliance on external features for unfamiliar faces. Thus, our first hypothesis was that participants would create fewer piles (i.e., perceive fewer identities) for familiar faces in the internal feature condition when compared to the external feature condition; in contrast, participants would create fewer piles (i.e., perceive fewer identities) for unfamiliar faces in the external feature condition when compared to the internal feature condition. Our second hypothesis was that there would be a three-way interaction between feature, familiarity, and race, such that there would be performance differences for recognition of own- and other-race unfamiliar faces in the internal versus external feature conditions. Specifically, we predicted that in the internal feature condition participants would create fewer piles (i.e., perceive fewer identities) for own-race unfamiliar than other-race unfamiliar faces; in the external feature condition, participants would create fewer piles (fewer identities) for other-race unfamiliar faces than for own-race unfamiliar faces.

## Method

### Participants

White North American adults ( $n = 258$ ;  $M$  age = 26.9 years,  $SD = 6.5$ ; 159 women, 86 men, 11 non-binary, 1 genderqueer, 1 not specified) were recruited from an undergraduate psychology research pool and Prolific, an online recruitment platform. Only White individuals (254 reported White, 2 reported White and Indigenous and 2 reported White and East Asian) were recruited for this study because of the limited stimuli that were available to serve as own-race and other-race faces. Additionally, Prolific recruitment was restricted to adults between the ages of 18 and 40 from the United States and Canada to better match the undergraduate sample. Undergraduate participants ( $n = 74$ ) received course credit and Prolific participants ( $n = 184$ ) received \$5.20 as compensation. PANGEA (Westfall, 2015) indicated that 240 participants in total would provide enough power to detect a medium effect size (considering the main effects of Face Familiarity, Face Race, and Features, and the interaction between the three factors, power = 0.809, Cohen's  $d = 0.45$ ,  $p$ -value = .05). A medium effect size was expected based on the findings of past literature with similar research questions (Kramer et al., 2018; Want et al., 2003; Zhou & Mondloch, 2016). An additional 89 participants completed the study but were excluded from data analysis due to (1) reporting ethnicity other than White on a demographic questionnaire ( $n = 22$ ); (2) reporting that they were familiar with any of the identities in the unfamiliar condition or that they were unfamiliar with any of the identities in the familiar condition ( $n = 41$ ); (3) failing to complete the sorting task or to form piles ( $n = 18$ ); or (4) completing the incorrect familiarity questionnaire, such that we could not determine their familiarity with the identities seen in the sorting task ( $n = 8$ ).

### Stimuli

The images used in this study were of celebrities chosen to be familiar or unfamiliar to the participants. Images of White American (familiar own-race), Black American (familiar other-race), White

Dutch (unfamiliar own-race), and Black British (unfamiliar other-race) female celebrities were sourced from Google Images. All images were high quality and showed the face in frontal view. In order to choose celebrities that were high and low in familiarity with our population of interest (young adults in North America), a separate online validation study ( $n = 40$  adults) was conducted through Qualtrics. Participants were shown images of numerous Black and White celebrity faces one at a time and asked, ‘How familiar is this face?’ They reported their answer on a scale of 1 (not at all familiar) to 7 (highly familiar). Next, they were asked to provide identifying information for each face. Only celebrities whose average rating was 1–2 and for whom participants did not provide identifying information (i.e., name, occupation) were chosen as unfamiliar stimuli. Only celebrities whose average rating was 6–7 and for whom the majority (80% and above) of participants provided identifying information were chosen as familiar stimuli. Based on these ratings, we chose four identities for each of the four conditions (own-race familiar; own-race unfamiliar; other-race familiar; other-race unfamiliar). These four identities were divided into two pairs, to create two sets of stimuli for each condition (see Table 1 for listing of pairs in each condition).

For each identity, we sourced 20 colour images of the individual from Google Images. In these 20 images, the face varied naturally in appearance. Images were cropped to  $400 \times 400$  pixels, and so that the face filled most of the frame. Each set of images for an identity was manipulated using Adobe Photoshop CS6 to produce whole face, internal feature, and external feature versions. The internal feature images were a modified version of Kramer et al. (2018); instead of cropping out the internal features and pasting them onto a white background, a white rectangle-shaped mask was used to disguise the external features (hair, face contour) whilst keeping the eyes, eyebrows, nose, and mouth, as well as the background information, visible. This maintained consistency with the external feature images, in that both sets of images retained background information. The external feature images contained an oval mask (approximately  $100 \times 120$  px) that covered the internal features, whilst keeping the ears, hair, face contour and background information visible (see Figure 1, e.g., of the stimuli).

## Procedure

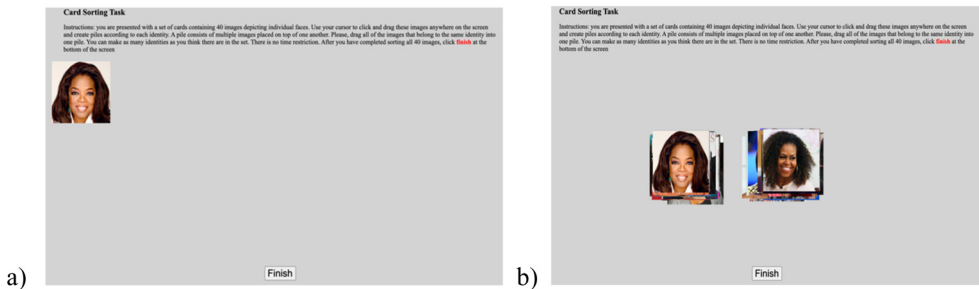
We created an online version of the card sorting task (Jenkins et al., 2011) using JavaScript, PHP, and a local server (accessed through Cyberduck; a Libre server and cloud storage browser). The online task was linked to a Qualtrics survey containing questionnaires associated with the study. Participants were randomly assigned to one of four conditions: own-race (White) familiar, own-race unfamiliar, other-race (Black) familiar, or other-race unfamiliar. They saw one of three feature conditions: whole face, external features only or internal features only. After completing the consent form through Qualtrics, participants were provided with detailed instructions on how to complete the card sorting task. Once participants read the instructions, they were provided with an external URL to complete the task. Like the in-person version of the card sorting task, in the online version, participants were provided with a set of 40 shuffled photographs (i.e., using JavaScript, images were automatically randomized at the start of each new session) of two

**Table 1.** List of identities in each condition of the card sorting task.

| Card sort conditions  | Card set 1 identities               | Card set 2 identities              |
|-----------------------|-------------------------------------|------------------------------------|
| Familiar own-race     | Emma Watson and Anne Hathaway       | Miley Cyrus and Jennifer Lawrence  |
| Unfamiliar own-race   | Marije Zuurveld and Davina Michelle | Ilse DeLange and Victoria Koblenko |
| Familiar other-race   | Rihanna and Raven Symone            | Oprah and Michelle Obama           |
| Unfamiliar other-race | Camilla Beeput and Freema Agyeman   | Riann Steele and Pearl Mackie      |



**Figure 1.** This is not one of the identities used in the study because publication of the images used in the study is restricted. Examples of the feature conditions: (a) whole face image, (b) external features only image, and (c) internal features only image. Stimulus images were sourced from Google Images and modified in Photoshop CS6.



**Figure 2.** Online card sorting task: (a) card sorting task with instructions at the top of the screen and a shuffled set of cards at the top left of the screen and (b) example of a card set of 40 images sorted into 2 piles.

unique individuals (20 photos of Person A and 20 photos of Person B). This card set was presented at the top left of the screen (Figure 2), with the 40 images placed on top of one another. Participants were instructed to sort the cards into piles representing each identity present in the set and to take as much time as they needed. They were not told how many identities were present in the set. Participants used their cursor to click on the first image, drag the image (revealing a new image underneath), and move it anywhere on the screen (in the grey space) to form a pile. They repeated this step until all 40 images were sorted into piles such that images representing the same identity were stacked over one another. Participants could also move images around after the initial placement (e.g., if they decided that a previous image represented a different identity, they could go back to that image and move it to create a new pile). Once participants were done sorting the images, they clicked the ‘finish’ button at the bottom of the screen and returned to Qualtrics to complete the study questionnaires. The ‘finish’ button was coded in HTML to save participant coordinate data on Cyberduck for all images sorted by each participant. This coordinate data represented the location of each image on the screen upon completion of the task, thus reflecting the number of piles created by each participant.

Participants completed three questionnaires. The familiarity questionnaire assessed their familiarity with the identities of the individuals in the card sets. Participants saw images of the celebrity faces used in the card sorting task and were asked ‘Are you familiar with the above individual?’

After indicating familiarity with each face with a ‘yes’ or ‘no’ answer, participants were then asked to provide more information about that individual (i.e., their name or occupation) if they knew it. A demographic questionnaire gathered information about the participant’s age, ethnicity, and gender, as all of these characteristics have been known to influence face perception (Ellis et al., 1979; Laurence et al., 2016), as well as where the participants had lived and for how many years, since early intergroup contact has been associated with a weaker other-race effect (McKone et al., 2021; Zhou et al., 2019). Lastly, the participant completed the Intergroup Relations Contact Quantity Contact Quality (CQCQ) scale (Islam & Hewstone, 1993). This is a validated questionnaire used to measure the level of contact between social groups. It includes several questions about the participant’s current intergroup interactions at university, as neighbours, and as close friends. Participants in the own-race condition completed the CQCQ-White scale, which measured contact with White individuals, whereas participants in the other-race condition completed the CQCQ-Black scale, which measured contact with Black individuals. Results from the CQCQ scale can be found in supplemental material.

### *Data Analysis Approach*

Data from both the pilot study and the main study were collapsed across the two face pairs in each condition and were analysed using three metrics of performance: number of piles (i.e., perceived identities), sensitivity ( $d'$ ), and response criterion ( $c$ ). To calculate the number of perceived identities, we simply counted the final number of piles (i.e., groups of images) created by each participant at the end of the card sorting task. To calculate the signal detection metrics ( $d'$  and  $c$ ), we calculated the proportions of ‘different person/same group’ errors (sorting images of different people into the same pile) and ‘same person/different group’ errors (sorting images of the same person into different piles). The different person/same group error rate, also known as misses, is calculated by identifying the piles that consist of two identities, then multiplying the image frequency of each identity together (i.e., the number of images of identity A multiplied by the number of images of identity B placed in each pile), summing these products to calculate the total number of misses, and dividing this value by the total number of different-person pairs (400 pairs for a card sorting task with 2 identities and 20 images per identity). The same person/different group error rate, also known as false alarms, is calculated by identifying the piles with identity A and the piles with identity B. For each identity, we calculate the product of all pairs (i.e., the number of images of identity A in one pile multiplied by the number of images of identity A in each of the other piles) and sum the products to calculate the total number of false alarms. We then divide this total number of errors by the total number of same-person pairs (380 pairs for a card sorting task with 2 identities and 20 images per identity). Using these two types of error, sensitivity ( $d'$ ) can be calculated using the formula  $[z(\text{Hits}) - z(\text{False alarms})]$ , where Hits equals  $1 - \text{‘different person/same group’ error rate}$  and False alarms equals  $\text{‘same person/different group’ error rate}$ . Response criterion ( $c$ ) can be calculated using the formula  $-0.5 * [z(\text{Hits}) - z(\text{False alarms})]$  (Balas & Pearson, 2017).

## **Results**

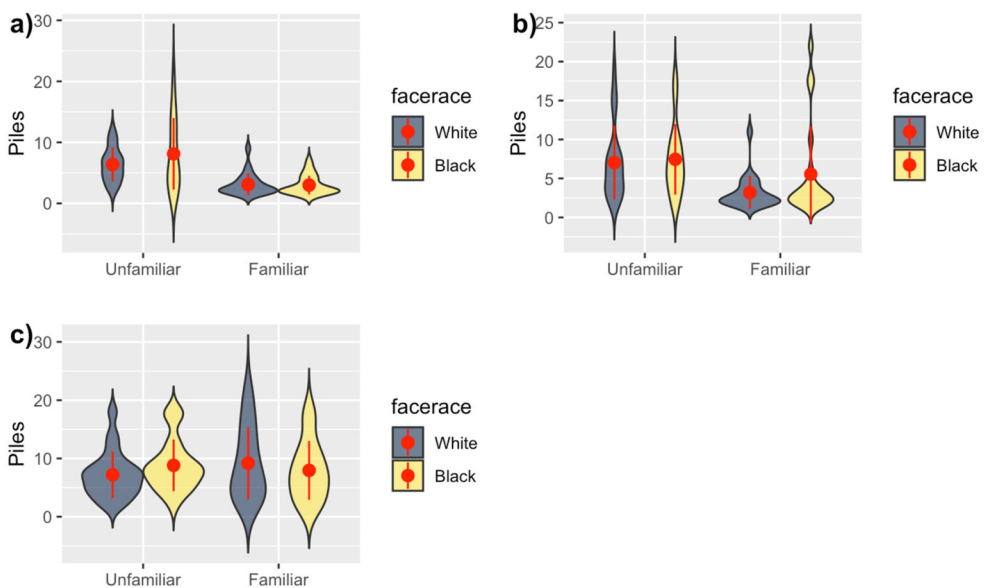
### *Pilot Study*

Results from the pilot study can be found in supplemental material. Overall, participants performed similarly in this online card sorting task compared to previous in-person card sorting tasks conducted both in our lab and other labs. Participants tended to create more piles than there were identities present in the set, and the average numbers of piles created were similar to previous literature.

## Main Study

To identify outliers and their effect on the relationship between variables, we conducted the Bonferroni Outlier Test to automatically compute extreme scores. This test revealed two outliers. The removal of these outliers did not influence the assumption violations. Thus, outliers were included in the analyses. For all analyses we report, we conducted tests of normality, sphericity, and homogeneity of variance. In cases where the Mauchly's test of Sphericity revealed significant violations, we applied the Greenhouse–Geisser correction.

**Number of Piles.** Data for the number of piles formed were analysed using a  $2 \times 2 \times 3$  between-subjects analysis of variance, where familiarity (familiar, unfamiliar), face race (own-race, other-race), and feature condition (whole, external and internal) varied between participants. We found a significant main effect of familiarity,  $F(1, 246) = 16.72, p < .001, \eta_p^2 = .07$ , with fewer piles created for familiar faces ( $M = 5.28, SD = 4.86$ ) compared to unfamiliar faces ( $M = 7.53, SD = 4.53$ ). We also found a significant main effect of feature condition,  $F(2, 246) = 11.54, p < .001, \eta_p^2 = .09$ : significantly more piles were created for external feature images ( $M = 8.30, SD = 4.94$ ) compared with both internal feature ( $M = 5.86, SD = 4.82, t(166) = 3.25, p = .001$ , and whole face images ( $M = 5.26, SD = 4.18, t(161) = 4.34, p < .001$ ). There was no significant difference between performance on internal feature images and whole face images,  $t(168) = .88, p = .38$ . There was no main effect of face race,  $F(1, 246) = 2.25, p = .14, \eta_p^2 = .009$ . Although there was a significant familiarity X feature condition interaction,  $F(2, 246) = 6.63, p = .002, \eta_p^2 = .05$ , the familiarity X face race interaction was not significant,  $F(1, 246) = .70, p = .41, \eta_p^2 = .003$ , nor was the three-way familiarity X face race X feature condition interaction,  $F(2, 246) = 1.76, p = .18, \eta_p^2 = .01$  (Figure 3).

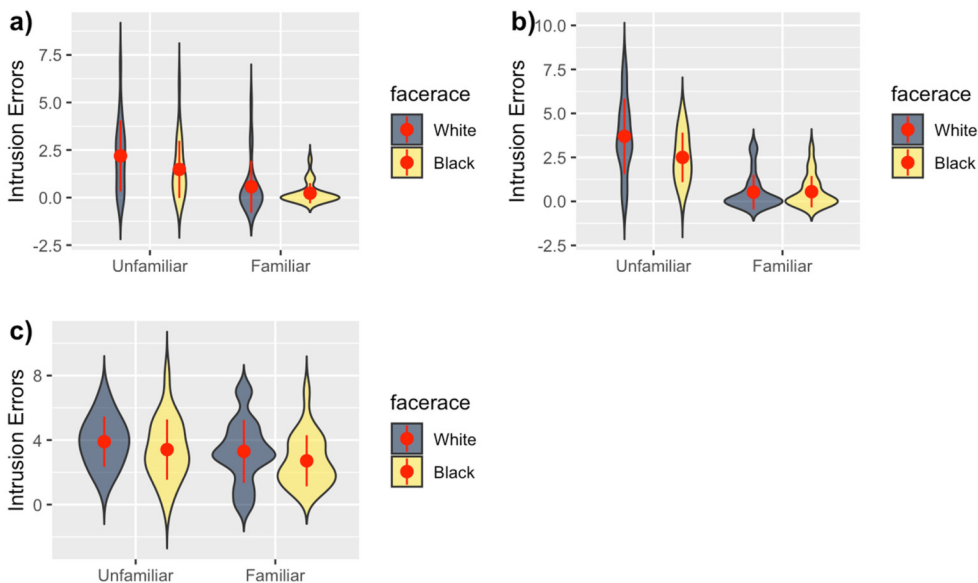


**Figure 3.** These violin plots present the number of piles created in (a) whole face condition; (b) internal features only condition; and (c) external features only condition. The whiskers (shown as the red lines in the centre of the plot) represent the SD. The bolded points (shown in red) are the mean number of piles created in the respective card sorting condition.



To further test our hypotheses, we carried out planned comparison  $t$ -tests to break down the significant two-way interaction between familiarity X feature condition. Planned comparisons revealed that for familiar faces, participants created fewer piles in both the whole face ( $M = 3.05$ ),  $t(47) = -6.07$ ,  $p < .001$ , and the internal features only conditions ( $M = 4.34$ ),  $t(77) = -3.71$ ,  $p = .0004$  when compared to the external features only condition ( $M = 8.56$ ). No difference was found in the number of piles created for familiar faces in the whole face condition than in the internal features only condition,  $t(49) = -1.68$ ,  $p = .099$ . In contrast, for unfamiliar faces, there were no statistically significant differences among the three feature conditions in the number of piles created. Thus, participants perceived fewer identities in the internal features only condition than in the external features only condition for familiar faces (Hypothesis 1). For unfamiliar faces, participants perceived a similar number of identities across all feature conditions.

**Intrusion Errors.** For the number of intrusion errors, we found a significant main effect of familiarity,  $F(1, 246) = 63.56$ ,  $p < .001$ ,  $\eta_p^2 = .04$ , with fewer errors created for familiar faces ( $M = 1.30$ ,  $SD = 1.76$ ) compared to unfamiliar faces ( $M = 2.83$ ,  $SD = 1.93$ ). We also found a significant main effect of race,  $F(1, 246) = 8.80$ ,  $p < .01$ ,  $\eta_p^2 = .01$ : significantly more errors were made for White faces ( $M = 2.37$ ,  $SD = 2.17$ ) compared to Black faces ( $M = 1.82$ ,  $SD = 1.78$ ). We also found a significant main effect of feature condition,  $F(2, 246) = 45.64$ ,  $p < .001$ ,  $\eta_p^2 = .06$ : significantly more errors were made for external feature images ( $M = 3.33$ ,  $SD = 1.77$ ) compared with both internal feature ( $M = 1.88$ ,  $SD = 1.99$ ),  $t(163) = 4.87$ ,  $p < .001$ , and whole face images ( $M = 1.24$ ,  $SD = 1.57$ ),  $t(164) = -8.55$ ,  $p < .001$ . There was also a significant difference between performance on internal feature images and whole face images,  $t(160) = 2.82$ ,  $p < .01$ . Although there was a significant familiarity X feature condition interaction,  $F(2, 246) = 8.21$ ,  $p < .001$ ,  $\eta_p^2 = .06$ , the familiarity X face race interaction was not significant,  $F(1, 246) = 2.01$ ,  $p = .16$ ,  $\eta_p^2 = .00$ , nor was the three-way familiarity X face race X feature condition interaction,  $F(2, 246) = .99$ ,  $p = .37$ ,  $\eta_p^2 = .01$  (Figure 4).



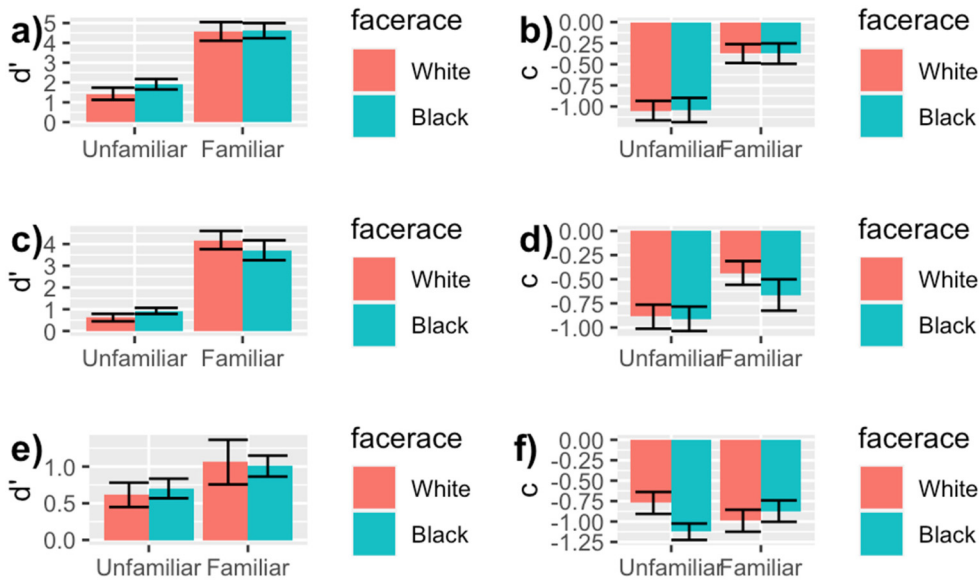
**Figure 4.** These violin plots present the number of intrusion errors created in (a) whole face condition; (b) internal features only condition; and (c) external features only condition. The whiskers (shown as the red lines in the centre of the plot) represent the SD. The bolded points (shown in red) are the mean number of piles created in the respective card sorting condition.

To further test our hypotheses, we carried out planned comparison *t*-tests to break down the significant two-way interaction between familiarity X feature condition. Planned comparisons revealed that for familiar faces, participants created fewer errors in both the whole face ( $M = .41$ ),  $t(65) = -8.16$ ,  $p < .001$ , and the internal features only conditions ( $M = .55$ ),  $t(61) = 7.85$ ,  $p < .001$  when compared to the external features only condition ( $M = 2.93$ ). No difference was found in the number of errors created for familiar faces in the whole face condition than in the internal features only condition,  $t(80) = .67$ ,  $p = .25$ . For unfamiliar faces, participants created fewer errors in the whole face condition ( $M = 1.84$ ) when compared to the internal features only ( $M = 3.18$ ),  $t(86) = 3.55$ ,  $p < .001$ , and external features only conditions ( $M = 3.68$ ),  $t(83) = -4.98$ ,  $p < .001$ . No difference was found in the number of errors created for unfamiliar faces in the external features only condition than in the internal features only condition,  $t(83) = 1.29$ ,  $p = .10$ . Thus, participants made fewer recognition errors in the internal features only condition than in the external features only condition for familiar faces (Hypothesis 1). For unfamiliar faces, participants made a similar number of recognition errors across both the internal features only and the external features only conditions.

**Sensitivity Analysis ( $d'$ ).** For sensitivity ( $d'$ ), there was a main effect of familiarity,  $F(1, 246) = 154.09$ ,  $p < .001$ ,  $\eta_p^2 = .27$ , and a main effect of feature condition,  $F(2, 246) = 58.16$ ,  $p < .001$ ,  $\eta_p^2 = .20$ , but no main effect of face race,  $F(1, 246) = .21$ ,  $p = .65$ ,  $\eta_p^2 = .00$ . The familiarity X feature condition interaction was statistically significant,  $F(2, 246) = 25.74$ ,  $p < .001$ ,  $\eta_p^2 = .09$ , whereas both the familiarity X face race interaction,  $F(1, 246) = 1.94$ ,  $p = .17$ ,  $\eta_p^2 = .003$ , and the familiarity X face race X feature condition interaction,  $F(2, 246) = .27$ ,  $p = .76$ ,  $\eta_p^2 = .00$  were not.

Lower sensitivity was found for the unfamiliar conditions ( $M = 1.05$ ,  $SD = 1.09$ ) compared with familiar conditions ( $M = 3.22$ ,  $SD = 2.31$ ), and for external feature images ( $M = .84$ ,  $SD = .89$ ) compared to both internal feature ( $M = 2.29$ ,  $SD = 2.17$ ),  $t(114) = -5.70$ ,  $p < .001$ , and whole face images ( $M = 3.10$ ,  $SD = 2.21$ ),  $t(118) = -8.85$ ,  $p < .001$ . As well, sensitivity was greater for whole face images than internal feature images,  $t(173) = -2.44$ ,  $p = .02$ . We conducted planned comparison *t*-tests to break down the two-way interaction between familiarity X feature condition. Planned comparisons revealed that for familiar faces, there was greater sensitivity in both the whole face,  $t(65) = 10.49$ ,  $p < .001$ , and the internal features only conditions,  $t(61) = 8.38$ ,  $p < .001$  when compared to the external features only condition, with no difference between the whole face and the internal features only conditions,  $t(82) = 1.50$ ,  $p = .14$ . For unfamiliar faces, greater sensitivity was found in the whole face condition when compared to both the internal features only,  $t(70) = 4.02$ ,  $p = .0002$ , and the external features only condition,  $t(67) = 4.56$ ,  $p < .001$ . Sensitivity did not differ between the internal features only and the external features only conditions,  $t(85) = .70$ ,  $p = .48$ , for unfamiliar faces. Overall, participants made fewer recognition errors (false alarms and misses) when viewing familiar faces in the whole face and the internal features only conditions than in the external features only condition. When viewing unfamiliar faces, participants made fewer recognition errors in the whole face condition than in both the internal features only and the external features only conditions.

**Response Criterion ( $c$ ).** For response criterion ( $c$ ), the results were similar to  $d'$  such that there was a main effect of familiarity,  $F(1, 246) = 22.96$ ,  $p < .001$ ,  $\eta_p^2 = .08$ , a main effect of feature condition,  $F(2, 246) = 3.93$ ,  $p = .02$ ,  $\eta_p^2 = .03$ , and a familiarity X feature condition interaction,  $F(2, 246) = 6.25$ ,  $p = .002$ ,  $\eta_p^2 = .04$ . All other effects were not statistically significant. Planned comparisons revealed a more negative response criterion for familiar faces in the external features only condition when compared to both the whole face,  $t(80) = 4.50$ ,  $p < .001$ , and the internal features only conditions,  $t(79) = 2.79$ ,  $p = .007$ , with no difference between the whole face and the internal features only conditions,  $t(78) = 1.33$ ,  $p = .19$ . For unfamiliar faces, response criterion did not differ



**Figure 5.** Summary of the sensitivity scores and response criterion for all of the feature conditions: (a) and (b) whole faces, (c) and (d) internal features only, and (e) and (f) external features only. Error bars represent standard error.

among any of the three feature conditions. Thus, participants assumed greater variability and made more false alarm errors when viewing familiar faces in the external features only condition than in the whole face and the internal features only conditions. When viewing unfamiliar faces, participants did not differ in their tendency to make false alarm errors across all feature conditions. See Figure 5 for a summary of the sensitivity and response criterion findings.

## Discussion

Previous research has extensively examined familiar and unfamiliar face recognition, as well as own- and other-race face recognition; however, few studies have investigated their interaction or their impact on adults' reliance on different features for recognition (Havard, 2021; Wong et al., 2020). Thus, in the current study, we investigated how race and familiarity affect card sorting behaviour when participants are presented with full faces or are limited to only the internal or external features.

Across both the pilot and the main study, we found a robust main effect of familiarity: participants showed better recognition, as measured by the number of piles, number of intrusion errors, and sensitivity ( $d'$ ), of familiar faces when compared to unfamiliar faces. These findings are consistent with previous literature using a range of face recognition tasks (Kramer et al., 2018; Want et al., 2003; Wong et al., 2020). Both children and adults can more accurately recognize familiar faces when compared to newly encountered unfamiliar faces (Baker et al., 2017; Ellis et al., 1979), and are more likely to sort familiar faces into the correct number of piles than unfamiliar faces (Andrews et al., 2015; Jenkins et al., 2011; Kramer et al., 2018; Young & Burton, 2017; Zhou & Mondloch, 2016). That participants created fewer piles for familiar faces indicates better ability to generalize across different images of the same identity. This is consistent with theories of familiar versus unfamiliar face recognition that suggest that changes in pose, expression and context can impair recognition of unfamiliar but not familiar faces (Johnston & Edmonds, 2009).

In the current study, we also found that, overall, participants created fewer piles for images of the full face and for images including only the internal features (i.e., eyes, nose and mouth), regardless of face category, indicating a smaller number of perceived identities for these conditions compared to the condition with images of only the external features (i.e., hair, face contour and ears). This is consistent with previous literature that suggests poorer recognition of faces when only the external features are present (Kramer et al., 2018; Tanaka & Sengco, 1997). However, our analyses of sensitivity ( $d'$ ) revealed a somewhat different story: sensitivity was lower for images containing only internal features compared to the images of the whole face, indicating better face recognition when both internal and external features are made available. Sensitivity was lower for images containing only external features compared to both images containing only internal features and images of the whole face. These results highlight the importance of the identity information conveyed by the external features, since excluding these features (i.e., in the images containing only internal features) led to poorer recognition compared to the whole face. Similarly, both Havard (2021) and Wong et al. (2020) found better recognition of faces when external features were presented along with internal features (i.e., the whole face) when compared to when only the internal features are presented.

We also found a more negative response criterion ( $c$ ) when only the external features were presented when compared to both when only the internal features were presented and when the whole face was presented. There was no difference in response criterion between when only the internal features were presented and when the whole face was presented. Thus, there was a larger bias towards reporting a greater number of identities for the external features only condition compared to the internal features only or whole face condition (Balas & Pearson, 2017).

There was also a two-way interaction between familiarity and feature condition. Participants displayed better performance for familiar faces, as measured by the number of piles and sensitivity ( $d'$ ), for both the images containing only internal features and images of the whole face, than images containing only the external features. Past research suggests an increased reliance on internal features for familiar face recognition (Di Oleggio Castello et al., 2017; Ellis et al., 1979; Megreya & Bindemann, 2009), and an increased reliance on external features for unfamiliar face recognition (Bonner et al., 2003; Liu et al., 2013; Want et al., 2003). Similarly, we found an internal feature advantage for familiar faces such that there was better performance when only the internal features were presented than when only the external features were presented. In contrast, we found no difference in performance for unfamiliar faces when only the internal features were visible and when only the external features were visible. This finding is not consistent with previous literature suggesting increased reliance on external features for unfamiliar face recognition. However, not all studies have found differential reliance on external features for familiar versus unfamiliar faces. Thus, it is possible that external features alone may not contain enough identity information to accurately recognize unfamiliar faces and that adults may require the whole face (including both internal and external features) to improve unfamiliar face recognition. Moreover, past research suggests that as familiarity increases, adults become better equipped to recognize familiar faces even with limited facial information (i.e., using only the internal features).

In the intrusion error analysis, we found a main effect of race. Participants created more errors for own-race faces when compared to other-race faces. However, we did not find a main effect of face race on any other measure, nor did we find any interactions involving race. Participants created similar numbers of piles for own- and other-race faces, indicating a similar number of perceived identities for both groups, and their sensitivity and response criterion for own- and other-race faces did not differ significantly. The direction of the effect of race in the intrusion error analysis is also inconsistent with previous literature. Previous research suggests more recognition errors for other-race faces than own-race faces. The lack of other-race effect in the current data is inconsistent with a large literature demonstrating an other-race effect in face recognition (for a review, see

Meissner & Brigham, 2001), including studies using the card sorting task. Laurence and colleagues (2016) found that both East Asian and White participants created twice the number of piles for other-race unfamiliar faces than own-race unfamiliar faces. Although we observed a similar pattern, where more piles were created for the other-race unfamiliar faces ( $M = 8.12$ ) than for own-race unfamiliar faces ( $M = 6.38$ ), this difference was not significant. Additionally, the lack of interaction between face race and face familiarity in the current study is inconsistent with Zhou and Mondloch (2016), who found that East Asian participants showed a robust other-race effect when sorting unfamiliar faces, but no other-race effect when sorting familiar (celebrity) faces. Finally, the lack of interaction between face race and feature condition is inconsistent with previous literature (Havard, 2021; Wong et al., 2020), who found greater performance for other-race faces when external features were presented along with internal features (i.e., whole face condition), and better recognition performance for own-race faces when internal features were presented alone.

Furthermore, the discrepancy in results between our three metrics of accuracy—number of piles created, number of intrusion errors made and sensitivity scores—reflects the importance of more fine-grained analyses. Because the number of piles metric does not account for both types of recognition error in a card sorting task, and the intrusion errors metric only accounted for one type of error (different person/same group error),  $d'$  is a more sensitive measure of overall recognition performance.

There are several possible explanations for the lack of an other-race effect in the current study. Previous literature suggests that the other-race effect is a statistically small effect, and the face-familiarity effect is a much larger effect found in both face memory and face recognition tasks (Meissner & Brigham, 2001; Rossion & Michel, 2012). Zhou and colleagues (2021) conducted an old/new recognition task and found that the effect size for face-familiarity ( $n^2 = 0.33$ ) was approximately 10 times greater than that for racial group membership ( $n^2 = 0.03$ ). Moreover, individuals with greater exposure to other-race faces (i.e., individuals from multiracial societies) generally show a weaker other-race effect (McKone et al., 2019; Wright et al., 2001; Zhou et al., 2019). Previous studies that show a robust other-race effect recruited samples from primarily monoracial communities. In the current study, almost one-third of participants were undergraduate students living in Greater Toronto Area, a large, highly multiracial city (Statistics Canada, 2016).

Furthermore, the current study is the first to conduct an online card sorting task to assess adult face recognition. The inherent methodological differences between an in-lab and an online card sorting task may result in differences in findings that can be associated with variations in task demands and other environmental variables. Although the current study found comparable results to in-person card sorting tasks (Jenkins et al., 2011; Kramer et al., 2018), previous studies have reported issues with online testing, including minimal control over the research environment, and loss of data or low-quality responses (Kinney, 2001; Lochner, 2016). The card sorting task requires detailed comprehension of task instructions for participants to accurately complete the task and form individual piles for each identity. These instructions were either ignored or misunderstood by some of the participants, who were then excluded from the analysis ( $n = 26$ ). This loss of data can be minimized in an in-person card sorting task as participants are potentially more attentive in lab and can gain clarification from the researcher regarding any aspect of the study.

In this study, we analysed performance on the card sorting task by collapsing across the two face pairs in each condition. Most face pairs within a condition (e.g., Black familiar) were not significantly different from one another in recognition performance; however, there was a significant difference in the number of piles (perceived identities) created for the two Black unfamiliar identity pairs,  $t(21) = 2.13$ ,  $p = .045$ . This may be due to a difference in similarity between identities within the face pairs. Previous studies suggest better discrimination of dissimilar identity pairs

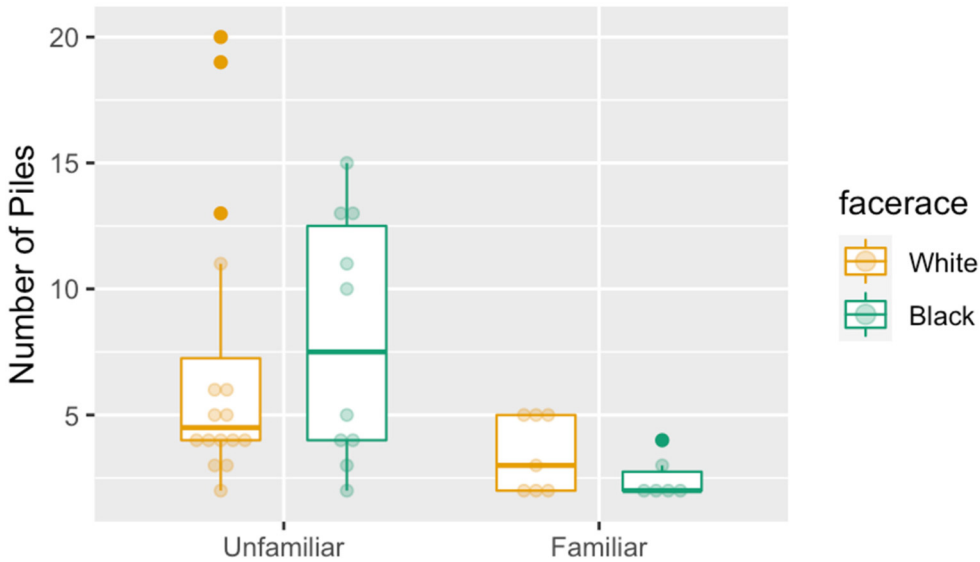
when compared to similar identity pairs (Davies et al., 1979; Ellis et al., 1992). This can lead to an increased sensitivity to between-person differences which may reduce the number of piles created in the card sorting task. Thus, it is possible that one pair of the two Black unfamiliar identity pairs were more dissimilar in appearance leading to significantly fewer piles. Moreover, Freema Agyeman, one of the Black unfamiliar celebrities, is on the show *New Amsterdam*, which at the time was in the top 10 list on Netflix Canada. Thus, it is possible that although participants reported no familiarity with Freema Agyeman, they may have had some exposure to her face and have been somewhat familiar with this identity. In the current study, we conducted an online validation study prior to beginning data collection, to assess familiarity ratings of target identities. In future studies, it would be important to conduct a validation study that assesses similarity ratings alongside familiarity ratings of each identity pair. This would ensure equivalent levels of similarity and familiarity for the target identity pairs within each condition, thereby allowing researchers to accurately assess the presence of the other-race effect during unfamiliar face recognition.

Future studies investigating the recognition of faces that vary in familiarity and race should consider the type of stimulus employed across studies when comparing the results. Much of the previous literature assessing facial recognition uses stimuli consisting of only internal features, excluding critical external information such as hair or face contour. The limited ecological validity of stimuli displaying only internal features has been criticized in previous research (e.g., Wong et al., 2020). Results based on face images without any cropping (e.g., with hair and facial contours included) and that incorporate natural variability (i.e., variations in lighting, setting and appearance) are arguably more representative of performance in real-world viewing conditions and may encourage reliance on external features for both unfamiliar faces and other-race faces (Megreya and Bindemann, 2009). In contrast, images without external features may be more detrimental to the recognition of unfamiliar face categories and thus may be more sensitive to detecting a potentially subtle other-race effect (Wong et al., 2020).

This study extends previous research by examining the joint influence of race and familiarity on the recognition of faces that vary in the presentation of internal and external features. Overall, we found better performance for familiar than unfamiliar faces, and for faces that contained internal features compared to only external features. These findings replicate past research: when faces are unfamiliar, pictures of the same person appear to belong to several distinct identities (Jenkins et al., 2011). This study also supports the attentional shift towards internal features as face familiarity increases (Kramer et al., 2018; Wong et al., 2020), although these findings also highlight the importance of external features during facial recognition. Examining face perception in adults has important practical implications as it can inform in-person identification and can provide suggestions to improve the accuracy of face recognition in contexts where face identification is critical (e.g., border security).

## Pilot Study

We first conducted a pilot study with 39 adult participants (28 women, 9 men, 2 preferred not to answer;  $M$  age = 21 years,  $SD$  = 3.7, range: 17–32 years) recruited from SONA (Toronto Metropolitan's Psychology research pool) and through word of mouth. Only White individuals (38 reported White ethnicity and 1 reported White and Indigenous ethnicity) were recruited for this study. This study was conducted to assess the efficacy of the online card sorting task and to compare its findings to patterns found in previous literature using the in-person card sorting task. An additional five participants completed the pilot study but were excluded from data analysis due to reporting that they were familiar with one or more of the identities in the unfamiliar condition or that they were unfamiliar with one or more of the identities in the familiar condition. The stimuli and procedure were the same as in the main study. Participants were randomly



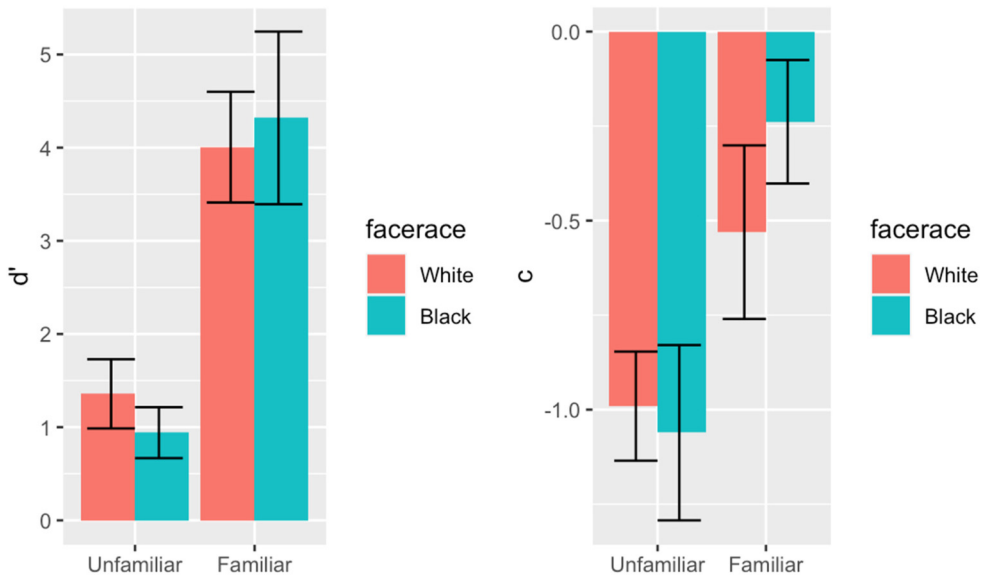
**Figure 6.** This boxplot presents the number of piles created in the whole face conditions for the pilot study. The whiskers (shown as the green and yellow lines extending from the box) represent the minimum and the maximum values, whilst the midline represents the median, and the box presents the interquartile range (first quartile, median and third quartile). The bolded points are outliers and the faded dots are the number of piles created by each participant in the card sorting task.

assigned to one of four card sorting conditions: own-race familiar, other-race familiar, own-race unfamiliar, and other-race unfamiliar. In the pilot study, all participants completed the whole face feature condition.

Data for the number of piles formed, sensitivity and response criterion were analysed separately using  $2 \times 2$  between-subjects analyses of variance, where both familiarity (familiar, unfamiliar) and face race (own-race, other-race) varied between participants. For the number of piles, we found a significant main effect of familiarity,  $F(1, 35) = 8.38, p = .007, \eta_p^2 = .19$ , with fewer piles created for familiar faces ( $M = 3.00, Mdn = 2.00, SD = 1.29$ ) compared to unfamiliar faces ( $M = 7.42, Mdn = 5.00, SD = 5.28$ ). We did not find a main effect of face race,  $F(1, 35) = .04, p = .84, \eta_p^2 = .007$  (Figure 6). As well, there was no familiarity  $\times$  face race interaction,  $F(1, 35) = .37, p = .55, \eta_p^2 = .01$ . However, there were more piles created in the unfamiliar other-race condition ( $M = 8.00, Mdn = 7.50, SD = 4.88$ ) when compared to the unfamiliar own-race condition ( $M = 7.06, Mdn = 4.50, SD = 5.64$ ), a pattern in line with our predictions.

For sensitivity ( $d'$ ), there was a main effect of familiarity,  $F(1, 35) = 32.90, p < .001, \eta_p^2 = .49$ : lower sensitivity was found for the unfamiliar conditions ( $M = 1.20, SD = 1.28$ ) compared with familiar conditions ( $M = 4.15, SD = 1.85$ ). Again, we did not find a main effect of face race,  $F(1, 35) = .11, p = .74, \eta_p^2 = .003$ , and there was no familiarity  $\times$  face race interaction,  $F(1, 35) = .49, p = .49, \eta_p^2 = .01$ . Similarly, for response criterion ( $c$ ) we found a main effect of familiarity,  $F(1, 35) = 9.14, p = .005, \eta_p^2 = .20$ , but did not find a main effect of face race,  $F(1, 35) = .08, p = .79, \eta_p^2 = .002$  (Figure 7) or a familiarity  $\times$  face race interaction,  $F(1, 35) = .76, p = .39, \eta_p^2 = .02$ . A more negative response criterion was found for the unfamiliar conditions ( $M = -1.02, SD = .63$ ) compared with familiar conditions ( $M = -.40, SD = .52$ ).

The goals of our pilot study were to (a) confirm that participants performed similarly in an online card sorting task compared to in-person card sorting tasks and (b) determine whether our stimulus



**Figure 7.** Mean sensitivity ( $d'$ ) and response criterion ( $c$ ) scores for the pilot study. Error bars represent standard error of the mean.

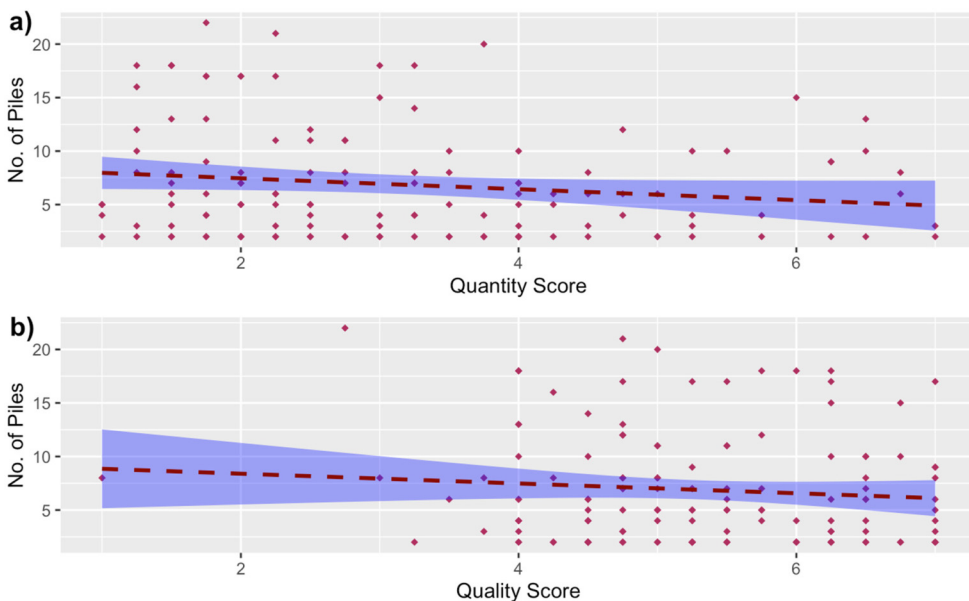
set elicited similar effects of familiarity and race as in previous literature. Participants tended to create more piles than there were identities present in the set, and the average numbers of piles created were similar to previous literature (i.e.,  $M=7.06$  piles created for unfamiliar own-race faces in the current study compared to  $M=7.28$  found for unfamiliar own-race faces in the Zhou and Mondloch, 2016, study). Additionally, participants created fewer piles for familiar faces than unfamiliar faces, and greater sensitivity was found for familiar than unfamiliar faces, which is consistent with previous literature (Jenkins et al., 2011; Kramer et al., 2018) and suggests a robust representation of familiar faces leading to greater recognition performance. As well, less negative response criterion was found for familiar than unfamiliar faces suggesting a bias towards reporting greater variation for unfamiliar faces (Balas & Pearson, 2017). Contrary to prior card sorting tasks, however, we did not find a statistically significant main effect of face race or an interaction between familiarity and face race, which is inconsistent with previous literature (Zhou & Mondloch, 2016). However, there was an indication of better recognition performance for unfamiliar own-race faces ( $Mdn=4.50$ ) when compared to unfamiliar other-race faces ( $Mdn=7.50$ ), suggesting that we might find an effect of face race with a larger sample. It is also possible that the typical other-race effect was diminished in the current sample. Previous studies with samples that have greater exposure to other-race faces (i.e., individuals from multiracial societies) have shown a weaker other-race effect in unfamiliar face recognition (De Heering et al., 2010; Wright et al., 2001). Wright et al. (2001) found that Black university students with greater exposure to White faces when compared to the general population, showed similar recognition accuracy for both White and Black faces. Similarly, the current pilot study only consisted of university students from the Greater Toronto Area, a highly multiracial metropolitan area, who may not show a large other-race effect due to greater exposure to other-race faces. Thus, we expanded recruitment for the main study beyond Toronto-area university students and instead included both Canadian and American residents between the ages of 18 and 47.



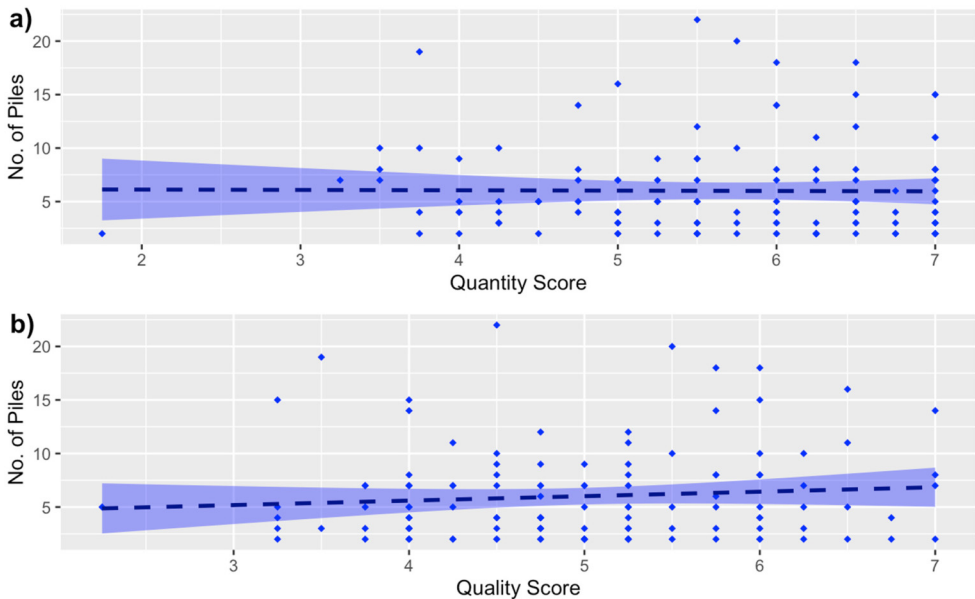
## Exploratory Analyses

**Intergroup Contact.** To assess the impact of contact between social groups on recognition performance, we conducted exploratory data analyses using participant responses on the Intergroup Relations CQCQ scale. The first question ‘how much contact do you have with X people **at university?**’ on the CQCQ scale was removed during analysis as this question did not apply to the entire sample (i.e., participants who were recruited from Prolific were not all university students). We computed two means for each participant: one for the quantity of contact with a group (i.e., Q2-6 in the CQCQ scale) and the other for the quality of contact with a group (i.e., Q6-10 in the CQCQ scale). These means were compared within groups of participants who saw White faces and participants who saw Black faces. On a Likert scale of 1 to 7, participants rated quantity of contact with White individuals significantly higher ( $M = 5.69, SD = 1.11$ ) than quantity of contact with Black individuals ( $M = 3.17, SD = 1.56$ ),  $t(237) = -14.98, p < .001$ . However, participants rated quality of contact with Black individuals significantly higher ( $M = 5.24, SD = 1.16$ ) than quality of contact with White individuals ( $M = 4.96, SD = .96$ ),  $t(251) = 2.10, p = .04$ . Next, we compared responses on the CQCQ scale to performance on the card sorting task using correlations and scatterplots. Both the Spearman rank-based correlation tests and the scatterplots did not show a relationship between the quantity ( $R = -.14, p = .12$ ) and the quality ( $R = -.05, p = .54$ ) of contact with Black individuals and the number of piles (i.e., perceived identity) created on the card sorting task by White participants who saw images of Black celebrities (Figure 8). As well, there was no correlation between the quantity ( $R = -0.04, p = .63$ ) and the quality ( $R = -0.07, p = .41$ ) of contact with White individuals and the number of piles created by White participants who saw images of White celebrities (Figure 9).

**Impact of Contact on Recognition.** Considering the absence of the other-race effect, it was important to investigate the contact our White participants had with the target other-race group (i.e., Black



**Figure 8.** These scatterplots present the mean scores of participants for the (a) quantity or (b) quality of contact with Black individuals and the number of piles created in the card sorting task when sorting images of Black celebrities.



**Figure 9.** These scatterplots present the mean scores of participants for the (a) quantity or (b) quality of contact with White individuals and the number of piles created in the card sorting task when sorting images of White celebrities.

individuals) compared to contact with White individuals. On the Intergroup Relations CQCQ scale, participants rated quantity of contact with White individuals significantly higher than quantity of contact with Black individuals, but rated quality of contact with Black individuals significantly higher than quality of contact with White individuals. We speculate that this discrepancy may be a result of the type of questions in the contact scales. Questions assessing quantity of contact are direct and ask participants about their day-to-day interactions with other-race individuals (i.e., ‘How much contact do you have with Black people as **close friends**?’), whereas questions assessing quality of contact are less direct and can heighten a participants’ awareness of their attitude and biases towards the target racial group (i.e., ‘To what extent did you experience contact with Black people as **pleasant**?’). Thus, White participants may have responded more positively to questions about quality of contact due to actively attending to and working against their racial biases.

Contrary to the contact hypothesis (Chiroro & Valentine, 1995), performance on the card sorting task (i.e., number of piles and sensitivity) did not positively correlate with the quality or quantity of contact with own- or other-race individuals. A similar result was found in Wong et al. (2020) where both self-reported quantity and quality of contact with other-race individuals on a social contact questionnaire, did not consistently predict how well other-race faces were recognized in a yes-no paradigm. In Meissner and Brigham’s (2001) meta-analysis of the other-race effect, self-reported contact with other-race individuals only accounted for 2% of the variance in the magnitude of the other-race effect. One possibility for the lack of relation between self-reported exposure and performance on other-race faces is that the CQCQ scale assesses current contact. There is substantial evidence that early exposure to other-race faces, during a sensitive period for the development of face perception, is instrumental in modifying the magnitude of the other-race effect (De Heering et al., 2010; McKone et al., 2019; Sangrigoli et al., 2005; Zhou et al., 2021; Zhou et al., 2019), whereas exposure in adulthood has little effect (McKone et al., 2019; Zhou et al., 2021; Zhou et al., 2019).


## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by the Natural Sciences and Engineering Research Council of Canada (grant number RGPIN-2019-05548).

## ORCID iD

Menahal Latif  <https://orcid.org/0000-0002-6010-662X>

## References

- Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, 68(10), 2041–2050. <https://doi.org/10.1080/17470218.2014.1003949>
- Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, 161, 19–30. <https://doi.org/10.1016/j.cognition.2016.12.012>
- Balas, B., & Pearson, H. (2017). Intra- and extra-personal variability in person recognition. *Visual Cognition*, 25(4–6), 456–469. <https://doi.org/10.1080/13506285.2016.1274809>
- Bonner, L., & Burton, A. M. (2004). 7–11-year-old children show an advantage for matching and recognizing the internal features of familiar faces: Evidence against a developmental shift. *The Quarterly Journal of Experimental Psychology Section A*, 57(6), 1019–1029. <https://doi.org/10.1080/02724980343000657>
- Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, 10(5), 527–536. <https://doi.org/10.1080/13506280244000168>
- Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology*, 59(2), 196–210. <https://doi.org/10.1006/jecp.1995.1009>
- 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>
- Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979). Similarity effects in face recognition. *The American Journal of Psychology*, 92(3), 507–523. <https://doi.org/10.2307/1421569>
- De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science*, 13(1), 181–187. <https://doi.org/10.1111/j.1467-7687.2009.00876.x>
- Di Oleggio Castello, M., Wheeler, K. G., Cipolli, C., & Gobbini, M. I. (2017). Familiarity facilitates feature-based face processing. *PLoS One*, 12(6), e0178895. <https://doi.org/10.1371/journal.pone.0178895>
- Ellis, H. D., Bruce, V., & De Schonen, S. (1992). The development of face processing skills [and discussion]. *Philosophical Transactions: Biological Sciences*, 335(1273), 105–111. <https://doi.org/10.1098/rstb.1992.0013>
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8(4), 431–439. <https://doi.org/10.1068/p080431>
- Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The functional neuroanatomy of human face perception. *Annual Review of Vision Science*, 3(1), 167–196. <https://doi.org/10.1146/annurev-vision-102016-061214>
- 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. <https://doi.org/10.1348/000712607X199981>
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330–337. [https://doi.org/10.1016/S1364-6613\(00\)01519-9](https://doi.org/10.1016/S1364-6613(00)01519-9)

- Havard, C. (2021). The importance of internal and external features in matching own and other race faces. *Perception, 50*(10), 861–875. <https://doi.org/10.1177/03010066211043464>
- Islam, M. R., & Hewstone, M. (1993). Dimensions of contact as predictors of intergroup anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality and Social Psychology Bulletin, 19*(6), 700–710. <https://doi.org/10.1177/0146167293196005>
- Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323. <https://doi.org/10.1016/j.cognition.2011.08.001>
- Johnston, R. J., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory (Hove, England), 21*(5):577–596. <https://doi.org/10.1080/09658210902976969>
- Kinney, N. E. (2001). A guide to design and testing in online psychology courses. *Psychology Learning & Teaching, 1*(1), 16–20. <https://doi.org/10.2304/plat.2001.1.1.16>
- Kramer, R. S. S., Manesi, Z., Towler, A., Reynolds, M. G., & Burton, A. M. (2018). Familiarity and within-person facial variability: The importance of the internal and external features. *Perception, 47*(1), 3–15. <https://doi.org/10.1177/0301006617725242>
- Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology, 107*(2), 374–388. <https://doi.org/10.1111/bjop.12147>
- Liu, S., Anzures, G., Ge, L., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Development of recognition of face parts from unfamiliar faces. *Infant and Child Development, 22*(2), 165–179. <https://doi.org/10.1002/icd.1781>
- Lochner, K. (2016). Online assessment and online surveys. In K. Lochner (Ed.), *Successful emotions: How emotions drive cognitive performance* (pp. 7–20). Springer Fachmedien. [https://doi.org/10.1007/978-3-658-12231-7\\_1](https://doi.org/10.1007/978-3-658-12231-7_1)
- McKone, E., Dawel, A., Robbins, R. A., Shou, Y., Chen, N., & Crookes, K. (2021). Why the other-race effect matters: Poor recognition of other-race faces impacts everyday social interactions. *British Journal of Psychology, 112*(1). <https://doi.org/10.1111/bjop.12508>
- McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., & Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports, 9*(1), 12820. <https://doi.org/10.1038/s41598-019-49202-0>
- Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception, 38*(12), 1831–1848. <https://doi.org/10.1068/p6385>
- 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>
- Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition, 21*(2), 139–253. <https://doi.org/10.1080/13506285.2013.772929>
- Rossion, B., & Michel, C. (2012). An experience-based holistic account of the other-race face effect. In *Oxford Handbook of Face Perception* (pp. 215–244). <https://doi.org/10.1093/oxfordhb/9780199559053.013.0012>
- Sandford, A., & Ritchie, K. L. (2021). Unfamiliar face matching, within-person variability, and multiple-image arrays. *Visual Cognition, 29*(3), 143–157. <https://doi.org/10.1080/13506285.2021.1883170>
- Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*(6), 440–444. <https://doi.org/10.1111/j.0956-7976.2005.01554.x>
- Statistics Canada (2016). Ethnic origin and visible minorities, 2016 Census, Retrieved from <http://www.statcan.gc.ca/bso/c/olc-cel/olc-cel?lang=eng&catno=97-562-X>.
- Suhrke, J., Freitag, C., Lamm, B., Teiser, J., Poloczek, S., Fassbender, I., Teubert, M., Voehringer, I., Keller, H., Knopf, M., Lohaus, A., & Schwarzer, G. (2015). Experience with headwear influences the other-race effect in 4-year-old children. *Journal of Experimental Child Psychology, 138*, 2015.03.011. <https://doi.org/10.1016/j.jecp.2015.03.011>
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583–592. <https://doi.org/10.3758/BF03211301>
- Tüttenberg, S. C., & Wiese, H. (2019). Learning own- and other-race facial identities from natural variability. *Quarterly Journal of Experimental Psychology, 72*(12), 2788–2800. <https://doi.org/10.1177/1747021819859840>

- Want, S. C., Pascalis, O., Coleman, M., & Blades, M. (2003). Recognizing people from the inner or outer parts of their faces: Developmental data concerning ‘unfamiliar’ faces. *British Journal of Developmental Psychology*, 21(1), 125–135. <https://doi.org/10.1348/026151003321164663>
- Westfall, J. (2015). PANGEA: Power ANalysis for GEneral Anova designs. PhD Proposal.
- 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>
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law* 7(1):119–133. <https://doi.org/10.1037/1076-8971.7.1.119>
- Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science*, 26(3), 212–217. <https://doi.org/10.1177/0963721416688114>
- Zhou, X., Burton, A. M., & Jenkins, R. (2021). Two factors in face recognition: Whether you know the person’s face and whether you share the person’s race. *Perception*, 50(6), 524–539. <https://doi.org/10.1177/03010066211014016>
- Zhou, X., Elshiekh, A., & Moulson, M. C. (2019). Lifetime perceptual experience shapes face memory for own- and other-race faces. *Visual Cognition*, 27(9–10), 687–700. <https://doi.org/10.1080/13506285.2019.1638478>
- Zhou, X., & Mondloch, C. J. (2016). Recognizing “bella swan” and “hermione granger”: No own-race advantage in recognizing photos of famous faces. *Perception*, 45(12), 1426–1429. <https://doi.org/10.1177/0301006616662046>