Exploring the contribution of motivation and experience in the post-pubescent own-gender bias in face recognition

Peter J Hills[1],

Bournemouth University

J Michael Pake[2], Jack R Dempsey[2],

Anglia Ruskin University

& Michael B Lewis[3]

Cardiff University

1 - Correspondence Author
Department of Psychology
Bournemouth University
Talbot Campus
Fern Barrow
Poole, Dorset
BH12 5BB
phills@bournemouth.ac.uk

2 - Department of Psychology
Anglia Ruskin University
East Road
Cambridge
CB1 1PT

3 - School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff
CF10 3AT

Word Count: 12057

10.1037/xhp0000533

Abstract

The own-gender bias in face recognition has been hypothesised to be the result of extensive experience with own-gender faces, coupled with a motivation to process own-group faces more deeply than other-group faces. We test the effect of experience and motivation in four experiments employing standard old/new recognition paradigms. In Experiment 1, no own-gender recognition bias was observed following an attractiveness-rating encoding task regardless of school type (single- or mixed-sex). Experiment 2, which used a distinctiveness-rating encoding task, did find a significant own-gender bias for all groups of participants. Experiment 3 on adults found that the own-gender bias was not affected by self-reported contact with the other-gender, but the encoding task did moderate the size of the bias. Experiment 4 revealed that participants with an own-gender sexual orientation showed a stronger own-gender bias. These results indicate that motivational factors influence the own-gender bias whereas no evidence was found for perceptual experience.

Public Significance Statement:

This study suggests that the biases that exist in face perception (the bias toward recognising faces of one's own gender) may be due to how interested we are in processing those types of faces. To reduce biases toward recognising faces of one's own group, we must have sufficient motivation to process other-group faces deeply.

Keywords

own-gender bias; own-group biases; face recognition; experience; motivation

PsychINFO Classification Codes 2323, 2340

Exploring the contribution of motivation and experience in the post-pubescent own-gender bias in face recognition

One of the most reliable findings in the face recognition literature is that of the own-group bias[1]. This is where participants demonstrate superior recognition for own-group (be that, ethnicity, gender, or age), relative to other-group, faces (e.g., Blandón-Gitlin, Pezdek, Saldivar, & Steelman, 2014; Brigham & Malpass, 1985; Wright & Sladden, 2003; Bäckman, 1991; Bartlett & Leslie, 1986; Rule, Ambady, Adams, & Macrae, 2007). The own-ethnicity bias[2] is the most extensively researched and theorised of these biases. The focus of the current research, however, is the own-gender bias. This bias is particularly interesting because, although male and female faces differ, they have similar physiognomic variablity within their categories – unlike faces of different ethnicities. Moreover, gender, like ethnicity, is a particularly salient classifier for one's own identity and the identity of others.

Here, it is explored how the theoretical models devised to explain the own-ethnicity bias can be used to explain the the own-gender bias. Researchers have indicated perceptual experience (e.g., Hayward, Crookes, & Rhodes, 2013, or Valentine, 1991) and socio-cognitive motivational mechanisms (e.g., Sporer, 2001) explain the own-ethnicity bias. Theoretical models of the own-group biases indicate that we employ some form of expert or deeper level of encoding for own-group faces relative to other-group faces: This expert or deeper level of encoding used for own-group faces is due to experience and/or motivation. However, direct tests comparing the relative involvement of both in a single study have not been forthcoming. Understanding the cause of the own-group biases in face recognition are important in constructing accurate models of face recognition generally and to develop methods for reducing them in settings where accurate

[1] The term 'own-group bias' is preferred to describe the effect (rather the often used 'cross-group effect', 'other-group effect' or 'in-group bias') because the biases do not always 'cross' groups and that this is a bias due to one's perception of one's own group rather than the effect of another group.

[2] Some authors refer to the own-ethnicity bias as the own-race bias, however, since there is only one human sub species (*homo sapiens*), it is a misnomer to refer to this as a racial bias. The term 'race' may be acceptable to refer to major anthropological groups, it is incorrect (as is common in the literature) when referring to ethnic groups (such as 'Hispanic' who are Caucasian and therefore the same race as 'Whites' Valentine, Lewis, & Hills, 2016).

recognition of all faces is important. Typically, experience-based explanations suggest it is the number or quality of encounters with faces of different groups that predict the observed bias (Valentine, 1991) whereas motivation-based explanations suggest that there are cultural reasons why own-group faces are processed more deeply (Hugenberg, Young, Bernstein, & Sacco, 2010). Below, it is described how these explanations account for existing findings in the perception of faces before a series of experiments are presented that aimed to distinguish between them in the context of own-gender bias.

Experience-based exlanations of the own-group bias suggest that experience leads to the development of expertise for processing own-group faces. One model used to explain expertise in face recognition is Valentine's (1991) face-space model in which every face is stored in a multidimensional space defined by facial dimensions. This model has been influential in many aspects of face processing (see Valentine, Lewis, & Hills, 2016). It is a model that clearly describes how experience affects face perception. Each dimension represents the physiognomy of the most frequently encountered faces (Lewis, 2004; McKone, Aitkin, & Edwards, 2005). Dimensions of face-space can be added to help differentiate between faces more frequently encountered: If two faces cannot be discriminated between easily, perceptual learning theory (e.g., McLaren, 1997; Mundy, Honey, & Dwyer, 2007) suggests that people will focus on what differentiates the two and inhibit the common elements (Hall, 1991). Perceptual learning is more effective when learning to discriminate between similar stimili (e.g., Mundy, Dwyer, & Honey, 2006) suggesting dimensions will be added to distinguish between the most frequently encountered faces (Valentine & Endo, 1992). Dimensions of face-space can also be removed if they no longer discriminate between faces that are frequently encountered (Hills, Holland, & Lewis, 2011) as in discrimination learning (Sheffield, 1965; Saarinen & Levi, 1995). Evidence for this stems from evidence indicating younger children being able to discriminate between pairs of monkey faces that adults are less able to discriminate between (Lewkowicz & Ghazanfar, 2006; Pascalis, de Haan, & Nelson, 2002; Scott, Shannon, & Nelson, 2005, 2006) and be adapted to facial distortions that adults cannot be adapted to (Hills, Holland, & Lewis, 2010). With increased expertise, it is easier to rule out and ignore less relevant information, creating a characteristically efficient expert processing style (Charness, 1981; Saariluoma, 1994; Simon & Barenfeld, 1969; Simon & Gilmartin, 1973; Wagner & Scurrah, 1971). The result of extensive experience causes the face-space to have less frequently encountered groups of faces stored in a densely-packed cluster (because the dimensions are not appropriate to discriminate between them) far from the centre of the face-space (Johnston & Ellis, 1995). This model therefore accounts for own-group biases by extensive experience with own-group faces refining the space.

The expert nature of face processing is thought to be based on holistic processing (Diamond & Carey, 1977) and some authors have indicated that this type of processing is deployed more readily for own-group faces relative to other-group faces (Michel, Corneille, & Rossion, 2007). According to Maurer, Le Grand, and Mondloch (2001) holistic processing is a form of configural encoding in which the entire face is encoded as a gestalt (e.g., Rossion, 2008, 2009). Some authors describe holistic processing as a failure of selective attention to attend to parts of images (Richler, Gauthier, Wenger, & Palmeri, 2008). Inversion disrupts and delays the use of holistic processing (Richler, Mack, Palmeri, & Gauthier, 2011). Holistic processing is based on our visual experience (Carey, de Schonen, & Ellis, 1992; Le Grand, Mondloch, Maurer, & Brent, 2001) and is typically unconscious, hard to verbalise, and involves chunking. In this way, it is similar to many forms of expertise and tacit knowledge (e.g., Newell & Simon, 1972). Configural processing also entails processing the first- and second-order relationships between facial features but these are less likely to be involved in expert facial discrimination (Burton, Schweinberger, Jenkins, & Kaufmann, 2015). The engagement of holistic processing for own-group faces is thought to be automatic.

The reduced reliance on holistic processing for other-ethnicity faces is evidenced in the face-composite task: a task in which comparisons of the top half of a face are influenced by differences in the bottom half of the face (see Young, Hellawell & Hay, 1987). The interference of bottom halves on identfiying that the top halves are the same is a result of the holistic processing of faces (Rossion, 2008). This face-composite effect is smaller for other-ethnicity faces than own-ethnicity faces (Michel, et al., 2007), which leads to the suggestion that the processing of other-ethnicity faces is less holistic than own-ethnicity faces. Why different processing styles are used for own- and other-ethnicity faces is not clear but might have something to do with experience, as suggested by Michel, Corneille, and Rossion (2007), or motivation, as suggested by Hugenberg et al (2010).

Motivation to process own-group faces more deeply is also a critical factor in the magnitude of the own-group biases according to the categorisation-individualisation model proposed by Hugenberg and colleagues (2010). They suggest that motivation and experience may affect the deployment of expert face processing. The motivation-based component of own-group bias is demonstrated by evidence suggesting that simply classifying faces as one's own group (either experimentally-induced; or due to religious or university affiliation) leads them to be recognised more accurately than classifying the same faces as out-group (Bernstein, Young, & Hugenberg, 2007; Hehman, Mania, & Gaertner, 2010; Rule, Garrett, & Ambady, 2010; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008; Young, Bernstein, & Hugenberg, 2010). Indeed, labelling androgynous faces as either 'male', 'female', or 'faces' led to Rehnman and Herlitz's (2007) female participants showing an own-gender bias for the faces labelled as own-group.

Consistent with this motivation-based account for the biases, motivation has been shown to reduce the own-ethnicity bias: if participants are expecting to interact with a face later on, they are more motivated to process it deeply (Wilson, See, Bernstein, Hugenberg, & Chartier, 2014). Similarly, the own-ethnicity bias is smaller for faces that are perceived to be of a higher status, presumably due to motivation to process them more accurately (Ratcliffe, Hugenberg, Shriver, & Bernstein, 2011; Shriver et al., 2008).

The motivation-based explanation can also explain why the task used during learning the faces affects the size of the bias. The own-gender bias is larger for male faces when they are rated for male-traits (such as dominance) in male participants than when they are rated for more neutral traits (such as likeability; Motta-Mena, Picci, & Scherf, 2016). These tasks may direct the participants' attention towards more socially relevant engagement with the to-be-remembered faces. Motivation can account for why the own-age bias is also enlarged in pubescent children relative to younger children (Picci & Scherf, 2016): older children may not desire to interact with younger children to the same extent that younger children may wish to interact with older children. This highlights developmental changes in the function of face processing (Scherf, Behrmann, & Dahl, 2012) that may be beyond simple experience or age-of-acquisition effects (Scherf & Scott, 2012). Collectively, these results indicate that motivation to engage in deeper processing is an issue in creating the own-group biases and also potentially important for reducing them.

Contrasting with motivation-based accounts, experience-based explanations have been effective in explaining factors affecting the own-ethnicity bias. Cross, Cross, and Daly (1971) demonstrated that the own-ethnicity bias was larger in white children that lived in ethnically-segregated neighbourhoods compared to children that lived in desegregated neighbourhoods. Chiroro and Valentine (1995) found that the own-ethnicity bias was larger in White British and Black Zimbabwean participants from villages with less other-ethnicity contact than participants from cities with higher other-ethnicity contact. Sporer and Horry (2011) reported that Turks living in Germany recognised German and Turkish faces equally well, but German participants showed a recognition advantage for German faces relative to Turkish faces. Walker and Hewstone (2006) found an own-ethnicity bias among white U.K. residents in their recognition of South Asian faces, but not for Asian people living in the U.K. Self-reported contact with people from other-ethnicities reduced the own-ethnicity bias (Brigham, Maass, Snyder, & Spaulding, 1982, but see Brigham & Barkowitz, 1978, Luce, 1974, Malpass & Kravitz, 1969), though the evidence appears to be that it is the quality of the contact rather than the sheer quantity of experience that matters most (Brigham & Malpass, 1985; Forsyth, Cinque, & Bukach, 2017).

Experientially, the own-gender bias is somewhat different from the own-ethnicity bias in that for most people, they will encounter an equal number of male and female faces during their lifetime. Nevertheless, studies into the developmental trajectory of the own-gender bias show that girls demonstrate the own-gender bias from the age of 6-years, but boys recognise women's faces more accurately than male faces (Feinman & Entwisle, 1976; Ge et al., 2008). In fact, the own-gender bias is reliably found in women but less so in men (Lewin & Herlitz, 2002; Lovén, Herlitz, & Rehnman, 2011; Rehnman & Herlitz, 2006, 2007). Herlitz and Lovén (2013) suggest that experience can be used to explain the findings. There are two elements to the explanation. First, females have a general advantage in face processing skills possibly as a result of greater mutual eye contact with females (Ashear & Snortum, 1971; Exline, 1963; Exline, Gray, & Schuette, 1965; Field, Cohen, Garcia, & Greenberg, 1984; Levine & Sutton-Smith, 1973; Osofsky & O'Connell, 1977; Rennels & Davis, 2008). Second, there is greater early experience with female faces for people of both genders (Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2000; Hittelman & Dickes, 1979; Leeb & Rejskind, 2004). For boys, this early experience is at least partially counteracted by later significant interactions with own-group peer groups.

There is also, however, a motivation-based explanation for Herlitz and Lovén's (2013) results. It is possible that the females are more motivated to process same-gender faces than other-gender faces for cultural reasons. The male participants may not possess a similar differential motivation. This pattern of motivations would explain Herlitz and Lovens' findings and so they, on their own, cannot distinguish between an experience- and a motivation-based account of the own-gender bias. Nevertheless, the own-gender bias can be found in both male and female participants (Man & Hills, 2016) depending on the task instructions during learning (Motta-Mena et al., 2016).

This background indicates that the own-gender bias in expert face processing is potentially based on either motivation or experience (as suggested by Hugenberg et al., 2010). While these factors have been considered in isolation, it is possible that these factors interact. For example, motivation may lead one to seek out experiences with particular groups or experience of groups may affect the way that social categorisations are made. Evidence for such an interaction, and to establish the relative importance of experience and motivation is crucial to advance theoretical models of the own-group biases.

The current study aimed to explore the roles of these factors in the own-gender bias in four experiments. Experiment 1 explored the role of objectively-measured experience on the own-gender bias as the experience-based account would predict that those with a greater degree of experience with the other-gender will show a smaller own-gender bias. Experiment 2 explored whether the failure to find an own-gender bias in Experiment 1 could be explained in terms of a motivation-based account. It did this by looking at the effect of changing the encoding instructions from attractiveness ratings (Experiment 1), which may benefit other-gender faces, to distinctiveness ratings (Experiment 2), which are more neutral in context. Experiment 3 directly contrasted the experience-, expertise-, and motivation-based accounts of the own-gender bias by using both the attractiveness ratings and the distinctiveness ratings at the encoding phase. Moreover, experience of the other gender was accessed using a self-report measure. Finally, Experiment 4 tested the motivation-based explanation by contrasting the size of the own-gender bias using attractiveness ratings at encoding in groups of people who self-reported as either being gay or straight. In these experiments, we primarily measured recognition accuracy, but also explored response times, since increased motivation to process faces might be reflected in longer processing times (Crookes & Rhodes, 2017), though, it might also reflect more difficult processing.

Experiment 1

The purpose of Experiment 1 was to explore the role of experience on the own-gender bias. Experience has been suggested to be important in the development of the own-gender bias (Herlitz & Lovén, 2013); however, a test of real-world quantifiable experience has yet to be conducted. To do this, a group of participants who have significantly more perceptual experience with own-gender faces was tested. It was hypothesised that, if experience were a critical factor then students (aged 17 years) who have been educated in single-sex schools would show a larger own-gender bias than those educated in mixed-sex schools when tested in a standard old/new recognition experiment using attractiveness ratings at the encoding stage. Participants at the age of 17 were tested because they will have had the maximum impact of a single- or mixed-sex schooling system.

To assess whether the own-gender bias differed in magnitude to other own-group biases, we compared the own-gender bias in this group to the own-age and own-ethnicity bias. We anticipated that all our participants would have limited experience with other-age and other-ethnicity faces and so would therefore show the standard own-age and own- ethnicity biases. It was hoped that these biases could be used to contrast the relative size of any own-gender biases observed.

Method

Participants

Twenty-six boys and 27 girls took part in this study from four single-sex schools in addition to 23 boys and 22 girls from three mixed-sex schools in South Wales. Sample size was determined through a power analysis to observe the interaction between participant gender and face gender (thereby revealing the own-gender bias). With a η_p^2 of .08 (from Herlitz and Lovén's, 2013, meta-analysis), to achieve a power of .90, 97 participants were required (calculated using PANGEA, https://jakewestfall.shinyapps.io/pangea/, and Hulley, Cummings, Browner, Grady, & Newman, 2013). Many single-sex schools have higher socioeconomic status and student outcome than mixed-sex schools. To explore the demographic details and student outcomes across these schools, we compared school rating data for the single- and mixed-sex schools based on data from school comparison websites (https://www.compare-school-performance.service.gov.uk/ and https://www.estyn.gov.wales/). These comparisons are shown in Table 1. There were no statistically significant differences across these (though this is due to the small number of the schools). All participants self-reported that they had normal or corrected vision and were considered typically developing by their schools. All were 17 years old at the time of testing and self-reported that they were ethnically White. All participants reported spending at least 5 years in single- or mixed-sex school (depending on group).

Materials

One-hundred and twenty faces from the Minear and Park (2004) database were used for this Experiment. Of these, 40 (20 female) were ethnically-Black (aged 20- to 22-years old), 40 (20 female) were ethnically-White (aged 20- to 22-years old), and 40 (20 female) were ethnically-White (aged 65- to 85-years old). Participants only saw other-age and other-ethnicity faces of their own gender, therefore each participant saw 80 faces during the Experiment. All faces had similar hairstyles and did not have jewelry, distinctive marks, nor paraphernalia (e.g., glasses, beards, or ear-rings). Two images of each face were used (displaying slightly different expressions), one presented during the learning phase and one presented during the test phase to ensure face recognition rather than picture recognition. The images were counterbalanced across participants. The faces were presented 100 mm by 110 mm dimensions in 72 dpi resolution and were presented using Superlab Pro 2TM Research Software using a Toshiba Tecra M4™ Tablet PC.

Design

A 2 (type of school: single-sex or mixed-sex) by 2 (gender of participants: female or male) by 4 (type of face: young White male, young White female, young Black, or old White) Mixed design was employed, whereby gender of participants and schooling was varied between-subjects. The faces were counterbalanced such that each face was a target as often as it was a distracter and such that each participant saw the same number of male, female, Black, and old faces. Participants saw other-age and other- ethnicity faces of their own-gender. This was done to ensure that any additive effects of combining biases was prevented. Recognition accuracy was measured using the signal detection theory (e.g., Swets, 1966) measure of stimulus discriminability, *d′*. This study received ethical approval from Cardiff University's Research Ethics Committee.

Procedure

A standard old/new recognition paradigm was employed. Participants were tested individually. Participants responded verbally and the experimenter entered the responses into a standard computer keyboard. Participants sat 50 cm from the computer screen. The Experimenter was blind to the contents of the screen, since it was turned away from him. Thus, the Experimenter could not influence the participants' responses.

The experimental procedure involved three consecutive phases: learning, distraction, and test. The learning phase involved showing the participants half of the set of faces (*N* = 40). Participants were instructed to rate each face for how attractive they thought the face was using a 1-9 Likert-type scale, where 1 was ugly and 9 was beautiful (Light, Hollander, & Kayra-Stuart, 1981). If a participant did not understand the scale, it was explained to them using alternative synonyms. The purpose of this scale was to ensure that the participants attended to the faces. The presentation of each face was response terminated. Between each face was a random-noise mask lasting 500 ms.

Immediately after this presentation, participants were given some control demographic questions asking them how long they have been in single-sex education for. These questions took no longer than 60 seconds to administer.

Following this, the participants were given the test phase. In this, the participants saw all 40 target faces and 40 distractor faces sequentially in a random order. For each face, participants were requested to make an old/new recognition judgement verbally. The experimenter keyed in the participants' response. Each presentation was response terminated. Between each face, participants were presented with a random-noise mask for 500ms. Once this phase was completed, participants were thanked and debriefed.

Results

We first analysed the recognition data before exploring the learning rating data. The old/new recognition responses were converted into the signal detection theory measure of stimulus discriminability, *d'*, using the MacMillan and Creelman (2005) method. Mean recognition accuracy is presented in Figure 1. These data were subjected to a 2 x 2 x 4 mixed-subjects ANOVA with the factors: type of school (between-subjects; mixed or single-sex), gender (between-subjects; male or female) and type of face (within-subjects; same gender, other gender, other age or other ethnicity). The only significant effect in this analysis was the main effect of type of face, $F(2.64, 248.55$[3]$) = 42.55$, $MSE = 0.65$, $p < .001$, $\eta_p^2 = .31$. Bonferroni-Šídák corrected pairwise comparisons revealed that own-age male and female own-ethnicity faces were recognised more accurately than other-age and other ethnicity faces (all *p*s < .001) – results consistent with there being own-age and own-race biases. No other main effects, nor interaction was significant: the main effect of participant gender, $F(1, 94) = 0.03$, $MSE = 0.66$, $p = .862$, $\eta_p^2 < .01$; the main effect of type of school, $F(1, 94) = 0.04$, $MSE = 0.66$, $p = .811$, $\eta_p^2 < .01$; the interaction between participant gender and type of school, $F(1, 94) = 0.52$, $MSE = 0.66$, $p = .475$, $\eta_p^2 < .01$; the interaction between type of face and participant gender, $F(2.64, 248.55) = 1.01$, $MSE = 0.65$, $p = .383$, $\eta_p^2 = .01$; the interaction between type of face and type of school, $F(2.64, 248.55) = 0.10$, $MSE = 0.65$, $p = .939$, $\eta_p^2 < .01$; the three-way interaction, $F(2.64, 248.55) = 0.30$, $MSE = 0.65$, $p = .802$, $\eta_p^2 < .01$.

Subsequently, we assessed the magnitude of each of the biases (the own-gender, own-age, and own-ethnicity bias) to establish if any were present in our participants. Given that participants only viewed other-age and other-ethnicity faces of their own gender, this metric did not combine multiple biases in one score. The own-group biases were calculated by subtracting the other-group recognition accuracy score from the own-group recognition accuracy score. Therefore, the own-group faces were used in the calculation of each of the other biases. This calculation gave three

[3] Mauchley's test of sphericity was significant, $W(5) = .81$, $p = .002$, therefore the Greenhouse-Geisser correction was applied to the degrees of freedom.

scores representing the three own-group biases: Positive scores indicate better recognition for own-group faces than other-group faces and negative numbers the reverse. The size of these biases are demonstrated in Figure 2 separated for each school type.

The results summarised in Figure 2 were subjected to a 2 x 3 ANOVA: type of school (between subject) by bias (within subject). This revealed a main effect of bias, $F(2, 192) = 45.22$, $MSE = 0.75$, $p < .001$, $\eta_p^2 = .32$. Bonferroni-Šídák corrected pairwise comparisons were run to compare the magnitude of the biases. The own-age bias ($M = 1.00$, $SE = 0.12$) was significantly larger than the own-gender bias ($M = 0.17$, $SE = 0.09$), $t(97) = 6.79$, $p < .001$, Cohen's $d = 0.79$. The own-ethnicity bias ($M = 1.31$, $SE = .09$) was significantly larger than the own-gender bias, $t(97) = 10.16$, $p < .001$, Cohen's $d = 1.28$, and approaching significantly greater than the own-age bias, $t(97) = 2.33$, $p = .060$, Cohen's $d = 0.30$. One-sample *t*-tests confirmed that the own-gender bias was not significantly different from zero, $t(97) = 1.37$, $p = .17$. Both the own-age and the own-ethnicity biases were significantly different from zero, $t(97) = 8.67$, $p < .001$ and $t(97) = 13.28$, $p < .001$, respectively. The main effect of participant gender, $F(1, 94) = 2.34$, $MSE = 1.52$, $p = .130$, $\eta_p^2 = .02$, type of school $F(1, 94) = 0.01$, $MSE = 1.52$, $p = .944$, $\eta_p^2 < .01$, nor the interactions between participant gender and type of school, $F(1, 94) = 0.13$, $MSE = 1.52$, $p = .722$, $\eta_p^2 < .01$, between bias type and type of school, $F(2, 188) = 0.14$, $MSE = 0.71$, $p = .870$, $\eta_p^2 < .01$, and the three-way interaction, $F(2, 188) = 0.18$, $MSE = 0.71$, $p = .834$, $\eta_p^2 < .01$, were not significant.

Figure 3 shows the learning data for attractiveness ratings during the learning task. The attractiveness data were subjected to a parallel 2 x 2 x 4 ANOVA. This analysis revealed a main effect of type of face, $F(2.56, 240.91$[4]$) = 84.79$, $MSE = 0.54$, $p < .001$, $\eta_p^2 = .47$. Bonferroni-Šídák corrected pairwise comparisons revealed that female faces ($M = 3.66$, $SE = .13$), were rated as more attractive than other-age faces ($M = 3.27$, $SE = .13$), other-ethnicity faces ($M = 2.92$, $SE = .12$), and male faces ($M = 2.17$, $SE = .11$): all comparisons were significant (all *p*s < .001). This main effect

[4] Mauchley's test of sphericity was significant, $W(5) = .77$, $p < .001$, therefore the Greenhouse-Geisser correction was applied to the degrees of freedom.

interacted with participant gender, $F(2.56, 240.91) = 5.51$, $MSE = 0.54$, $p = .002$, $\eta_p^2 = .06$. Bonferroni-Šídák corrected *t*-tests revealed that male faces were rated as more attractive by girls than by boys, $t(69.76$[5]$) = 3.41$, $p = .001$, Cohen's $d = 0.69$. There were no gender differences in the attractiveness ratings for all other types of faces (all *p*s > .330).

No other main effects nor interactions were significant: participant gender, $F(1, 94) = 1.87$, $MSE = 4.33$, $p = .174$, $\eta_p^2 = .02$; type of school, $F(1, 94) = 0.05$, $MSE = 4.33$, $p = .820$, $\eta_p^2 < .01$; participant gender by type of face, $F(1, 94) = 0.02$, $MSE = 4.33$, $p = .876$, $\eta_p^2 < .01$; type of face by type of school, $F(2.56, 240.91) = 0.82$, $MSE = 0.54$, $p = .468$, $\eta_p^2 = .01$; nor the three-way interaction, $F(2.56, 240.91) = 0.13$, $MSE = 0.54$, $p = .846$, $\eta_p^2 < .01$.

Discussion

The results from Experiment 1 indicate that the own-age and own-ethnicity bias were significant, but the own-gender bias was not present in any of our participants. We anticipated that the own-gender bias would be larger in our participants who had greater exposure to own-gender faces, however, this was not borne out by our data. Indeed, we failed to show any own-gender bias. These results are inconsistent with many of the published studies reporting the own-gender bias (e.g., Lewin & Herlitz, 2002; Lovén, Herlitz, & Rehnman, 2011; Man & Hills, 2016; McKelvie, 1987; Rehnman & Herlitz, 2006, 2007). We have shown significant own-age and own-ethnicity biases in our participants although in a limited sense. The experiment only tested one age-group and own-ethnicity and so it does not show full bidirectional cross-over interactions here as has been demonstrated elsewhere (e.g., Hills & Pake, 2013). This, therefore, means we cannot rule out the fact that these biases occurred due to stimulus differences (for example, in distinctiveness).

One methodological difference between the present study and previous work (excluding the face database used, which will be ruled out in subsequent experiments) is that, in those, participants are

[5] Levene's test of equality of variances was significant, $F = 42.98$, $p < .001$, therefore the degrees of freedom were adjusted accordingly.

instructed to learn the faces, make distinctiveness judgements during learning, or simply told to view the faces. It is conceivable that making an attractiveness judgement may cause participants to encode other-gender faces more deeply as they might be more motivated to process them (at least in heterosexual participants – sexuality of the participants was not recorded). This is plausible given the importance of attractiveness in sexual and romantic interest (Buss, 1985; Rhodes, Simmons, & Peters, 2005). This might therefore indicate that the motivation to process other-gender faces over-rides that of perceptual experience. Alternatively, it may be that judging attractiveness of other-sex faces is more typical than judging attractiveness of same-sex faces, especially for these 17-year old participants. This typicality may be reflected in employing appropriate processing when making attractiveness judgements of other-sex faces whereas less appropriate processing would be employed when making attractiveness judgements of same-sex faces.

Our attractiveness rating data may support this perspective, given that the female participants did rate male faces as more attractive than male participants did. This potentially supports the notion that they were more interested in these stimuli. However, the converse was not true - male participants did not find female faces as more attractive than female participants. Anecdotally, this may be the result of male participants less willing to rate male faces as attractive and therefore creating a floor effect in this factor. The issue of whether making attractiveness judgements alters the way faces were processed was directly assessed in the subsequent Experiments.

Experiment 2

Experiment 1 failed to find an own-gender bias even in participants who had greater exposure to own-gender faces than other-gender faces. One reason for this failure might be because participants were requested to make attractiveness judgements when they first viewed the faces. To explore this, Experiment 2 was run in an identical manner to Experiment 1 except that participants were requested to make distinctiveness judgements, rather than attractiveness judgements, during the learning phase of the experiment. A face's distinctiveness is related to its attractiveness (Vokey &

Read, 1992) but it is more gender-neutral than attractiveness. As such, it is possible to use distinctiveness ratings as an encoding task that will not benefit the other-gender faces.

Method

Participants

Participants were students (aged 15 to 16 years) from two single-sex schools and from three mixed-sex schools in South Wales. Forty-nine were female and 46 were male and were recruited from a sample that returned consent forms to their schools. All were considered typically developing by their school and self-reported that they had normal or corrected vision and were ethnically White. The sample size was selected to be similar to Experiment 1.

Materials, Design, and Procedure

The materials used, design, and procedure were identical to that of Experiment 1, except that during the learning phase of the experiment, participants were requested to judge the faces for 'how easy they would be to spot in a crowd' on a 1 (very difficult, therefore typical face) to 9 (very easy, therefore distinctive face) Likert-type scale as used by Light, Kayra-Stuart, & Hollander (1979, Rhodes, Sumich, & Byatt, 1999). This study received ethical approval from Cardiff University's Research Ethics Committee.

Results

The mean recognition accuracy is presented in Figure 4 and was subjected to a 2 x 2 x 4 mixed-subjects ANOVA with the between-subjects factors of type of school (mixed- or single-sex) and participant gender (male or female) and the within-subjects factor of face type (same gender, other gender, other age, or other ethnicity). This analysis revealed a main effect of type of face, $F(3, 273) = 20.29$, $MSE = 0.67$, $p < .001$, $\eta_p^2 = .18$. Bonferroni-Šídák corrected pairwise comparisons revealed that own-age and -ethnicity male and female faces were recognised more accurately than other-age and other-ethnicity faces (all *p*s < .001). This main effect interacted with participant gender, $F(3, 273) = 8.93$, $MSE = 0.67$, $p < .001$, $\eta_p^2 = .09$. A series of Bonferroni-Šídák corrected between-subjects *t*-tests were run to compare the recognition of each type of face for boys and girls. These revealed that girls recognised female faces ($M = 2.62$, $SE = 0.08$) more accurately than male faces ($M = 1.89$, $SE = 0.14$), *t*(71.81[6]) = 4.47, $p < .001$, Cohen's $d = 0.93$, and boys recognised male faces ($M = 2.32$, $SE = 0.11$) more accurately than female faces ($M = 1.84$, $SE = 0.15$), $t(85.26) = 2.52$, $p = .013$, Cohen's $d = 0.52$. There was no difference in the recognition of other-age nor other-ethnicity faces, $t(81.57) = 0.18$, $p = .860$, Cohen's $d = 0.04$, and $t(93) = 0.28$, $p = .780$, Cohen's $d = 0.06$, respectively.

The main effect of type of school approached significance, $F(1, 91) = 3.69$, $MSE = 0.65$, $p = .058$, $\eta_p^2 = .04$, in which participants from single-sex schools recognised faces ($M = 1.92$, $SE = 0.06$) somewhat more accurately than those from mixed-sex schools ($M = 1.76$, $SE = 0.06$). The main effect of participant gender was not significant, $F(1, 91) = 0.44$, $MSE = 0.65$, $p = .508$, $\eta_p^2 = .01$, nor the interaction between participant gender and type of school, $F(1, 91) = 0.07$, $MSE = 0.65$, $p = .796$, $\eta_p^2 < .01$, nor the interaction between type of face and type of school, $F(3, 273) = 1.77$, $MSE = 0.67$, $p = .154$, $\eta_p^2 = .02$, nor was the three-way interaction, $F(3, 273) = 0.25$, $MSE = 0.67$, $p = .864$, $\eta_p^2 < .01$.

Similar to Experiment 1, we compared the magnitude of the own-group bias between the participants in the single-sex schools (assumed to have more exposure to own-gender faces) and mixed-sex schools. The means are shown in Figure 2. This revealed a main effect of bias type, *F*(1.88, 174.46[7]) = 4.99, $MSE = 0.80$, $p = .009$, $\eta_p^2 = .05$, in which the own-age bias ($M = 1.00$, $SE = 0.11$) was larger than the own-gender bias ($M = 0.61$, $SE = 0.11$), $t(94) = 2.28$, $p = .019$, Cohen's $d = 0.36$. No other pairwise comparisons were significant. The interaction between bias type and type of school, $F(1.88, 174.46) = 1.82$, $MSE = 0.80$, $p = .167$, $\eta_p^2 = .02$, and the main effect of type of school, $F(1, 93) = 0.57$, $MSE = 2.00$, $p = .450$, $\eta_p^2 = .01$, were not significant. In this experiment, all three biases were

[6] Levene's Test of Equality of Variances was significant for the comparison of female faces, $F = 6.07$, $p = .016$, for other-age faces, $F = 12.29$, $p = .001$, and for male faces, $F = 8.37$, $p = .005$, therefore the degrees of freedom were adjusted appropriately.

[7] Mauchly's Test of Spericity was significant, $W(2) = .93$, $p = .043$, indicating the assumption of Sphericity was not met, therefore the Greenhouse-Geisser correction was applied to the degrees of freedom.

significantly different to zero, $t(94) = 5.37$, $p < .001$, $t(94) = 9.96$, $p < .001$, $t(94) = 8.56$, $p < .001$ for the own-gender, -age, and -ethnicity bias respectively. This confirmed that we found an own-gender bias in Experiment 2.

Given that we observed an own-gender bias in Experiment 2 and we explicitly predicted that the own-gender bias would interact with type of school, we compared the size of the own-gender bias across the two types of school. This analysis was not significant, $t(93) = 0.51$, $p = .613$, Cohen's $d = 0.11$.

Figure 5 shows the learning data for distinctiveness ratings. The distinctiveness rating data were subjected to a parallel 2 x 2 x 4 ANOVA. This analysis revealed a main effect of type of face, $F(2.47, 224.53$[8]$) = 13.08$, $MSE = 0.94$, $p < .001$, $\eta_p^2 = .13$. Bonferroni-Šídák corrected pairwise comparisons revealed that other-ethnicity faces ($M = 2.84$, $SE = .12$) were rated as significantly (all ps < .001) less distinctive than female faces ($M = 3.60$, $SE = .14$), male faces ($M = 3.43$, $SE = .10$), and other-age faces ($M = 3.32$, $SE = .13$). No other comparisons were significant. This main effect interacted with participant gender, $F(2.47, 224.53) = 3.37$, $MSE = 0.94$, $p = .027$, $\eta_p^2 = .04$. Bonferroni-Šídák corrected *t*-tests revealed that male faces were rated as more distinctive by boys than by girls, $t(80.31$[9]$) = 2.70$, $p = .008$, Cohen's $d = 0.56$. There were no gender differences in the attractiveness ratings for all other types of faces (all ps > .430).

No other main effects nor interactions were significant: participant gender, $F(1, 91) = 0.15$, $MSE = 3.49$, $p = .704$, $\eta_p^2 < .01$; type of school, $F(1, 91) = 0.69$, $MSE = 3.49$, $p = .408$, $\eta_p^2 = .01$; participant gender by type of face, $F(1, 91) = 0.09$, $MSE = 3.49$, $p = .768$, $\eta_p^2 < .01$; type of face by type of school, $F(2.47, 224.53) = 0.42$, $MSE = 0.94$, $p = .679$, $\eta_p^2 = .01$; nor the three-way interaction, $F(2.47, 224.53) = 0.18$, $MSE = 0.94$, $p = .880$, $\eta_p^2 < .01$.

---

[8] Mauchley's test of sphericity was significant, $W(5) = .74$, $p < .001$, therefore the Greenhouse-Geisser correction was applied to the degrees of freedom.

[9] Levene's test of equality of variances was significant, $F = 5.37$, $p = .023$, therefore the degrees of freedom were adjusted accordingly.

Discussion

In Experiment 2, we found an own-gender bias in all our participants. The own-gender bias did not differ between the groups that had much greater everyday exposure to other-gender faces and the group with less everyday exposure to other-gender faces. We had anticipated that the own-gender bias to be greater in the participants from the single-sex school if experience was the major contributor the own-gender bias as expected given the theory presented by Herlitz and Lovén (2013). Instead, we found experience did not modulate the own-gender bias.

Our distinctiveness data indicates that own-ethnicity faces were found to be more distinctive than other-ethnicity faces, and that male participants find male faces to be more distinctive than female faces. These results are consistent with the notion that participants use more individuation for own-group faces than other-group faces (Levin, 2000) especially for own-ethnicity faces. Given parallel effects were not observed for the own-age bias, nor for female participants, indicates that these biases may be based on different mechanisms. Similarly, the results are indicative of female participants processing faces more deeply than male participants (Rennels & Davis, 2008). Nevertheless, the pattern of significance is not the same as the recognition accuracy data, which therefore indicates that differences in perception of distinctiveness is not solely what is driving the own-gender bias.

The own-gender bias was found to be greater in Experiment 2 than in Experiment 1 (as indicated by a significant *t*-test result comparing the magnitude of the own-gender bias in Experiment 1 with that in Experiment 2, $t(191) = 3.09$, $p = .002$, Cohen's $d = 0.44$). The results of Experiment 2 suggest that the own-gender bias was not found in Experiment 1 because we unintentionally manipulated the participants' motivation to process other-gender faces deeply. Requiring participants to make attractiveness judgements while they learn faces causes them to process own- and other-gender faces to the same level of depth, thus removing the own-gender bias. In addition to motivational differences, this may be the result of attractiveness judgements requiring faces to be judged according to a sex-specific norm rather than an overall face norm (see e.g., Ellis, 1986). Distinctiveness judgements, on the other hand, are based on comparing a face to the overall face norm (Valentine, 1991). This effect of making attractiveness ratings was an unintended consequence of the design of Experiment 1. Experiment 3 was designed to more directly test the effect that the encoding instructions have on the own-gender bias.

Experiment 3

Experiments 1 and 2 combined indicate that the own-gender bias is not based on peer-group experience, but rather on motivation to process faces more or less deeply: attractiveness encoding nullified the own-gender effect observed with distinctiveness encoding. One limitation of Experiments 1 and 2, however, were that the experience factor was based on type of school attended. Participants in same-sex schools may spend 8 hours a day with same-sex peers (and this represents, on average 60%[10] of daily interactions, Hofferth & Sandberg, 2001). However, they will interact with teachers, parents, and siblings who may not be same-sex. While parents and teachers are other-age and may therefore be processed using an age-classifying feature (Levin, 2001; Hills, 2012), siblings could potentially be of a similar age. Potentially, many of the participants may have friends outside of the school environment (including sports and arts clubs) that are other-gender. Therefore, we may not have captured experience completely by using this measure. Experiment 3 aims to address this oversight and replicate the findings regarding the role of motivation from Experiments 1 and 2 in a different sample of participants. Just like participants reported their contact with other-ethnicity faces in Brigham and Malpass's (1985) study, participants in Experiment 3 quantified their experience with the other gender. In this way, the effect of encoding type (attractiveness versus distinctiveness) can be contrasted with other-gender experience to explore which, if either, affects the own-gender bias.

[10] This number is likely to be a low estimate given that since the work was conducted, data has revealed that there is a larger number of hours spent online and using electronic devises. In addition, the remaining interactions are split by involving family members (who are a limited number of people) and peers who are likely to be same-sex.

In Experiment 3, we also explored the face-inversion effect with own- and other-gender faces. It has been argued that expert processing can be measured using the face-inversion effect (Valentine, 1988). This is where the recognition of inverted faces is disproportionately worse than that of upright faces relative to inversion effects in objects (Yin, 1969). The face-inversion effect is a well-established measure of expert processing (Edmunds & Lewis, 2007; Freire, Lee, & Symons, 2000; Gauthier, Tarr, Moylan, Skudlarski, Gore, & Anderson, 2000) and is known to disrupt configural processing (Tanaka & Farah, 1993). It was hypothesised, therefore, that the face-inversion effect would be larger in own-gender faces than other-gender faces and that experience with other-gender faces would moderate this effect (given that experience moderates the face-inversion effect in the own-age bias, Harrison & Hole, 2010; Hills & Lewis, 2011; Hills & Willis, 2016; Kuefner, Macchi Cassia, Picozzi, & Bricolo, 2008). In Experiment 3, we also recorded response times, given that increased response time to process faces can reflect the amount of effort required and the motivation to process faces (Crookes & Rhodes, 2017).

Method

Participants

One hundred and five (81 female, mean age = 21 years) undergraduate students from Anglia Ruskin University took part in this study as part of a course requirement. All participants self-reported that they had normal or correct vision, were ethnically-White, were brought up in a European country, and were sexually attracted to the opposite sex. The sample size was slightly larger than in Experiments 1 and 2 to obtain sufficient male participants.

Materials and Procedure

One hundred and twenty (60 female) faces were taken from the Minear and Park (2004) database. These had a mean age of 21-years (matching that of the participants to avoid the own-age bias).

These were constrained in the same way as in Experiment 1. The stimuli were presented on high-resolution LCD screens using E-Prime Pro 2™.

The procedure for this experiment employed a similar old/new recognition procedure to that of Experiment 1. In the learning phase, participants were presented with 60 faces (30 were female and 30 were male, half of each were inverted). Half the participants rated attractiveness (as in Experiment 1) and half rated distinctiveness (as in Experiment 2). Participants responded themselves using a standard computer keyboard rather than verbalising their responses.

During the distractor phase, participants were asked to provide their age, gender, ethnicity, country they grew up in, and sexuality. In addition, they were asked for how much contact they had with other-gender people and how many other-gender friends they had on a seven-point Likert-type scale (similar to Brigham & Malpass, 1985) to measure other-gender contact. Scores ranged from 1 to 4 (M = 3.51, SE = 0.06) and 1 to 5 (M = 4.25, SE = 0.09) on each measure respectively. Finally, they were asked how much they valued being either female or male (depending on their gender), how proud they were to be a female or male (depending on their gender), and how much belonging to their gender was important to their identity in a similar way as Van Bavel and Cunningham (2012). These questions were asked on a 6-point Likert-type scale with anchor points of 'strongly disagree' or 'very little/few' to 'strongly agree' or 'very many'. During the test phase, participants saw 120 faces (with the same proportion of female/male and inverted/upright faces).

Design

Experiment 3 employed a 2 (instructions: either make attractiveness or distinctiveness judgements) by 2 (face gender: coded as either own-gender or other-gender) by 2 (face orientation: upright and inverted) mixed design. Instructions were manipulated between-subjects, whereas face gender and orientation were manipulated within-subjects. Faces were counterbalanced such that they could appear as a target or distractor an equal number of times and they could appear upright or inverted an equal number of times. Faces were presented in a random order in both the learning and recognition phases of the experiment. In this Experiment, we recorded response time in addition to recognition accuracy. This study received ethical approval from Anglia Ruskin University's Research Ethics Committee.

Results

The old/new responses were converted into *d'* in the same way as in the previous experiments. Mean recognition accuracy is presented in Figure 6 and was subjected to a 2 x 2 x 2 mixed-subjects ANOVA with the factors: instruction (between-subjects), face gender, and face orientation (both within-subjects). This analysis revealed that own-gender faces (M = 1.17, SE = 0.06) were recognised more accurately than other-gender faces (M = 0.80, SE = 0.07), $F(1, 103) = 17.20$, MSE = 0.83, $p < .001$, $\eta_p^2 = .14$. Upright faces (M = 1.28, SE = 0.06) were recognised more accurately than inverted faces (0.69, SE = 0.05), $F(1, 103) = 98.97$, MSE = 0.37, $p < .001$, $\eta_p^2 = .49$. The main effect of instructions was not significant, $F(1, 103) = 0.81$, MSE = 1.00, $p = .372$, $\eta_p^2 = .01$. Finally, the three-way interaction was significant, $F(1, 103) = 7.94$, MSE = 0.29, $p = .006$, $\eta_p^2 = .07$. To explore this interaction, we ran Bonferroni-Šídák corrected (α = .013) within-subjects *t*-tests between the recognition of own- and other-gender faces separately for upright and inverted faces. When making attractiveness judgements, the own-gender bias (M_{own} = 1.43, SE = 0.11; M_{other} = 1.24, SE = 0.09) was not significant for upright faces, $t(54) = 1.58$, $p = .121$, Cohen's d = 0.25, whereas it was when making distinctiveness judgements (M_{own} = 1.47, SE = 0.11; M_{other} = 0.96, SE = 0.13), $t(49) = 3.39$, $p = .001$, Cohen's d = 0.60. For inverted faces, the own-gender bias (M_{own} = 0.98, SE = 0.11; M_{other} = 0.45, SE = 0.09) was significant when making attractiveness judgements, $t(54) = 3.06$, $p = .003$, Cohen's d = 0.71, whereas it was not when making distinctiveness judgements (M_{own} = 0.79, SE = 0.10; M_{other} = 0.54, SE = 0.08), $t(49) = 1.92$, $p = .061$, Cohen's d = 0.39.

To assess the role of expert face processing, we explored the magnitude of the face-inversion effect across each condition. When making attractiveness judgements, the magnitude of the face-inversion effect (calculated as the recognition accuracy for inverted faces subtracted from that of upright faces) was not significantly different between own- (*M* = 0.68, *SE* = 0.11) and other-gender (*M* = 0.43, *SE* = 0.11) faces, $t(49) = 1.77$, $p = .083$, Cohen's $d = 0.32$. However, when making distinctiveness judgements, the face-inversion effect was larger for own-gender faces (*M* = 0.79, *SE* = 0.13) than other-gender faces (*M* = 0.45, *SE* = 0.09), $t(49) = 2.23$, $p = .030$, Cohen's $d = 0.43$. The interaction between face gender and instructions was not significant, $F(1, 103) = 0.01$, $MSE = 0.83$, $p = .941$, $\eta_p^2 < .01$, nor were the interactions between orientation and instructions, $F(1, 103) = 0.35$, $MSE = 0.37$, $p = .557$, $\eta_p^2 < .01$, and face gender and orientation, $F(1, 103) = 0.16$, $MSE = 0.29$, $p = .693$, $\eta_p^2 < .01$. To assess whether self-reported contact and quality of contact affected the magnitude of the own-gender bias, we ran non-parametric correlations between contact and quality of contact and the own-gender bias (calculated as subtracting the recognition accuracy for other-gender faces from own-gender faces). These revealed that neither contact, nor quality of contact, correlated with the own-gender bias for upright faces, $r(48) = .05$, $p = .712$ and $r(48) = .02$, $p = .885$, respectively following making distinctiveness judgements and $r(53) = .06$, $p = .641$ and $r(48) = .06$, $p = .671$ following making attractiveness judgements (nor for inverted faces $r(48) = .16$, $p = .283$ and $r(48) = .04$, $p = .783$ respectively following making distinctiveness judgements and $r(48) = .02$, $p = .912$ and $r(48) = .14$, $p = .322$ following making attractiveness judgements).

We assessed recognition response times in a parallel 2 x 2 x 2 analysis. Means are shown in Figure 6. This revealed that own-gender faces (*M* = 1158, *SE* = 22) were viewed for longer than other-gender faces (*M* = 1097, *SE* = 24), $F(1, 103) = 16.15$, $MSE = 24447$, $p < .001$, $\eta_p^2 = .14$. This effect interacted with task instructions, $F(1, 103) = 7.87$, $MSE = 24447$, $p = .006$, $\eta_p^2 = .07$. Paired-sample *t*-tests showed that when making distinctiveness judgements, participants were quicker at recognising other- (*M* = 1057, *SE* = 36) than own-gender (*M* = 1161, *SE* = 31) faces, $t(49) = 4.25$, $p < .001$, Cohen's $d = 0.44$, but the difference in response time for own- (*M* = 1156, *SE* = 32) and other-gender (*M* = 1137, *SE* = 31) faces was not significant when making attractiveness judgements, $t(49) = 0.99$, $p =$

.327, Cohen's d = 0.09. Participants were also faster at recognising upright faces (M = 1095, SE = 20) than inverted faces (M = 1161, SE = 25), $F(1, 103) = 26.59$, MSE = 14351, $p < .001$, $\eta_p^2 = .21$.

The main effect of task instructions was not significant, $F(1, 103) = 0.74$, MSE = 196631, $p = .391$, $\eta_p^2 = .01$. The interaction between orientation and task instructions was not significant, $F(1, 103) = 0.85$, MSE = 14351, $p = .360$, $\eta_p^2 = .01$, nor were the interactions between face gender and orientation, $F(1, 103) = 0.48$, MSE = 15279, $p = .490$, $\eta_p^2 = .07$, and the three-way interaction, $F(1, 103) = 0.15$, MSE = 15279, $p = .699$, $\eta_p^2 < .01$.

The learning phase rating data were subjected to a parallel analysis, shown in Figure 7. This revealed a main effect of face gender, $F(1, 103) = 15.90$, MSE = 1.08, $p < .001$, $\eta_p^2 = .13$, with own-gender faces (M = 3.47, SE = 0.12) receiving higher ratings than other-gender faces (M = 3.07, SE = 0.11). Upright faces (M = 3.52, SE = 0.12) also received higher ratings than inverted faces (M = 3.01, SE = 0.11), $F(1, 103) = 22.84$, MSE = 1.18, $p < .001$, $\eta_p^2 = .18$. There was a marginal interaction between these factors, $F(1, 103) = 3.68$, MSE = 1.30, $p = .058$, $\eta_p^2 = .03$. For own-gender faces, upright faces (M = 3.83, SE = 0.14) were rated higher than inverted faces (M = 3.11, SE = 0.14), $t(104) = 5.03$, $p < .001$, Cohen's d = 0.50. For other-gender faces, upright faces (M = 3.21, SE = 0.14) were not rated significantly higher than inverted faces (M = 2.92, SE = 0.14), $t(104) = 1.78$, $p = .079$, Cohen's d = 0.20.

The main effect of task instructions was not significant, $F(1, 103) < 0.01$, MSE = 4.48, $p = .984$, $\eta_p^2 < .01$. The interaction between face gender and task instructions was not significant, $F(1, 103) = 0.01$, MSE = 1.08, $p = .942$, $\eta_p^2 < .01$, nor was the interaction between orientation and task instructions, $F(1, 103) = 1.15$, MSE = 1.18, $p = .287$, $\eta_p^2 = .01$, nor the three-way interaction, $F(1, 103) = 0.54$, MSE = 1.30, $p = .465$, $\eta_p^2 = .01$.

The learning response time data, also shown in Figure 7, were subjected to a parallel 2 x 2 x 2 ANOVA. This revealed an identical pattern of significance as the recognition response time data. This revealed that own-gender faces (*M* = 1082, *SE* = 22) were viewed for longer than other-gender faces (*M* = 1020, *SE* = 24), $F(1, 103) = 14.79$, $MSE = 28011$, $p < .001$, $\eta_p^2 = .13$. This effect interacted with task instructions, $F(1, 103) = 4.38$, $MSE = 28011$, $p = .039$, $\eta_p^2 = .04$. Paired samples *t*-tests showed that when making distinctiveness judgements, participants were quicker at recognising other- (*M* = 981, *SE* = 36) than own-gender faces (*M* = 1078, *SE* = 32) faces, $t(49) = 3.87$, $p < .001$, Cohen's d = 0.40, but the difference in response time for own- (*M* = 1086, *SE* = 32) and other-gender (*M* = 1057, *SE* = 32) faces was not significant when making attractiveness judgements, $t(49) = 1.35$, $p = .183$, Cohen's d = 0.13. Participants were also faster at recognising upright faces (*M* = 1016, *SE* = 20) than inverted faces (*M* = 1085, *SE* = 25), $F(1, 103) = 26.86$, $MSE = 18554$, $p < .001$, $\eta_p^2 = .21$.

The main effect of task instructions was not significant, $F(1, 103) = 0.95$, $MSE = 195759$, $p = .331$, $\eta_p^2 = .01$. The interaction between orientation and task instructions was not significant, $F(1, 103) = 0.86$, $MSE = 18554$, $p = .357$, $\eta_p^2 = .01$, nor were the interactions between face gender and orientation, $F(1, 103) = 1.07$, $MSE = 17992$, $p = .303$, $\eta_p^2 = .01$, and the three-way interaction, $F(1, 103) = 0.09$, $MSE = 17992$, $p = .763$, $\eta_p^2 < .01$.

Discussion

The results of Experiment 3 are consistent with Experiments 1 and 2 in terms of the fact that perceptual experience (this time measured by self-report) did not moderate the effect of the own-gender bias. Furthermore, the own-gender bias was only present when participants were making distinctiveness judgements and not when making attractiveness judgements. Since response time can be considered a metric for effort of processing (Crookes & Rhodes, 2017), our response time data indicate that participants put more effort into recognising own-gender faces than other-gender faces when making distinctiveness judgements, but not when making attractiveness judgements. The rating data is consistent with this view, that participants rated own-gender faces as more distinctive/attractive than other-gender faces. These results reinforce the fact that the own-gender bias is the result of motivation to process own-gender faces more deeply than other-gender faces.

By providing participants a reason to engage in elaborative processing for other-gender faces, the own-gender bias can be reduced. Experience, either due to exposure in schools (Experiments 1 and 2) or self-reported contact (Experiment 3) do not relate to the magnitude of the own-gender bias. This deeper processing is revealed through the interaction with face orientation. Expert processing (as indexed by the face-inversion effect) is employed for own-gender faces and other-gender faces when making attractiveness judgements, but when making distinctiveness judgements, the expert processing is employed only for own-gender faces and not other-gender faces.

Experiment 4

Thus far, we have demonstrated that the own-gender bias is not due to increased exposure to own-gender faces. Instead, the effect is based on motivation to process own-gender faces more deeply than other-gender faces. The motivation can be negated by encouraging participants to process other-gender faces more deeply by asking participants to provide attractiveness judgements. Attractiveness judgements encourage focusing on other-gender faces because of the importance of facial attractiveness in sexual and romantic engagement (Buss, 1985; Rhodes, et al., 2005). To confirm this hypothesis, it was tested whether gay participants would show an own-gender-bias even when making attractiveness ratings. If sexual or romantic interest is driving the deeper processing of other-gender faces, then participants who have sexual or romantic interest in own-gender people will show a larger own-gender bias than participants who have sexual or romantic interest in other-gender people. This hypothesis was directly tested in Experiment 4. Adult participants were employed as their sexuality will be more likely to be established and stable than for children.

Method

Participants

An opportunity sample of 48 participants were recruited from Lesbian, Gay, and Bisexual groups and other non-sexuality-focused student groups in the Cambridge area. Twenty-four (13 female) self-defined as gay (age range 19 - 42 years, mean = 26 years). Twenty-four (11 female) self-defined as straight (age range 18 - 54 years, mean = 33 years). All participants self-reported that they had normal or corrected vision and were ethnically White.

Materials

One hundred and sixty images of 80 faces (two images each) were collected from www.facebook.com and www.match.com in a similar manner to Rule and Ambady (2008). Only those in which the photograph displayed a frontal view, with a neutral or happy expression with no extraneous features (such as glasses, beards, or jewelry) were selected. These were compressed to 72 dpi resolution, edited such they only showed the head; all backgrounds were masked out using Photoshop CS™. Faces were presented 150 mm by 100 mm during the learning phase of the experiment and 100 mm by 67 mm in the recognition phase. All stimuli were of White males and females from the UK but not local to the University in which the experiment was conducted. All of the faces were ethnically White. All stimuli were presented using E-Prime Pro 2 on 19 inch colour LCD monitor from a Compaq Presario CQ81-405SA; the screen resolution was 1366 x 768 px.

Design and Procedure

A 2 (participant sexuality: gay or straight) by 2 (face gender: own- or other-gender) mixed-subjects design was employed. Participant sexuality was a between-subjects variable and face gender was a within-subjects variable. Recognition accuracy (*d'*) was measured. Faces were counterbalanced such that they appeared as a target and as a distractor an equal number of times. Faces were presented in a random order. This study received ethical approval from Anglia Ruskin University's Research Ethics Committee.

A similar old/new recognition paradigm was employed in this Experiment as in Experiments 1-3. The only differences were the number of faces presented during the learning phase (40: 20 of which were female) and the test phase (80: 40 of which were target faces). During the learning phase, all participants rated the faces for attractiveness as it was expected that this condition should increase the likelihood of finding the own-gender bias in the gay participants and decrease the chance of finding it in straight participants. At the end of the Experiment, participants confirmed that none of the faces were familiar to them.

Results

Mean recognition accuracy is presented in Figure 8. These data were subjected to a 2 x 2 mixed-subjects ANOVA with the factors participant sexuality and face gender. This revealed that own-gender faces (M = 2.37, SE = 0.11) were recognised more accurately than other-gender faces (M = 1.66, SE = 0.11),$F(1, 46) = 24.24$, MSE = 0.50, $p < .001$, $\eta_p^2 = .35$. This effect interacted with participant sexuality, $F(1, 46) = 11.78$, MSE = 0.50, $p = .001$, $\eta_p^2 = .20$. Within-subjects t-tests revealed that the own-gender bias (M_{own} = 2.59, SE = 0.14, M_{other} = 1.38, SE = 0.15) was significant for gay, $t(23) = 5.45$, $p < .001$, Cohen's d = 1.69, but not straight participants (M_{own} = 2.15, SE = 0.18, M_{other} = 1.93, SE = 0.15), $t(23) = 1.16$, $p = .258$, Cohen's d = 0.27. The main effect of participant sexuality was not significant, $F(1, 46) = 0.11$, MSE = 0.66, $p = .740$, $\eta_p^2 < .01$.

Mean recognition response time data are presented in Figure 8. These data were subjected to a parallel 2 x 2 mixed-subjects ANOVA. This revealed a main effect of face gender, $F(1, 46) = 6.95$, MSE = 25455, $p = .011$, $\eta_p^2 = .13$, in which own-gender faces (M = 1074, SE = 30) were recognised slower than other-gender faces (M = 988, SE = 23). The main effect of participant sexuality was not significant, $F(1, 46) = 0.10$, $MSE = 41950$, $p = .921$, $\eta_p^2 < .01$, nor was the interaction, $F(1, 46) = 0.57$, $MSE = 25450$, $p = .454$, $\eta_p^2 = .01$.

The distinctiveness rating data, summarised in Figure 9, were subjected to parallel mixed-subjects ANOVA. This revealed that own-gender faces ($M = 5.54$, $SE = 0.17$) were rated as more distinctive than other-gender faces ($M = 5.00$, $SE = 0.16$), $F(1, 46) = 10.40$, $MSE = 0.66$, $p = .002$, $\eta_p^2 = .18$. The main effect of participant sexuality was not significant, $F(1, 46) = 0.18$, $MSE = 1.94$, $p = .678$, $\eta_p^2 < .01$, nor was the interaction between face gender and participant sexuality, $F(1, 46) = 0.67$, $MSE = 0.66$, $p = .419$, $\eta_p^2 = .01$.

The learning response time data, summarised in Figure 9, were subjected to a parallel ANOVA. This revealed that participants were faster to respond to other-gender ($M = 911$, $SE = 23$) than own-gender ($M = 1004$, $SE = 31$) faces, $F(1, 46) = 7.99$, $MSE = 25646$, $p = .007$, $\eta_p^2 = .15$. The main effect of participant sexuality was not significant, $F(1, 46) = 0.02$, $MSE = 46053$, $p = .890$, $\eta_p^2 < .01$, nor was the interaction, $F(1, 46) = 0.57$, $MSE = 25646$, $p = .453$, $\eta_p^2 = .01$.

Discussion

Experiment 4 replicated the findings that attractiveness judgements led to an absence of the own-gender bias for the heterosexual participants. However, the gay participants demonstrated a significant own-gender bias. This further reinforces the notion that the own-gender bias can be manipulated through motivation. Consistently, the reaction time and rating data indicate that more effort was engaged in when processing own-gender relative to other-gender faces given the longer time engaged in rating and recognising these faces. When considering faces for attractiveness, one is more motivated to process those that one is romantically and sexually interested in to more depth. Since gay people are attracted to own-gender people, they are typically more motivated to process own-gender faces more deeply than other gender faces. Straight participants are motivated to process other-gender faces more deeply when making attractiveness judgements, thereby removing the own-gender bias. While we have made this claim, we cannot rule out that this effect is due to increased experience with faces of one's own-gender in gay participants. Such an experience-base argument would require that gay people have considerably more social contact with their own gender than straight people have with their own gender. We are not aware of any evidence that this is the case.

The results from Experiment 4 are consistent with the theoretical approach devised by Scherf et al. (2012) and tested by Picci and Scherf (2016). In their approach, they indicated that pubertal changes in hormones cause changes in the role and function of face perception. Specifically, prior to puberty, face recognition is driven by the need for attachment to the primary caregiver. However, during puberty, sexual and romantic interest develops and this causes face perception to be primarily tuned toward peers. The results of experiment 4 are consistent with this approach as it highlights that sexual and romantic interest of gay participants causes them to attend to own-gender faces more so than other-gender faces.

General Discussion

Across four experiments, we have shown that the own-gender bias can be manipulated by the encoding instructions that participants are provided with. Making neutral distinctiveness judgements results in a significant own-gender bias. This bias disappears if participants are asked to make attractiveness judgements. It is argued that people are motivated to process other-gender faces more deeply when making attractiveness ratings rather than distinctiveness ratings. Response time data in Experiments 3 and 4 are consistent with this particular view: participants spent more effort in encoding and recognising own-gender faces than other-gender faces. Experiments 1 to 3 evaluated the effect of participants' experience with other-gender faces on the size of the own-gender bias. The size of the own-gender bias was not significantly affected by perceptual experience of the other gender - whether this was determined by the school attended (single- versus mixed-sex) or self-reported contact with the other gender. Experiment 4 demonstrated that it is the potential sexual and romantic interest (Buss, 1985; Rhodes, et al., 2005) that leads to the moderating effect that making attractiveness ratings has on the own-gender bias. If the rater does not have a potential sexual or romantic interest in the person being rated then making attractiveness ratings does not decrease the own-gender bias.

Herlitz and Lovén (2013) reported that extensive and intense developmental experience (Ashear & Snortum, 1971; Exline, 1963; Exline, et al., 1965; Field, et al., 1984; Levine & Sutton-Smith, 1973; Osofsky & O'Connell, 1977; Rennels & Davis, 2008) with female faces leads to girls and women showing a large own-gender bias. Own-gender peer groups may negate this experience for boys and men, but not always (Feinman & Entwisle, 1976; Ge et al., 2008). The crux of Herlitz and Lovén's (2013) theory is that experience is a main driving force behind the own-gender bias. In the present work, experience was not found to moderate the magnitude of the own-gender bias. To reconcile the differences between the present data and the conclusions of Herlitz and Lovén, we must explore what the empirical data presented in Herlitz and Lovén's meta-analysis show.

Much empirical data highlights that the primary caregiver of newborn infants is a female. Most early intense contact is between a mother and a newborn infant (Connellan, et al., 2000) leading to increased eye contact between a mother and child (Hittelman & Dickes, 1979). This occurs because much feeding of a newborn infant is done by the mother, and the optimal visual acuity of a newborn infant is to the face of the mother during breast feeding (Dobson & Teller, 1978). Therefore, this increased experience and exposure to the mother's face also increases interest with it. This pattern continues as many caregivers (and primary school teachers) are female during early childhood (Ashear & Snortum, 1971; Exline, 1963; Exline, Gray, & Schuette, 1965; Field, Cohen, Garcia, & Greenberg, 1984; Levine & Sutton-Smith, 1973; Osofsky & O'Connell, 1977). While this does increase experience with female faces, the experience is partially balanced by extensive experience with peers who are male. In fact, it might be that this exposure means that children are more motivated to process female faces. Therefore, these previous studies might actually be confounding exposure with interest and motivation. Indeed, this exposure might lead to motivation to process female faces, and it is this motivation that remains, leading to the reliable own-gender bias in women. In men, the own-gender bias can also be caused by interest and motivation.

The finding that making attractiveness judgements affects face perception differently to making distinctiveness judgements is a novel finding, but one which could have been predicted. Establishing sexual and romantic relationships is an important part of human life (Bancroft, 2009) and attractiveness is an important part of finding a sexual and romantic partner (Buss, 1985; Rhodes, et al., 2005). Attractiveness is often seen as a sign of good genes (Scheib, Gangestad, & Thornhill, 1999), and creates a bias in which attractive people are considered more intelligent (Olsen & Inglehard, 2011), more successful (Easly, Ashmore, Makhijani, & Longo, 1991), to give better performances (Ryan & Costa-Giomi, 2004; Wapnick, Darrow, Kovacs, & Dalrymple, 1997), and less likely to be considered guilty in mock jury decisions when making snap judgements (Patry, 2008). This highlights an attractiveness bias (Shahani, Dipboye, & Gehrlein, 1993), where attractive people are considered of higher importance. Research has shown that when participants consider a face to be of higher value they are more likely to recognise it subsequently (Shriver et al., 2008, see also Motta-Mena et al., 2016). Furthermore, Feingold (1990) has identified from behavioural studies that judging attractiveness is vital for both men and women and there is not a large gender difference in the importance of attractiveness on mate choice. This means that judging attractiveness is an important ability for participants and attractive faces are likely to be processed more deeply. By extrapolation, making attractiveness judgements forces participants to consider this variable and therefore will show more interest in those faces. This view is consistent with the theory that puberty changes the way in which children attend to and process faces (Scherf et al., 2012). Indeed, Rule, Rosen, Slepian, and Ambady (2011) have found that women are better at judging male sexuality from faces when they are in the most fertile part of their ovulation cycle. This highlights how sexual interest can affect the cognitive and perceptual system.

In this study, we were able to manipulate interest in faces by making participants make attractiveness judgements. One aspect of the present results is that the study was conducted in participants whose mean age ranged between 15 and 21 years across these four Experiments. This age range, late adolescence, is associated with extensive interest in romantic and sexual relationships (Garcia & Fisher, 2006; Miller & Benson, 1999). Many adolescents report being involved in serious romantic or sexual relationships (Collins, 2003). In this age range, many people experience their first long-term relationship (Mercer et al., 2013). The present findings may therefore be limited to the present age group because their interest in developing romantic and sexual relationships is heightened. This empirical question can only be tested by exploring the own-gender bias in other age groups.

Returning to the issue that attractiveness judgements reduce the own-gender bias by enhancing recognition of other-gender faces, we must now explore why this happens. In the introduction, we highlighted one potential mechanism for enhancing expertise in face processing: holistic processing. Holistic processing occurs when the whole face is processed as a gestalt (Rossion, 2008). It is a type of configural processing (Maurer, et al., 2001). In Experiment 3, the face-inversion effect was observed for own-gender faces and other-gender faces. When making distinctiveness judgements, it was larger for own-gender than other-gender faces, replicating Man and Hills (2017). This indicates that there is more expert processing applied to own-gender faces. Interestingly, the face-inversion effect was larger for other-gender faces when making attractiveness judgements than when making distinctiveness judgements. This indicates that requiring participants to make attractiveness judgments during encoding of faces caused them to process other-gender faces with more appropriate expert processing (as identified by the face-inversion effect) than when they made distinctiveness judgements. This provides further evidence to support the fact that the own-gender bias is caused by participants engaging the less expert perceptual system for other-gender faces but shows that this can be moderated by asking people (if they are heterosexual) to make attractiveness ratings when encoding the faces.

Given the failure to find an effect of perceptual experience on the own-gender bias, we discount the experienced-based perceptual expertise accounts (e.g., Valentine, 1991) as the cause of the bias. It is suggested here that the own-gender bias is caused by a lack of motivation to apply expert processing to other-gender faces. Given that we manipulated when our participants engaged in expert processing by requiring them to make different types of judgements during encoding, we manipulated the amount of attention participants applied to face recognition through motivation. These results are consistent with Hugenberg et al.'s (2010) categorisation-individuation model. When participants are motivated to process other-gender faces deeply, they will engage the expert perceptual system to do so. This bias is, therefore, based primarily on motivation, and social categorisation (given results from Rehnman & Herlitz, 2007) rather than experience. These results highlight how flexible the face recognition system is: expert processing is not always applied: It is applied when the conditions (motivation, instructions, and context) allow for it.

In summary, this study has provided evidence that the own-gender bias is not caused by enhanced developmental experience with own-gender faces. Instead, the own-gender bias is caused by motivation to process own-gender faces more deeply than other-gender faces. Other-gender faces can be processed as deeply when straight participants are instructed to make attractiveness judgements as this enhances their interest in other-gender faces (the same manipulation causes a larger own-gender bias in gay participants). This enhanced processing is due to a greater reliance on expert face processing for own-gender than other-gender faces.

Acknowledgements

The authors would like to thank Bryn Hafren School in Barry, Monmouth School for Boys, Haberdasher’s Monmouth School for Girls, Queen Elizabeth Hospital, Bristol, Rumney High School, Llanishen High School, and Bishop Hedley Church of Wales School for allowing research to take place. Experiments 1 and 2 were supported by grant PTA-030-2003-00524 from the ESRC to Peter Hills.

References

Ashear, V., & Snortum, J. R. (1971). Eye contact in children as a function of age, sex, social and intellective variables. *Developmental Psychology*, *4*(3), 479.

Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition*, *19*(1), 63-71.

Bancroft, J. (2009). *Human sexuality and its problems*. Elsevier Health Sciences.

Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for faces versus single views of faces. *Memory & Cognition*, *14*(5), 371-381.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706-712.

Blandón-Gitlin, I., Pezdek, K., Saldivar, S., & Steelman, E. (2014). Oxytocin eliminates the own-race bias in face recognition memory. *Brain Research*.

Brigham, J. C. & Barkowitz, P. (1978). Do “They all look alike”? the effect of race, sex, experience, and attitudes on the ability to recognize faces*. Journal of Applied Social Psychology, 8,* 306 – 318.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673.

Brigham, J. C. & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41*, 139 – 155.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482-496.

Buss, D. M. (1985). Human mate selection. *American Scientist, 73,* 47-51.

Carey, S., De Schonen, S., & Ellis, H. D. (1992). Becoming a face expert [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *335*(1273), 95-103.

Charness, N. (1981). Search in chess: age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 467 – 476.

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*, *48*(4), 879-894.

Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of research on adolescence*, *13*(1), 1-24.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant behavior and Development*, *23*(1), 113-118.

Crookes, K. & Rhodes, G. (2017). Poor recognition of other-race faces cannot always be explained by a lack of effort. *Visual Cognition,* 1-12.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of face. *Perception & Psychophysics, 10,* 393 – 396.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of experimental child psychology, 23*(1), 1-22.

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vision Research*, *18*(11), 1469-1483.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin*, *110*(1), 109-128.

Edmunds, A. J. & Lewis, M. B. (2007). The effect of rotation of configural encoding in a face-matching task. *Perception, 36,* 446 – 460.

Ellis, H. D. (1981). Theoretical aspects of face recognition. In G.M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), *Perceiving and Remembering Faces.* London: Academic Press.

Exline, R. V. (1963). Explorations in the process of person perception: Visual interaction in relation to competition, sex, and need for affiliation. *Journal of personality*, *31*(1), 1-20.

Exline, R., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, *1*(3), 201.

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59*(5), 981-993. doi:10.1037/0022-3514.59.5.981

Feinman, S. & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506 – 510.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and development*, *7*(1), 19-25.

Forsyth, R., Cinque, C., & Bukach, C. M. (2017). Motivational instructions reduce the other-race effect for participants with low (but not high) quality other-race relationships. Poster presented at the 58th Annual Meeting of the Psychonomic Society, Vancouver, Canada, 8th-12th November 2017

Freire, A., Lee, K., & Symons, L. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception, 29*, 159 – 170.

Garcia, J. R. & Fisher, H. E. (2006). Why we hook up: Searching for sex or looking for love? In S. Tarrant (Ed.), *Gender, Sex, and Politics: In the streets and between the sheets in the 21st century.* New York: Routledge.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12,* 495 – 504.

Ge, L., Anzures, G., Wang, Z., Kelly, D. J., Pascalis, O., Quinn, P. C., ... & Lee, K. (2008). An inner face advantage in children's recognition of familiar peers. *Journal of Experimental Child Psychology*, *101*(2), 124-136.

Harrison, V., & Hole, G. J. (2009). Evidence for a contact-based explanation of the own-age bias in face recognition. *Psychonomic bulletin & review*, *16*(2), 264-269.

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21*(9-10), 1224-1247.

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology*, *46*(2), 445-448.

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: a meta-analytic review. *Visual Cognition*, *21*(9-10), 1306-1336.

Hall, G. (1991). *Perceptual and Associative Learning.* Oxford: Oxford University Press.

Hills, P. J. (2012). A developmental study of the own-age face recognition bias in children. *Developmental psychology*, *48*(2), 499.

Hills, P. J. & Lewis, M. B. (2011). The own-age face recognition bias in children and adults. *The Quarterly Journal of Experimental Psychology, 64,* 17 – 23. Doi: 10.1080/17470218.2010.537926

Hills, P. J., Holland, A. M., & Lewis, M. B. (2010). Aftereffects for face attributes with different natural variability: Children are more adaptable than adolescents. *Cognitive Development, 25,* 278-289.

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition*, *129*(3), 586-597.

Hills, P. J., & Willis, S. F. (2016). Children view own-age faces qualitatively differently to other-age faces. *Journal of Cognitive Psychology*, *28*(5), 601-610.

Hittelman, J. H., & Dickes, R. (1979). Sex differences in neonatal eye contact time. *Merrill-Palmer Quarterly of Behavior and Development*, *25*(3), 171-184.

Hofferth, S. L., & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family*, *63*(2), 295-308.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychological review*, *117*(4), 1168.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). *Designing clinical research : an epidemiologic approach* (4th Ed.) Philadelphia, PA: Lippincott Williams & Wilkins.

Johnston, R. A. & Ellis, H. D. (1995). Age effects in the processing of typical and distinctive faces. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 48,* 447-465.

Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 811.

Leeb, R. T., & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles*, *50*(1-2), 1-14.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature*, *410*(6831), 890-890.

Levin, D.T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129,* 559 – 574.

Levine, M. H., & Sutton-Smith, B. (1973). Effects of age, sex, and task on visual behavior during dyadic interaction. *Developmental Psychology*, *9*(3), 400.

Lewis, M. B. (2004). Face-space-r: towards a unified account of face recognition. *Visual Cognition, 11,* 29 – 70.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences*, *103*(17), 6771-6774.

Light, L. L., Hollander, S., & Kayra-Stuart, F. (1981). Why attractive people are harder to remember. *Personality and Social Psychology Bulletin, 7,* 269 – 276.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 212 - 228.

Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women's own-gender bias in face recognition memory. *Experimental psychology*.

Luce, T. S. (1974). The role of experience in inter-racial recognition. *Personality and Social Psychology Bulletin, 1,* 39 – 41.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of personality and social psychology*, *13*(4), 330.

Man, T. W., & Hills, P. J. (2016). Eye-tracking the own-gender bias in face recognition: Other-gender faces are viewed differently to own-gender faces. *Visual Cognition*, *24*(9-10), 447-458.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255 – 260.

McKelvie, S. J. (1987). Sex differences, lateral reversal, and pose as factors in recognition memory for photographs of faces. *The Journal of general psychology*, *114*(1), 13-37.

McKone, E., Aitkin, A., & Edwards, M. (2005). Categorical and coordinate relations in faces, or Fechnar's law and face-space instead? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1181 – 1198.

McLaren, I. P. L. (1997). Categorisation and perceptual learning: an analogue of the face inversion effect. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *50*, 257 – 273.

MacMillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Cambridge, UK: Cambridge University Press.

Mercer, C. H., Tanton, C., Prah, P., Erens, B., Sonnenberg, P., Clifton, S., ... & Copas, A. J. (2013). Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *The Lancet*, *382*(9907), 1781-1794.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science*, *31*(5), 911-924.

Miller, B. C., & Benson, B. (1999). Romantic and sexual relationship development during adolescence (pp 99-121). In W. Furman, B Bradford Brown, & C Feiring (Eds), *The development of romantic relationships in adolescence*. Cambridge, UK: Cambridge University Press.

Minear, M. & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research, Methods, Instruments, and Computers, 34,* 630 – 633.

Motta-Mena, N., Picci, G., & Scherf, S. (2016). Dominance elicits the own-gender bias in males. *Journal of Vision, 16,* 497. doi:10.1167/16.12.497

Mundy, M. E., Dwyer, D. M., & Honey, R. C. (2006). Inhibitory associations contribute to perceptual learning in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 178 – 184.

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 124 – 138.

Nelson, C. A. (2001). The development and neural bases in face recognition. *Infant and Child Development, 10,* 3 – 18.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Olsen, J. A., & Inglehart, M. R. (2011). Malocclusions and perceptions of attractiveness, intelligence, and personality, and behavioral intentions. *American Journal of Orthodontics and Dentofacial Orthopedics*, *140*(5), 669-679.

Osofsky, J. D., & O'Connell, E. J. (1977). Patterning of newborn behavior in an urban population. *Child Development*, 532-536.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296,* 1321 – 1323.

Patry, M. W. (2008). Attractive but guilty: Deliberation and the physical attractiveness bias. *Psychological reports*, *102*(3), 727-733.

Picci, G. & Scherf, K. S. (2016). From caregivers to peers: Puberty shapes human face perception. *Psychological Science, 27,* 1461 - 1473. doi: 10.1177/0956797616663142

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, *14*(3), 289-296.

Rehnman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta psychologica*, *124*(3), 344-355.

Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior and Development*, *31*(4), 665-678.

Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success?. *Evolution and human behavior*, *26*(2), 186-201.

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average configurations attractive only because of their symmetry? *Psychological Science, 10,* 52-58.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 328.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision research*, *51*(3), 333-342.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, *128*(2), 274-289.

Rule, N. O. & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology, 44,* 1100 - 1105. doi: 10.1016/j.jesp.2007.12.001

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review*, *14*(4), 687-692.

Rule, N. O., Garrett, J. V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PloS one*, *5*(12), e14241.

Rule, N. O., Rosen, K. S., Slepian, M. L., & Ambady, N. (2011). Mating interest improves women's accuracy in judging male sexual orientation. *Psychological Science, 22*(7), 881.

Ryan, C., & Costa-Giomi, E. (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education*, *52*(2), 141-154.

Saariluoma, P. (1994). Location coding in chess. *Quarterly Journal of Experimental Psychology, 47A,* 607 – 630.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research, 35*, 519 – 527.

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1431), 1913-1917.

Scherf, K. S., Behrmann, M., & Dahl., R. E. (2012). Facing changes and changing faces in adolescence: A new model investigating adolscent-specific interactions between pubertal brain and behavioral development. *Developmental Cognitive Neuroscience, 2,* 199-219.

Scherf, K. S., & Scott, L. S. (2012). Connecting developmental trajectories: Biases in face processing from infancy to adulthood*. Developmental psychobiology*, 54(6), 643-663.

Scott, L. S., Shannon, R. W., & Nelson, C. A. (2005). Behavioral and electrophysiological evidence of species-specific face processing. *Cognitive, Affective, and Behavioral Neuroscience, 5,* 405 – 416.

Scott, L. S., Shannon, R. W., & Nelson, C. A. (2006). Neural correlates of human and monkey face processing in 9-month-old infants. *Infancy, 10,* 171 – 186.

Shahani, C., Dipboye, R. L., & Gehrlein, T. M. (1993). Attractiveness bias in the interview: Exploring the boundaries of an effect. *Basic and Applied Social Psychology*, *14*(3), 317-328.

Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality and Social Psychology Bulletin*, *34*(2), 260-274.

Simon, H. A. & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review, 76,* 473 – 483.

Simon, H. A. & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5,* 29 – 46.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups – an integration of theories. *Psychology, Public Policy, and Law, 7,* 36 – 97.

Sporer, S. L., & Horry, R. (2011). Recognizing faces from ethnic in-groups and out-groups: Importance of outer face features and effects of retention interval. *Applied Cognitive Psychology*, *25*(3), 424-431.

Swets, J. A. (1966). *Signal detection and recognition by human observers.* New York: John Wiley & Sons Inc.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161 – 204.

Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: the effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A,* 671 – 703.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*(10), 1996-2019.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, 20(3), 291-302.

Wagner, D. A. & Scurrah, M. J. (1971). Some characteristics of human problem solving in chess. *Cognitive Psychology, 2,* 451 – 478.

Walker, P. M. & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology, 20*, 461 – 475.

Wapnick, J., Darrow, A. A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45*(3), 470-479.

Wilson, J. P., See, P. E., Bernstein, M. J., Hugenberg, K., & Chartier, C. (2014). Differences in anticipated interaction drive own group biases in face memory. *PloS one*, *9*(3), e90668.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta psychologica*, *114*(1), 101-114.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141 – 145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747-759.

Young, S. G., Bernstein, M. J., & Hugenberg, K. (2010). When do own-group biases in face recognition occur? Encoding versus post-encoding. *Social Cognition*, *28*(2), 240-250.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition a critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116-142.

Table 1.

Comparison of demographics across single-sex and mixed-sex schools for the school used in Experiment 1 and 2 (ranges are presented in parentheses). Percentage of students entitled to free-school meals was used as a proxy for social economic status. Official school rating is based on a score out of 4 (with 1 is high). Average school attainment is based on percentage of students achieving 3 A-C grades at A-Level.

	Single-sex schools		**Mixed-sex schools**	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2
Socio-economic status	7.7% (0-19)	11.5% (0-23)	15% (7-23)	13% (7-19)
Official School Rating	1.52 (1.29-1.86)	1.64 (1.57-1.64)	2.36 (1.71-3.00)	2.57 (1.71-3.00)
Average School Attainment	86.0% (77-94)	89.0% (71-94)	85.5% (77-94)	87.0% (79-94)

Figure Captions

Figure 1. Mean recognition accuracy (*d'*) for own- and other-gender, other-age, and other-ethnicity faces, split by type of school (Experiment 1). Error bars represent standard error of the mean.

Figure 2. Mean magnitude of the own-gender, own-age, and own-ethnicity bias for Experiment 1 (top panel) and Experiment 2 (bottom panel). Error bars represent standard error of the mean.

Figure 3. Mean attractiveness ratings for own- and other-gender, other-age, and other-ethnicity faces, split by type of school (Experiment 1). Error bars represent standard error of the mean.

Figure 4. Mean recognition accuracy (*d'*) for own- and other-gender faces (Experiment 2), split by type of school. Error bars represent standard error of the mean.

Figure 5. Mean distinctiveness ratings for own- and other-gender, other-age, and other-ethnicity faces, split by type of school (Experiment 2). Error bars represent standard error of the mean.

Figure 6. Mean (top panel) recognition accuracy (*d'*) and (bottom panel) response times (*ms*) for upright and inverted own- and other-gender faces, split by encoding instruction (Experiment 3). Error bars represent standard error of the mean.

Figure 7. Mean (top panel) distinctiveness/attractiveness ratings and (bottom panel) response times to make judgements (*ms*) for own- and other-gender, upright and inverted faces (Experiment 3). Error bars represent standard error of the mean.

Figure 8. Mean recognition accuracy (*d'*) for own- and other-gender faces, split by participant sexuality (Experiment 4). Error bars represent standard error of the mean.

Figure 9. Mean (top panel) distinctiveness ratings and (bottom panel) response times to make distinctiveness judgements (*ms*) for own- and other-gender faces split by participant sexuality(Experiment 4). Error bars represent standard error of the mean.

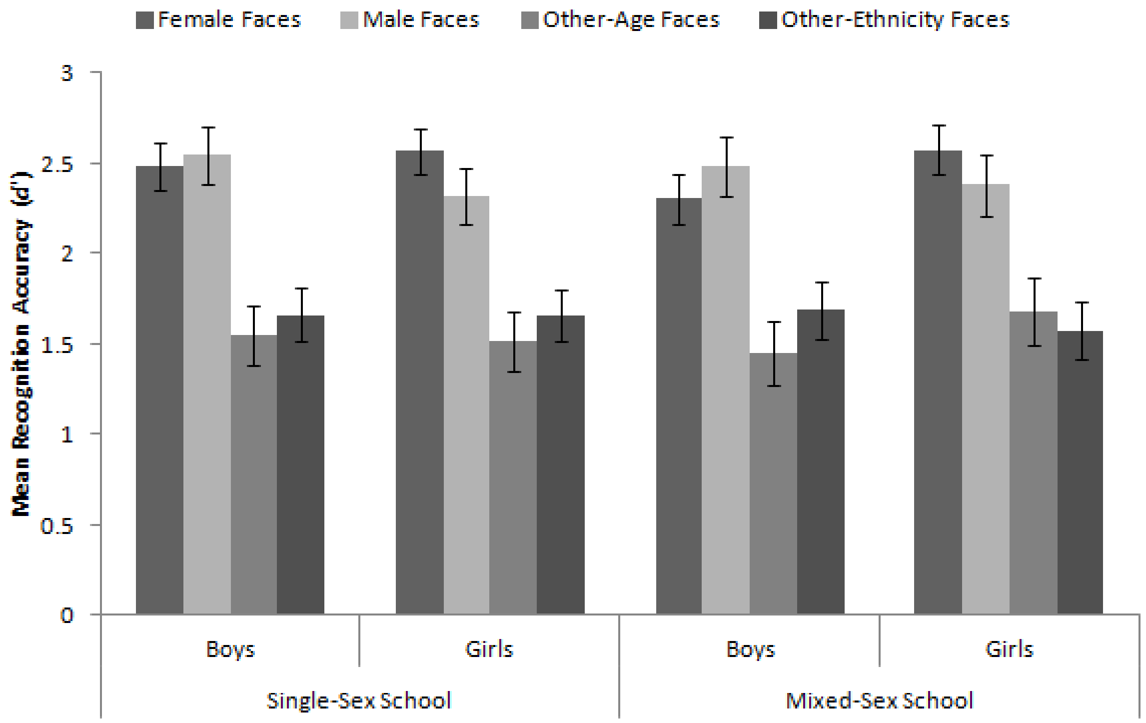

Female Faces
Male Faces
Other-Age Faces
Other-Ethnicity Faces
Mean Recognition Accuracy (d')
3
2.5
2
1.5
1
0.5
0
Boys
Girls
Boys
Girls
Single-Sex School
Mixed-Sex School

Figure 1.

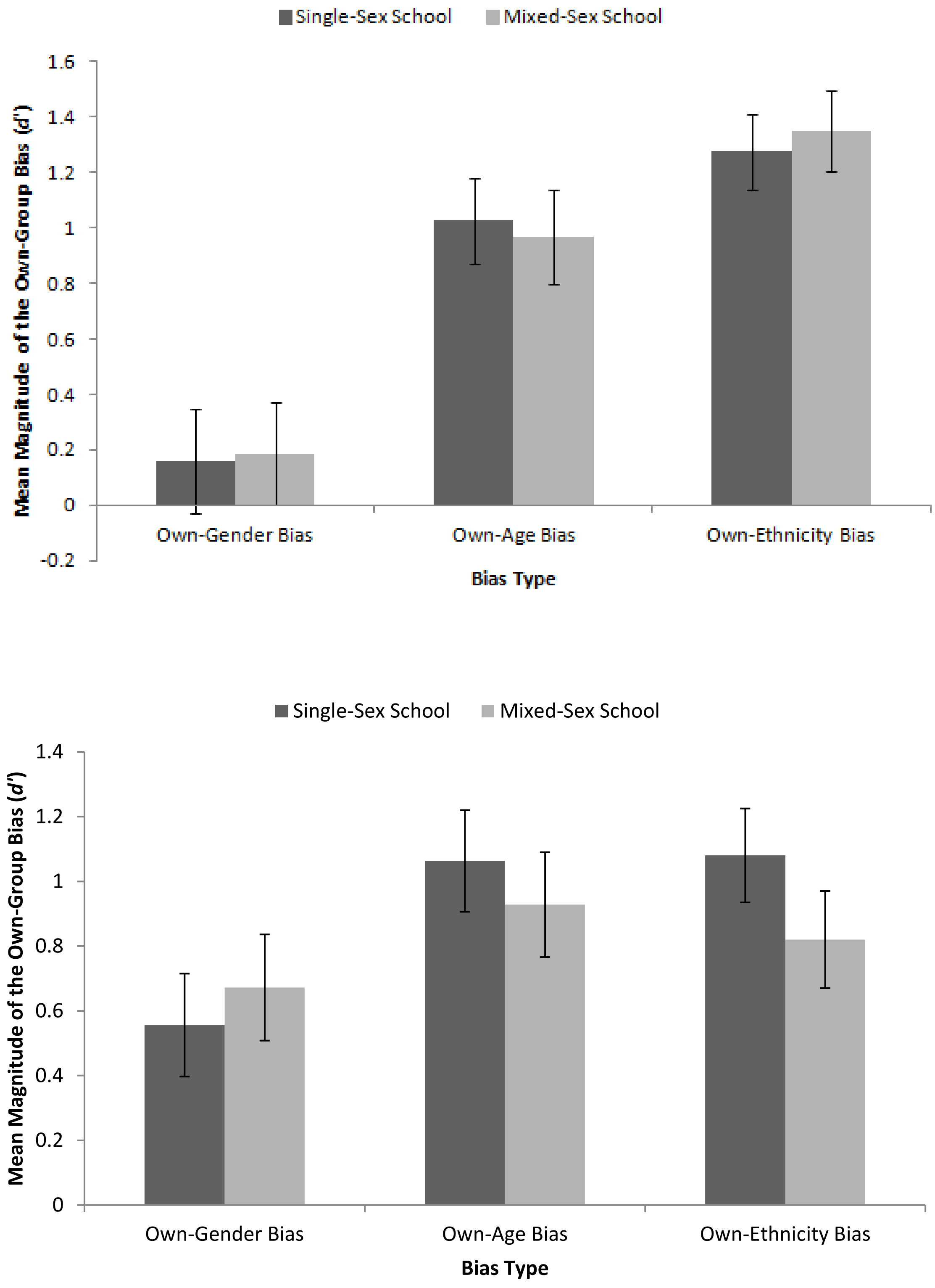

Single-Sex School
Mixed-Sex School
Mean Magnitude of the Own-Group Bias (d')
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Own-Gender Bias
Own-Age Bias
Own-Ethnicity Bias
Bias Type
Single-Sex School
Mixed-Sex School
Mean Magnitude of the Own-Group Bias (d')
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Own-Gender Bias
Own-Age Bias
Own-Ethnicity Bias
Bias Type

Figure 2.

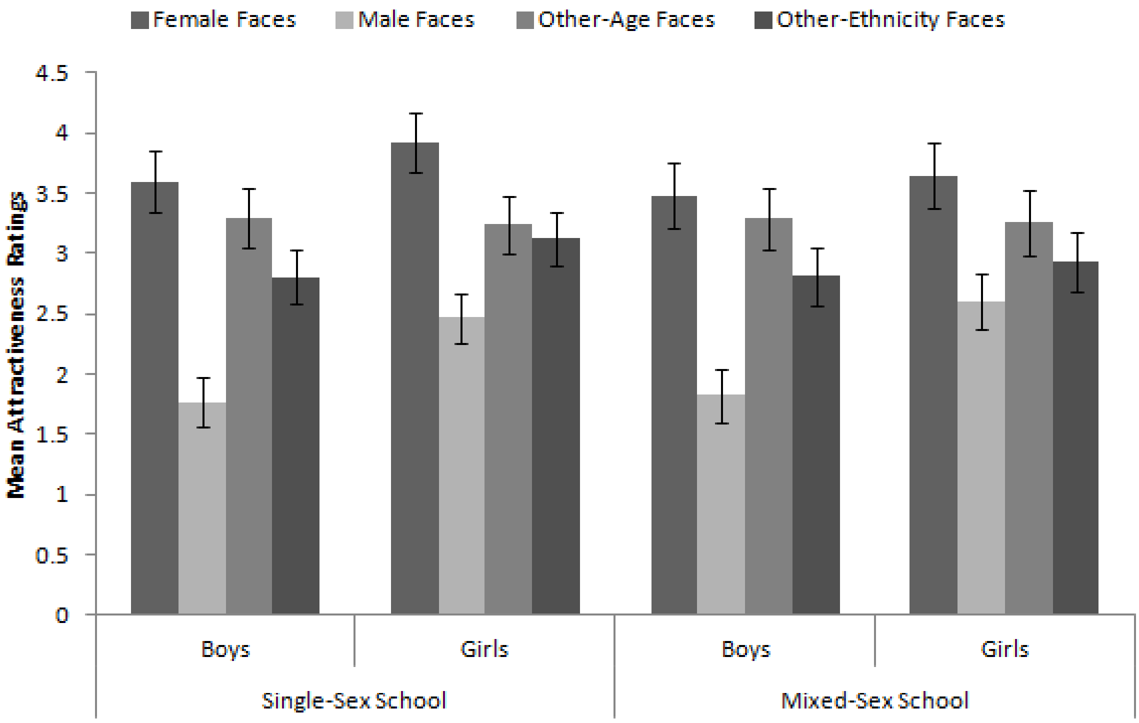

Female Faces
Male Faces
Other-Age Faces
Other-Ethnicity Faces
Mean Attractiveness Ratings
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Boys
Girls
Boys
Girls
Single-Sex School
Mixed-Sex School

Figure 3.

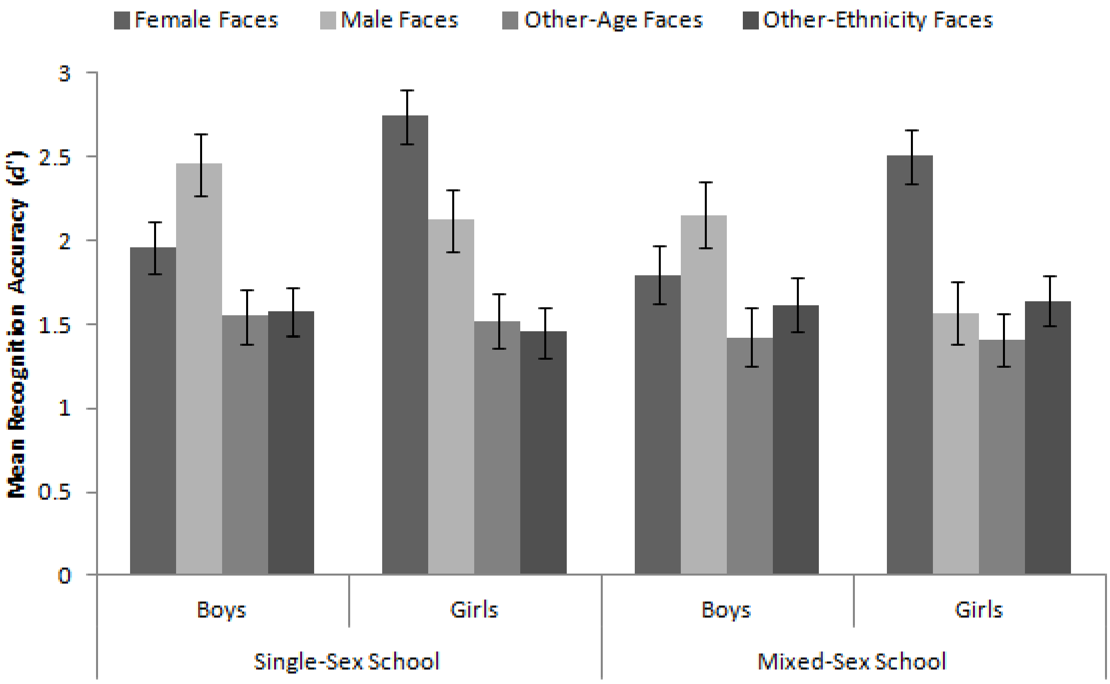

Female Faces
Male Faces
Other-Age Faces
Other-Ethnicity Faces
Mean Recognition Accuracy (d')
3
2.5
2
1.5
1
0.5
0
Boys
Girls
Boys
Girls
Single-Sex School
Mixed-Sex School

Figure 4.

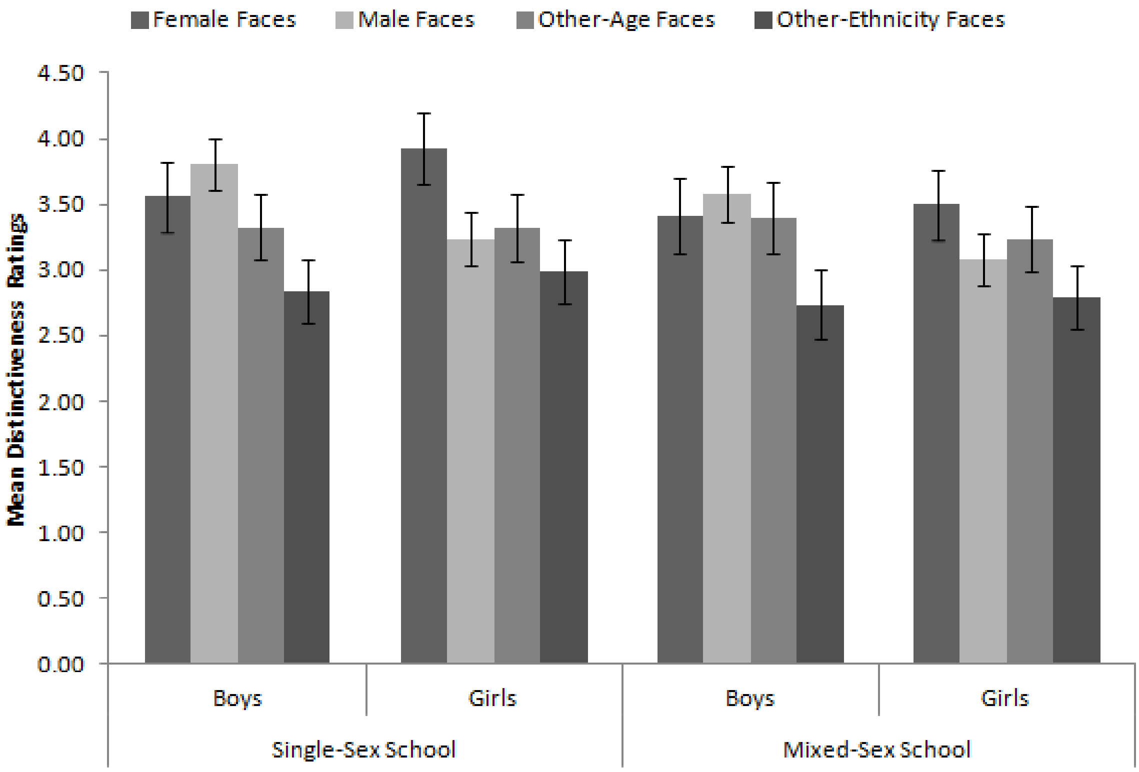

Female Faces
Male Faces
Other-Age Faces
Other-Ethnicity Faces
Mean Distinctiveness Ratings
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Boys
Girls
Boys
Girls
Single-Sex School
Mixed-Sex School

Figure 5.

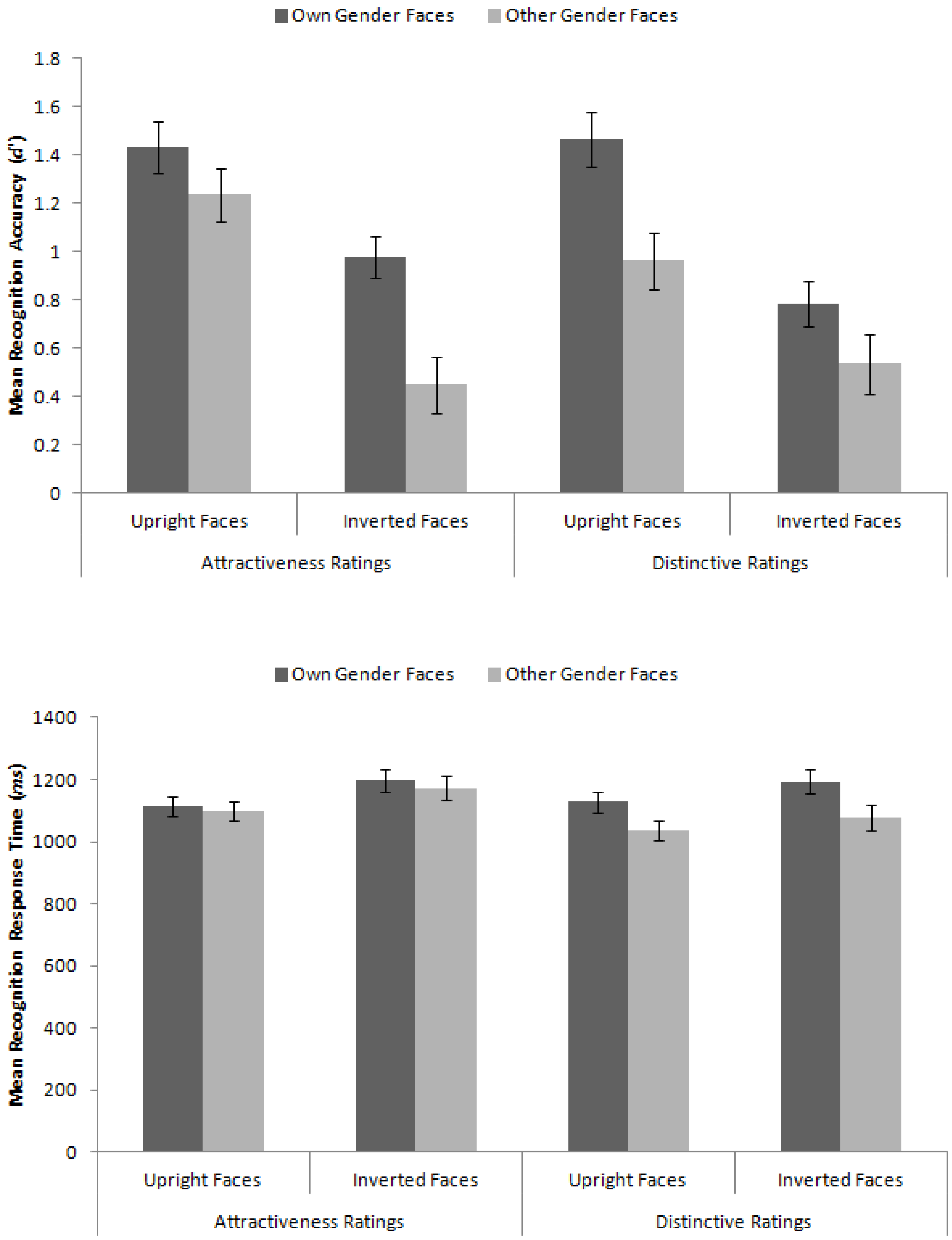
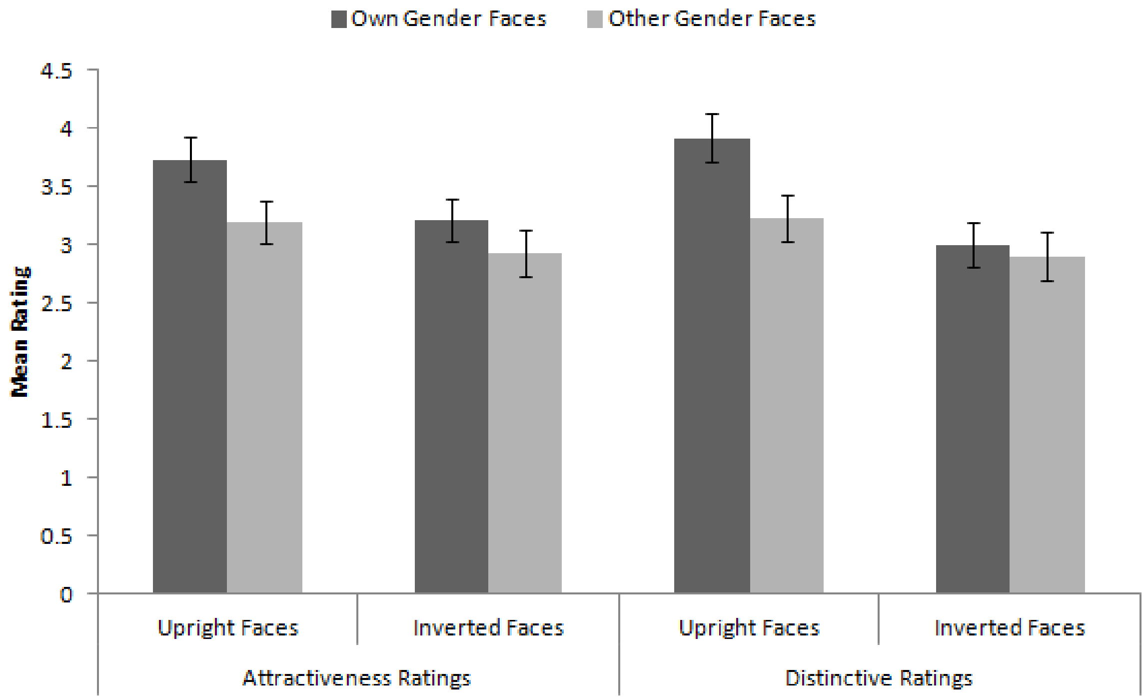

Own Gender Faces
Other Gender Faces
Mean Recognition Accuracy (d')
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Upright Faces
Inverted Faces
Upright Faces
Inverted Faces
Attractiveness Ratings
Distinctive Ratings
Own Gender Faces
Other Gender Faces
Mean Recognition Response Time (ms)
1400
1200
1000
800
600
400
200
0
Upright Faces
Inverted Faces
Upright Faces
Inverted Faces
Attractiveness Ratings
Distinctive Ratings

Figure 6.


Own Gender Faces
Other Gender Faces
Mean Rating
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Upright Faces
Inverted Faces
Upright Faces
Inverted Faces
Attractiveness Ratings
Distinctive Ratings

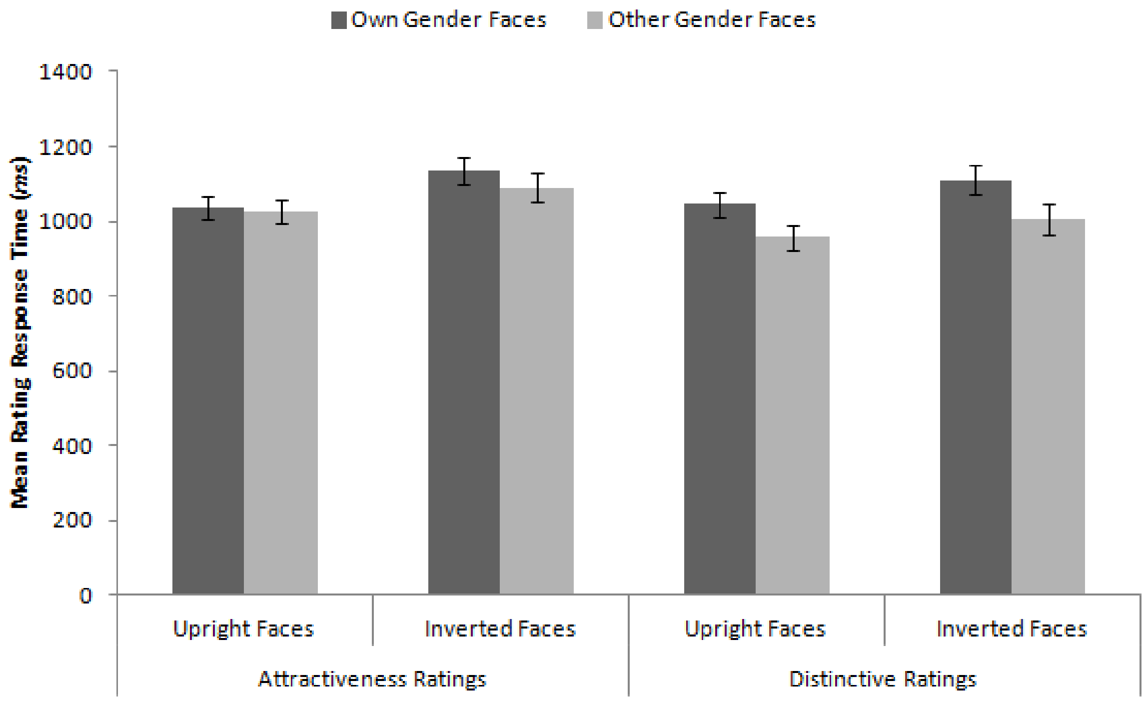

Own Gender Faces
Other Gender Faces
Mean Rating Response Time (ms)
1400
1200
1000
800
600
400
200
0
Upright Faces
Inverted Faces
Upright Faces
Inverted Faces
Attractiveness Ratings
Distinctive Ratings

Figure 7.


Own-Gender Faces
Other-Gender Faces
Mean Recognition Accuracy (d')
3
2.5
2
1.5
1
0.5
0
Gay
Straight
Participant Sexuality


Own-Gender Faces
Other-Gender Faces
Mean Recognition Response Time (ms)
1200
1000
800
600
400
200
0
Gay
Straight
Participant Sexuality

Figure 8.

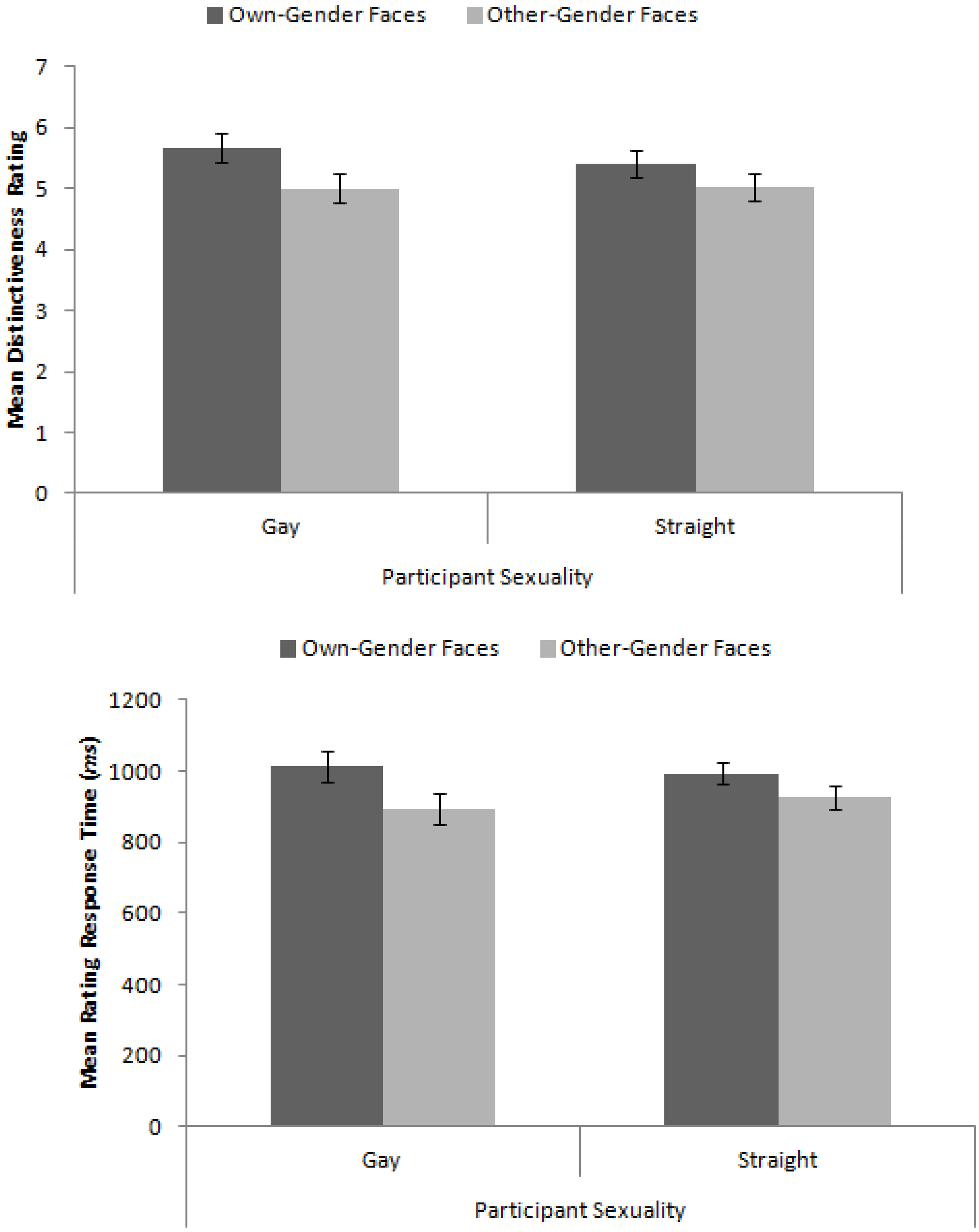

Own-Gender Faces
Other-Gender Faces
Mean Distinctiveness Rating
7
6
5
4
3
2
1
0
Gay
Straight
Participant Sexuality
Own-Gender Faces
Other-Gender Faces
Mean Rating Response Time (ms)
1200
1000
800
600
400
200
0
Gay
Straight
Participant Sexuality

Figure 9.