Cronfa - Swansea University Open Access Repository

_____

This is an author produced version of a paper published in :
Journal of Experimental Psychology - Human Perception and Performance

Cronfa URL for this paper:

_____

Paper:

Tree, J., Horry, R., Riley, H. & Wilmer, J. (in press). Are portrait artists superior face recognisers? Limited impact of adult experience on face recognition ability.. *Journal of Experimental Psychology - Human Perception and Performance*

_____

Are portrait artists superior face recognizers? Limited impact of adult experience on face recognition ability

Jeremy J. Tree[a], Ruth Horry[a], Howard Riley[b], and Jeremy B. Wilmer[c]

[a]Swansea University, UK
[b]University of Wales Trinity Saint David, UK
[c]Wellesley College, USA

Dr Jeremy J. Tree*, Department of Psychology, Vivian Tower, Park Campus, Swansea University, Swansea, SA2 8PP, UK. Email: j.tree@swansea.ac.uk

Dr Ruth Horry, Department of Psychology, Vivian Tower, Park Campus, Swansea University, Swansea, SA2 8PP, UK. Email: r.horry@swansea.ac.uk

Professor Howard Riley, Swansea College of Art, University of Wales Trinity Saint David, De La Beche Street, Swansea, SA1 3EU, UK. Email: howard.riley@uwtsd.ac.uk

Dr Jeremy B. Wilmer, Department of Psychology, Wellesley College, Wellesley, MA, USA. Email: j.wilmer@wellesley.edu

*Corresponding author

Abstract

Across two studies, we asked whether extensive experience in portrait art is associated with face recognition ability. In Study 1, 64 students completed a standardized face recognition test before and after completing a year-long art course that included substantial portraiture training. We found no evidence of an improvement in face recognition after training over and above what would be expected by practice alone. In Study 2, we investigated the possibility that more extensive experience might be needed for such advantages to emerge, by testing a cohort of expert portrait artists ($N = 28$), all of whom had many years of experience. In addition to memory for faces, we also explored memory for abstract art and for words in a paired-associate recognition test. The expert portrait artists performed similarly to a large, normative comparison sample on memory for faces and words, but showed a small advantage for abstract art. Taken together, our results converge with existing literature to suggest that there is relatively little plasticity in face recognition in adulthood, at which point our substantial everyday experience with faces may have pushed us to the limits of our capabilities.

Keywords: *art expertise, face recognition, individual differences, plasticity*

Statement of Public Significance: This study investigates the impact of expertise in portraiture on the ability to recognize faces, an important skill for navigating through a social world. Neither art students with one year of portraiture training, nor professional portrait artists, were better at recognizing faces than participants from comparison samples. By adulthood, we may have reached the limits of our capabilities, rendering face recognition ability relatively resistant to change even from intensive training.

Ideas about the plasticity of mind and brain have long driven individual and societal decisions across domains as diverse as education, health, and business. Such ideas gain ever greater currency as the global economy increasingly values knowledge and cognitive skills, as the cost of education reaches unprecedented levels, and as "brain games" become a billion dollar industry (http://sharpbrains.com/). There are substantial risks to either underestimating or overestimating the plasticity of a given cognitive capacity. On the one hand, if we underestimate its plasticity, we may bypass valuable opportunities for growth and development. On the other hand, if we overestimate its plasticity, we may waste personal or institutional resources on ineffective training interventions, while potentially discouraging the pursuit of alternatives such as creative compensatory strategies or common sense accommodations. Individuals and society thus benefit from science that accurately estimates the plasticity of cognitive capacities. In this study we focus on a key question: does extensive training, of an intensity and duration that is not possible in a laboratory setting, augment face recognition ability? We focus on artists who have received intensive training in portrait art (Study 1), or who have many years of professional experience in portraiture (Study 2). Below, we briefly review the literature on the plasticity of face recognition ability before outlining our hypotheses.

Plasticity in face recognition ability

Over recent years, individual differences in face recognition ability (FRA) have garnered considerable research interest (for a review, see Yovel, Wilmer, & Duchaine, 2014). It is clear that there is a wide distribution of face recognition skill in the general population. At one end of the spectrum, some individuals are unable to

recognize even the faces of close friends and family members, which can cause social difficulties (e.g., Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008). At the other end of the spectrum are "super-recognizers" (Russell, Duchaine, & Nakayama, 2009). These individuals claim that they can recognize faces of people they met in passing years later; they also score very highly on behavioral tests of face recognition and identity matching (e.g., Bobak, Bennetts, Parris, Jansari, & Bate, 2016; Bobak, Dowsett, & Bate, 2016; Bobak, Hancock, & Bate, 2015). The variability of FRA has created a burgeoning literature examining the correlates of this important skill (e.g., Rhodes, Jeffery, Taylor, & Ewing, 2013; Richler, Cheung, & Gauthier, 2011). A second literature has also developed around the plasticity of the human face recognition system.

Three complementary types of evidence bear importantly on the question of FRA's plasticity. First, twin studies estimate the relative contributions of genes, family environment, and non-family environment to individual differences in a population. Measured environmental influences in a twin study quantify the degree of plastic change at one or more stages of life. Second, intervention studies aim to induce change via training or other direct manipulations. Such induced changes document non-zero plasticity and may be used to identify specific causes of plasticity. Third, correlational studies of an experience too strong or prolonged to impose ethically or practically in the laboratory aim to document differences associated with that experience. Cases of no difference despite exceptional experience can provide powerful evidence *against* plasticity. We now review, in turn, how existing evidence

from each of these three types of studies relates to FRA's plasticity versus stability. We then turn to our own study of artists, which is of the third type.

The three existing twin studies of FRA, taken together, show surprisingly little evidence for environmental influence (Wilmer et al., 2010a; Shakeshaft & Plomin, 2015; Zhu et al., 2010); moreover, what environmental influence there was appeared to dissipate with age (Zhu et al., 2010). These twin results substantially constrain the overall amount of plasticity one would expect to see in FRA. That said, twin studies are relatively insensitive to plasticity that exists in relatively few individuals, or that results from relatively rare experiences (Plomin, DeFries, Knopik, & Neiderhiser, 2013). Guided by these twin studies, we therefore infer that if FRA plasticity exists, it is most likely constrained to unusual individuals or unusual experiences.

One unusual subset of individuals with regard to FRA is those whose FRA is clinically poor: those with acquired (brain damage induced) or developmental prosopagnosia (AP and DP, respectively). The plasticity of FRA in these groups is of particular interest for two reasons. First, those with FRA deficits are clearly not already at or near the ceiling of possible FRA; they could therefore conceivably have greater potential for improvement via interventions. Second, those with FRA deficits self-evidently have much potentially to gain, in terms of quality of life, from meaningful improvements in their FRA.

Efforts to train FRA in individuals with AP have yielded little overall improvement (Degutis, Chiu, Grosso, & Cohan, 2014). In some cases, these patients gained facility in recognizing particular faces, but these gains generalized poorly to novel faces (Degutis et al., 2014). Efforts to train those with DP have shown

somewhat greater promise (Degutis et al., 2014; Bate & Bennetts, 2014), though there is a need for follow-up work in this area. Of four published single-case training studies (Brunsdon, Coltheart, Nickels, & Joy, 2006; Schmalzl, Palermo, Breen, Brunsdon, & Coltheart, 2008; Dalrymple, Corrow, Yonas, & Duchaine, 2012; Degutis, Bentin, Robertson, & D'Esposito, 2007), one (Degutis et al., 2007) showed a generalized, though transient, improvement in objectively measured FRA that was reflected in neural measures. A major follow-up that applied the same training regimen to 24 new DP cases found mean improvements, relative to a waiting period control, in objective face perception tests and in subjective reports of FRA. This improvement did not generalize to face perception tests that required matching across different face views; the study also did not objectively test FRA, nor did it conduct a long-term follow-up to determine the persistence of the observed improvements (Degutis, Cohan, & Nakayama, 2014). The study's findings nevertheless suggested some degree of plasticity in face identity processing. Taken as a whole, these studies of DP suggest at least some plasticity in face processing, but an important challenge for future work is to demonstrate the repeatability of these initial findings and extend them to clarify the mechanisms and persistence of any plastic change.

Yet what about experiences that - qualitatively, quantitatively, or both - are truly exceptional? Might these experiences induce sustained, plastic changes in FRA, even if lesser experiences do not? This brings us to the third type of evidence mentioned above: correlational studies of those with exceptional experiences. Exceptional experiences can be either negative or positive. On the negative side, visual deprivation due to cataracts during infancy has been associated with reductions,

many years later, in sensitivity to face spacing (Le Grand, Mondloch, Maurer, & Brent, 2003) and in the strength of the composite face illusion (Le Grand, Mondloch, Maurer, & Brent, 2004). Though these studies did not measure FRA directly, they suggest that sustained, plastic, negative change in face processing is possible.

More relevant to our study, does FRA correlate with exceptional positive experiences? To our knowledge, two studies have asked this question. The first investigated forensic examiners with many years of on-the-job experience comparing face images for law enforcement and government agencies (White, Phillips, Hahn, Hill, & O'Toole, 2015). Though FRA was not directly assessed in this study, the forensic examiners performed unusually well in face matching; interestingly, these individuals also showed a decrease in their inversion effect, suggesting that their superior performance may have resulted from a part-based processing strategy. In another study of a different sort of expert, students at the Guangzhou Academy of Fine Arts, who had 2-16 years of experience drawing faces, were compared with age-matched controls (Zhou, Cheng, Zhang, & Wong, 2012). No difference between groups in FRA was found. A difference was found, however, in holistic processing of faces. This difference, similar to that in the forensic examiners, suggested a more part-based strategy amongst the art students.

Contemporary theories of face processing posit that face processing involves parallel encoding of two broad classes of information: part-based information (the shape and size of individual features) and configural information (spatial relationships between features and the holistic impression of the face) (Maurer, Le Grand, & Mondloch, 2002). Given that configural processing has long been considered a

hallmark of human face processing (Maurer, LeGrande & Mondloch, 2002), and given evidence that face processing deficits often coincide with reduced configural processing (e.g. Degutis et al, 2012), it may seem surprising that two groups with intensive experience with faces – forensic examiners and art students (White et al, 2015; Zhou et al, 2012) – employ a relatively part-based approach to face processing. Yet the apparent contradiction in the latter results diminishes if one considers that both configural and part-based processing contribute to accurate face recognition, and they do not necessarily trade off against one another in a zero-sum manner (Hayward, Crookes, & Rhodes, 2013; Mondloch et al., 2010). First, it remains possible that a capacity to flexibly adopt a part-based strategy when it is called for – such as in the composite task completed by Zhou et al's participants – could be beneficial for FRA. Second, a common assumption that configural processing contributes strongly to individual differences in FRA over the non-clinical range has failed to receive empirical support; individual differences studies have found relatively little relationship between configural processing and FRA (Yovel, Wilmer & Duchaine, 2014). In sum, it remains entirely plausible that improved part-based processing, with or without changes in configural processing, could bolster FRA.

Though a valuable initial investigation of FRA in artists, Zhou and colleagues' (2012) study had a number of limitations that we seek to tackle here. First, Zhou et al. used a novel, unvalidated measure of FRA, and provided no evidence for the sensitivity (reliability) of that measure. Our study uses the most well-validated measure of FRA, the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). The CFMT is widely used to assess FRA (Wilmer et al., 2010a; Wilmer et al.

2012) and to diagnose face recognition impairments in adults (e.g. Avidan, Tanzer, & Berhmann, 2011; Bate et al., 2014; Bowles et al., 2009; Rivolta, Schmalzl, Coltheart, & Palermo, 2010). The CFMT demonstrates high internal reliability (Duchaine & Nakayama, 2006), high test-retest reliability (Wilmer et al., 2010a), convergent validity with measures of face perception (Bowles et al, 2009) and famous face recall (Wilmer et al, 2010a; Wilmer et al, 2012), and divergent validity with object recognition (e.g., Dennett et al., 2012). The CFMT demonstrates desirable psychometric properties, such as the ability to precisely estimate FRA across a broad range of ability (Wilmer et al, 2012; Choo et al., 2015). Scores on the CFMT also correlate highly with self-reported face recognition difficulties (Shah, Gaule, Sowden, Bird, & Cook, 2015; Wilmer et al, 2010a) and with performance on face-matching tests (Palermo et al., in press). Crucially, large, normative datasets including test-retest performance were available for comparison with our sample of art students, which allowed us to separate any gains in FRA from task-specific practice effects.

Second, Zhou et al. (2012) tested art students only once; their study therefore cannot rule out the possibility that art students start out worse at FRA and attain FRA normalcy via their art training. Indeed, the famous portrait artist Chuck Close has a severe deficit in FRA, falsifying the notion that poorer FRA could not coexist with, or even potentially fuel, world-class face-related art achievement. The first part of our study thus tests a cohort of art students before and after their first year of intensive, university-level art training.

Third, and finally, it could be that even a year of university-level training is not enough to impact FRA. Rather, perhaps the more intensive, longer-term demands

of achieving a productive career as a professional portrait artist could, generally-speaking (Chuck Close's persisting FRA deficits notwithstanding), cause positive plastic changes in one's FRA. The second part of our study tested this hypothesis by assessing FRA in a cohort of 28 professional portrait artists.

In summary, our study aimed to explore the impact of unusual, prolonged, and intensive experience at scrutinizing faces on FRA. Though laboratory based interventions have not been particularly successful in augmenting face processing (e.g., Degutis et al., 2014; Dolzycka, Herzmann, Sommer, & Wilhelm, 2014), and though twin studies have suggested little environmental influence on FRA (Wilmer et al., 2010a; Shakeshaft & Plomin, 2015; Zhu et al., 2010), it is nonetheless possible that truly exceptional experience with the processing of faces will be associated with superior memory for face stimuli.

Study 1

In Study 1, art students completed the CFMT prior to and after completing a year-long art course, with substantial portraiture training. In addition, the participants completed the Cambridge Face Perception Test (Duchaine, Yovel, & Nakayama, 2007). The CFPT was completed with both upright and inverted faces.

Method

Participants and Design.

Sixty-four students enrolled on a university-level foundation art course took part. This foundation course included specific modules on portrait art, and included substantial practical components requiring drawing and painting faces. The art student cohort consisted of 21 males and 43 females (mean age = 19.35, SD = 2.85). The

comparison sample for the upright CFMT data consisted of 389 participants from Wilmer et al. (2010a). Our sample sizes were dictated by the size of the cohort enrolled on the course and by the size of the extant test-retest dataset obtained by Wilmer et al. A post-hoc power analysis using G*Power 3.0 (Faul, Erdfelder, Buchner, & Lang, 2007) indicated that our sample sizes produced 80% power for detecting group differences with an effect size of $f = 0.13$ (which is conventionally considered a small effect size; Faul et al, 2007).

The study followed a repeated measures design, with time (pre-training, post-training) as the independent variable. The dependent variables were percent correct on the CFMT (Duchaine & Nakayama, 2006) and CFPT (Duchaine et al., 2007).

Materials

Cambridge Face Memory Test. In the CFMT, participants learned six target faces and then completed 72 three-alternative forced choice trials in which they attempted to identify the targets. The CFMT had three phases of increasing difficulty. In the Learn phase, test images were identical to study images. In the Novel phase, the test images featured novel viewpoints and/or lighting. In the Noise phase, visual noise was added to the test images (for full procedural details, see Duchaine & Nakayama, 2006). The CFMT has demonstrated high reliability (Bowles et al., 2009; Wilmer et al., 2010b) and both convergent and divergent validity (Bowles et al., 2009; Dennett et al., 2012; Wilmer et al., 2010a; Wilmer et al., 2010b). Critically for our purposes, the fourth author had obtained a large dataset on test-retest performance (Wilmer et al., 2010a) which allowed us to account for any practice effects on our central measure.

Cambridge Face Perception Test. In the CFPT (Duchaine et al., 2007), participants were presented with a target face presented in three-quarter view, alongside six test faces presented in a frontal view. Each of the test faces was a morph between the target face and another face. To vary the similarity to the target, the six test faces fell at different points on the morphing continuum (88%, 76%, 64%, 52%, 40%, 28%). The participants' task was to rearrange the six targets in order of similarity to the target. The task was completed for eight different target identities, and participants had up to one minute to rearrange the six faces for each target. Participants completed the task twice: once with upright faces and once with inverted faces.

Procedure

Two testing sessions were administered. Session 1 was in week 1 of the arts foundation course. In this session, the participants completed a short questionnaire that collected demographic information (including, age, gender, handedness, evidence of neurological injury, experience of face processing impairments) and then undertook the CFMT and upright and inverted versions of the CFPT on site at the university. In Session 2 (approximately eight months later), the participants completed the tests again.

Results

Our primary question concerned whether participants would show improvement in their CFMT scores after receiving intensive art training over and above a general practice effect seen in a non-artistic control sample. We were able to compare our sample of art students with a normative sample of 389 participants from

Wilmer et al. (2010a), who completed the CFMT on two separate occasions, approximately six months apart. The participants in this normative sample did not undertake any specialist art training in the intervening period, and thus served as a baseline for practice effects. We conducted a 2 (Time: time 1, time 2) × 2 (sample: art students, comparison group) mixed ANOVA, with repeated measures on the first factor. As measures of effect size, we report partial eta-square for interaction terms, Cohen's *d* for between-subjects comparisons, and standardized mean change (*d*) for within-subjects comparisons. For ease of interpretation, the CFMT was scored as percent correct.
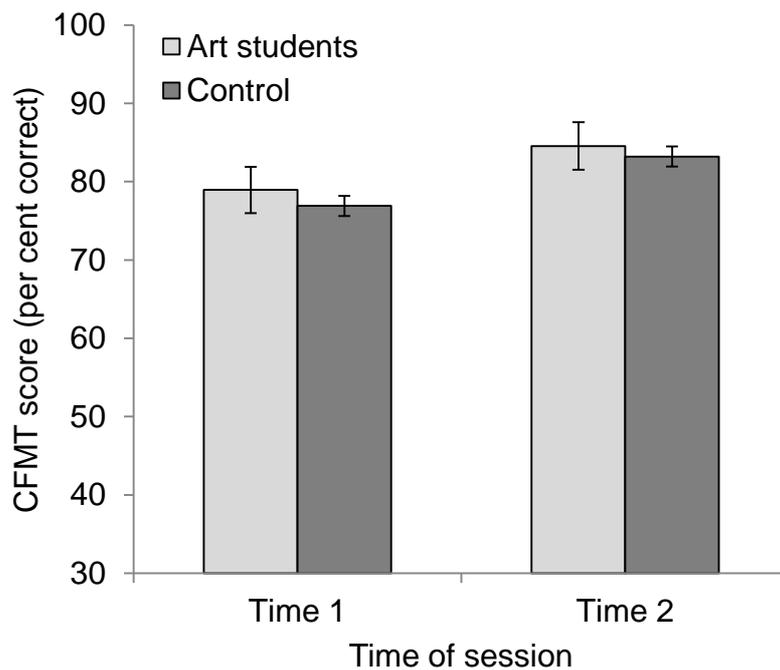

Figure 1. CFMT performance (per cent correct) at time 1 and time 2 for the art student and control samples. Error bars represent 95% confidence intervals.

Figure 1 shows mean CFMT performance (scored as per cent correct) for the artist and control samples at both time points. The main effect of group was not statistically significant, $F(1, 451) = 1.13$, $p = .29$, $d = 0.14$, 95% CI [-0.12, 0.41]; the mean score for the art students ($M = 81.7\%$, $SD = 11.3\%$) was very similar to the mean score of the comparison group ($M = 80.1\%$, $SD = 11.8\%$). The main effect of time was statistically significant, $F(1, 451) = 80.92$, $p < .001$, $d = 0.63$, 95% CI = [0.53, 0.73]; overall, accuracy increased from 77.2% ($SD = 12.7\%$) at time 1 to 83.4% ($SD = 12.8\%$) at time 2. However, the crucial interaction term was not statistically significant, $F(1, 451) = 0.27$, $p = .61$, $\eta_p^2 < .001$, 90% CI [.00, .01], which suggests that the increase in scores observed for the art students ($d = 0.66$, 95% CI [0.39, 0.93]) was entirely consistent with a practice effect ($d = 0.63$, 95% CI [0.52, 0.73] in the comparison sample). In both samples, CFMT accuracy was highly correlated from time 1 to time 2: art students, $r(62) = .76$, $p < .001$; comparison group: $r(387) = .70$, $p < .001$.

One potential problem with the preceding analysis is that the effectiveness of artistic training may have varied by baseline performance: that is, art students with poorer baseline face recognition ability may have benefited more from portraiture training than art students with higher baseline abilities. To control for the effects of baseline performance, we conducted an ANCOVA with time 2 accuracy as the dependent variable, group as the independent variable, and time 1 accuracy as a covariate. Group was not significantly associated with time 2 accuracy when controlling for time 1 accuracy, $F(1, 450) = 0.01$, $p = .95$, $\eta_p^2 < .001$, 90% CI [.00, .001]. These results converge with the previous analysis in suggesting that the

improvement in scores in the art students was entirely consistent with a task-specific practice effect.

Next, we examined our secondary measures: the upright and inverted versions of the CFPT. We had no large, normative samples for comparison, and thus, these analyses focus on change within the artist sample from time 1 to time 2. The CFPT is scored as the number of errors, and thus, lower scores indicate better performance. A 2 (orientation) × 2 (time) repeated measures ANOVA revealed a large effect of orientation, with better performance on the upright ($M = 35.72$, $SD = 10.01$) than inverted test ($M = 63.02$, $SD = 9.65$), $F(1, 63) = 440.17$, $p < .001$, $d = 2.60$, 95% CI [2.09, 3.12]. Performance did not significantly change from time 1 ($M = 48.58$, $SD = 10.55$) to time 2 ($M = 50.16$, $SD = 10.51$), $F(1, 63) = 0.97$, $p = .33$, $d = 0.12$, 95% CI [-0.12, 0.37], and the magnitude of the inversion effect did not significantly change over time, $F(1, 63) = 0.52$, $p = .47$, $\eta_p^2 = .008$, 90% CI [.00, .08]. Though we had no comparison sample, the lack of improvement in the CFPT suggests that portraiture training did not improve face perception ability. Furthermore, the lack of improvement in the inverted CFPT suggests that participants did not develop more featural processing strategies for faces over the course of their training.

Discussion

Experiment 1 examined whether receiving formal training in art, with a particular focus on portrait art, would enhance FRA. Though CFMT performance improved from time 1 to time 2, the rate of improvement was very similar for the art students as for a normative comparison group. The fact that we saw improvement in the comparison group rules out the possibility that face recognition *ability* had

improved between time 1 and time 2. Participants in the comparison group were simply going about their everyday lives in the interim period, and thus there would be no reason to expect any change in their FRA. Rather, the improvement is an artifact of completing the task twice, such that task-specific gains are made in recognizing the particular exemplar faces used in the CFMT. We should note, however, that there was a strong correlation between time 1 and time 2 scores, which suggests that, even on a second test, the CFMT is able to reliably measure FRA.

Face perception ability did not significantly improve from pre- to post-training, as measured by the upright version of the CFPT. Furthermore, there was no significant change in CFPT performance with inverted faces, which we would have expected if participants learnt to process faces in a more featural manner over the course of their studies (as suggested by the results of Zhou et al., 2012). We were unable to assess whether art students were generally more or less skilled at face perception tasks than the general population due to the lack of a large, normative sample. Previous studies that have used the CFPT have used quite small control samples, and as a consequence, there is substantial variability in control means across studies (e.g., Bowles et al., 2009; Duchaine et al., 2007).

Taken together with the results of Zhou et al. (2012), these data suggest that artistic training does not improve general FRA. However, it is possible that such benefits take many years to manifest themselves. Thus, one potential criticism of Experiment 1, and of Zhou et al., is that the samples were still artists-in-training, who lacked the sheer weight of experience required to boost face recognition abilities. For

this reason, our second study focused on a cohort of professional portrait artists with many years of experience.

Another limitation of Study 1 us that we examined the impact of art experience on a single category of visual stimuli – faces. In Study 2, we examine memory for faces and abstract art, along with verbal memory.

Study 2

Method

Participants and Design.

This study sought to determine whether highly experienced portrait artists showed any evidence of superior face recognition abilities. A cohort of 28 professional portrait artists (6 Male, 22 Female; mean age= 41.8, $SD = 15.6$) were recruited to take part in this study. All participants had considerable experience with portrait art, and worked in a professional capacity with a focus on this artistic discipline as teachers and/or in a commissioned capacity. Participants were recruited via a network of contacts available to the third author. Comparison data were taken from a large normative sample ($N = 1471$) from Wilmer et al. (2012). The mean age of the comparison sample was 23.6 ($SD = 10.1$); 67.64% of the comparison sample were female.

As in Study 1, our sample size was determined largely by the pragmatics of recruiting this specialist population. Using G*Power (Faul et al., 2007), we determined that our sample size produced 80% power to detect an effect size of $d =$

0.47 (d=0.50 is conventionally considered a medium effect size) at a p=0.05 (one-tailed) level (Faul et al., 2007).

The participants were tested on three standardized recognition tests: the CFMT, the Abstract Art Memory Test (AAMT; Wilmer et al., 2012), and the Verbal Paired-associates Memory Test (Wilmer et al., 2012). Scores were compared to a large, normative comparison sample (Wilmer et al., 2012).

Materials

In the VPMT, participants studied 25 word pairs (consisting of abstract nouns) and then completed a four-alternative recognition task. In each trial, the participants were presented with one word from each pair and were asked to identify the paired associate. The VPMT shows a high Cronbach's alpha reliability of 0.81 (Wilmer et al, 2012). The VPMT's convergent validity as a measure of verbal memory is supported by a robust 0.48 correlation with the Code-Learning Memory Test (CLMT), a test of verbal memory that differs markedly from the VPMT in its basic task structure (Wilmer et al, 2012). The VPMT's divergent validity as a measure of verbal memory is supported by its lower correlations with both CFMT (r=0.18) and AAMT (r=0.25) (Wilmer et al, 2012).

In the AAMT, participants studied 50 images of abstract art, and then completed a three-alternative forced choice recognition task. The AAMT shows a high Cronbach's alpha reliability of 0.80 (Wilmer et al, 2012). The AAMT's convergent validity as a measure of general visual memory is supported by a robust 0.68 correlation with the Object and Scene Memory Test (OSMT), a test of visual memory that differs markedly from the AAMT in the classes of stimuli used (Wilmer

et al, 2012). The AAMT's divergent validity as a measure of visual memory is supported by its lower correlations with both CFMT (0.26) and AAMT (r=0.25) (Wilmer et al, 2012). For further details of the VPMT and AAMT, see Wilmer et al. (2012).

Procedure

All tests were administered online. The expert artist participants were each emailed links to the three tests (the CFMT, AAMT, and VPMT) so that they could complete them in their own time. The order in which the three links were listed in the email was counterbalanced across participants, and the participants were asked to complete the tests in the order specified in the email. The participants also provided information about their portrait art experience in order to confirm their high level of expertise, in addition to standard demographic information.

Results

We tested a sample of expert portrait artists on three recognition tests: faces (using the CFMT), abstract art (using the AAMT), and words (using the VPMT). We compared their performance on each of these tasks to a large, normative sample (N = 1471) from Wilmer et al. (2012). Our artist sample was significantly older ($M =$ 41.79, $SD = 15.60$) than the normative sample ($M = 23.58$, $SD = 10.13$), $t(1497) =$ 9.31, $p < .001$, $d = 1.78$, 95% CI [1.40, 2.16], and a higher proportion of the artist sample was female (78.57%) than the normative sample (67.64%). We tested for age and sex effects within the control sample. Pearson's correlations between age and z scores for the control sample revealed small, positive relationships for each of the three tests: VPMT, $r = .048$; AAMT, $r = .056$; CFMT, $r = .074$. Cohen's d for sex

differences were calculated; women achieved higher z scores than men on every test, though the differences were very small: VPMT, $d = 0.14$, 95% CI [0.03, 0.25]; AAMT, $d = 0.12$, 95% CI [0.01, 0.23]; CFMT, $d = 0.15$, 95% CI [0.04, 0.26].

To control for age and sex effects, we converted the raw scores for each participant into z scores, controlling for age and sex.[1] Figure 2 shows the z scores of each artist in the sample. Because the distributions of z scores in the artist sample were not normal, we use non-parametric tests for all inferential comparisons. As measures of central tendency and dispersion, we report the median (Mdn) and the interquartile range (IQR). As a measure of effect size, we report r (Fritz, Morris, & Richler, 2012). Individual sex- and age-controlled z scores on each test are plotted in Figure 2.

As a test of face recognition ability, we compared the CFMT z scores of the artist sample ($Mdn = 0.32$, $IQR = 1.16$) with the normative sample ($Mdn = 0.04$, $IQR = 1.51$). The two groups did not significantly differ, $U = 17806.5$, $U_{CRIT} = 16146.32$, $p = .22$, $r = .03$.

Next, we compared performance on a test of memory for abstract art. It is important to note that our artist sample all considered themselves primarily as portrait artists, rather than abstract artists, and that none of them had any prior familiarity with any of the stimulus items. In contrast to the CFMT, the two groups were significantly different, though the effect size was very small, $U = 15942.5$, $U_{CRIT} = 16146.35$, $p = .04$, $r = .05$. The median z score in the artist group was 0.47 ($IQR = 1.31$) and the median z score in the comparison group was 0.03 ($IQR = 1.40$).
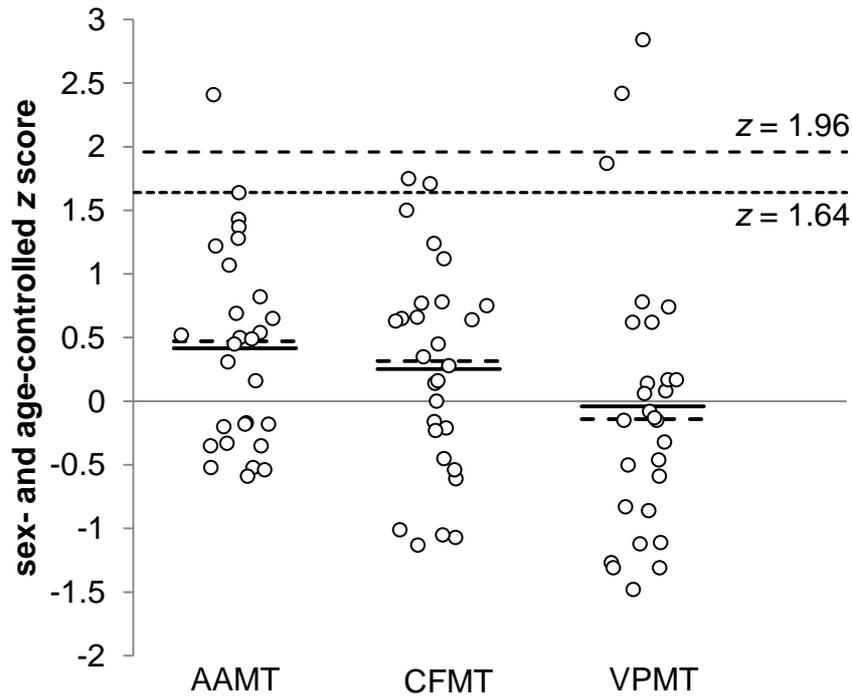
Figure 2. Sex- and age-controlled *z* scores for each expert artist participant on the Abstract Art Memory Test (AAMT), the Cambridge Face Memory Test (CFMT), and the Verbal Paired-associates Memory Test (VPMT).

Though the difference between the artist sample and the comparison sample was statistically significant for abstract art, but not for faces, it is apparent from Figure 2 that there is a great deal of overlap in the *z* scores of artists between the AAMT and CFMT. We explored whether the artist sample performed significantly better on the AAMT than the CFMT using a Wilcoxon signed-ranks test. The difference was not statistically significant, $T = 177$, $T_{CRIT} = 116.43$, $p = .56$, $r = .08$.

To ensure that our artist sample did not significantly differ from the normative sample in general memory ability, we compared VPMT accuracy of the artist sample ($Mdn = -0.14$, $IQR = 1.36$) and comparison sample ($Mdn = -0.17$, $IQR = 1.44$). The

two groups did not significantly differ, $U = 19391.5$, $U_{\text{CRIT}} = 15481.9$, $p = .84$, $r = .005$.

Finally, we compared each individual artist in our sample to the normative sample as if in a single case-study design (for a similar approach with "super-recognizers", see Bobak, Bennetts, et al., 2016; Bobak, Dowsett, et al., 2016; Bobak, et al., 2015). Given the large normative sample, it is appropriate to treat the normative data as a population parameter (Crawford & Howell, 1998). Thus, we used the z scores to estimate the percentage of the population that would likely fall beneath the participant's score. For example, a z score of 1.64 indicates that 95% of the population falls below the participant's score; a z score of 1.96 indicates that 97.5% of the population falls below the participant's score. For each test, we also estimated the probability of at least x number of participants exceeding the cutoff, even if we had sampled the participants randomly from a population with a mean z score of 0 (i.e., even if the artist population had the same parameters as the comparison population).

For the CFMT, none of the artists' z scores exceeded 1.96. Two of the artists (participants 24 and 25) surpassed a more lax criterion of $z \geq + 1.64$. Approximately 96.0% of the population would be expected to fall below participant 24's score, and approximately 95.6% of the population would be expected to fall below participant 25's score. The probability of at least two participants exceeding the 1.64 criterion is approximately 41.2%.

For the AAMT, one participant 9's z score exceeded 1.96 (99.2% of the population would be expected to fall below participant 9's score), and participant 1's z score equaled 1.64 (95.0% of the population would be expected to fall below

participant 1's score). Once again, the probability of at least two participants exceeding $z = 1.64$ is 41.2%. The probability of at least one participant exceeding $z = 1.96$ is 50.8%.

Finally, for the VPMT, two participants 4 and 12 had z scores greater than 1.96, and participant 1's z score greater exceeded 1.64. The percentage of the population that would be expected to fall below these scores were 99.8% for participant 12, 99.2% for participant 4, and 95.0% for participant 1. The probability of at least three participants exceeding $z = 1.64$ is 16.3%; the probability of at least two participants exceeding $z = 1.96$ is 15.4%.

Taken together, the single-case comparisons here provide no compelling evidence of "super" recognition abilities amongst our artists, for any category of stimuli.

Discussion

The current experiment explored whether face recognition memory performance for a cohort of professional portrait artists is superior to that of the general population. In addition we sought to determine whether performance for this cohort differs from that of the general population with a different category of expertise-relevant stimuli, abstract art. We also tested memory for words to rule out the possibility that artists have better memory abilities overall than the general population. Our findings suggested that (a) the artists performed in line with the general population with both faces and words and (b) on average, the artist sample performed better than the normative sample in memory for abstract art. These findings indicate that years of experience in professional portrait art does not lead to

robust, generalizable benefits for FRA, though experience may boost recognition memory for other types of expertise-relevant visual stimuli.

One potential question about this study is whether the statistical power associated with its *n* of 28 professional portrait artists is sufficient. We think it is instructive to consider this question in light of the literature on case-studies of neuropsychological patients. In studies, it is not uncommon for a single person to be studied when a reasonably large effect size is anticipated.

Granted, years of intensive training, followed by professional work, is no small experimental intervention. There is an analogy to be drawn here with studies of developmental prosopagnosia or super face recognition, where case studies of single individuals are often considered quite informative. Nevertheless, one might wonder at the degree to which our study could miss a subtle effect. Consequently, we compared the *z* score of each participant to the population as would be typical in single case studies in the neuropsychological literature (Crawford & Howell, 1998). Across all three tests, the majority of artists were in the typical range of the population (between -1.64 and +1.64), with only two to three participants in each test exceeding +1.64. Thus, the proportion of participants who significantly deviated from the population was in line with what would be expected from sampling error alone, if the artists had been randomly sampled from the general population. These results: i) add additional weight to the argument that portrait artists are not superior face recognizers; and ii) suggest that the average improvement in abstract art memory is a modest one, with most artists still within the typical range.

General Discussion

There exists substantial variability in face recognition ability (FRA) across the general population (e.g., Duchaine & Nakayama, 2006; Russell et al., 2009; Wilmer et al., 2010). A growing body of literature has sought to understand where this variability comes from and ask what, if any, practical opportunities may exist for improving FRA. Existing evidence from twin and intervention studies is consistent with the theory that for most of us, everyday experience with faces is so rich and varied that, for commonly experienced face types, we each actualize our natural FRA potential, and little additional upward plasticity exists. Existing evidence, however, has spoken relatively little to the potential impact of truly exceptional experiences on FRA. Here, we studied one of the most intensive, prolonged, and focused experiences one can have with faces: training and professional work in portraiture. We examined face recognition ability in two groups of portrait artists: those undergoing art training, and those with multiple years of professional experience. If such exceptional experience is capable of enhancing face recognition abilities, then we would expect these participants to perform unusually well on a sensitive, well-validated, normed FRA test (the CFMT; Duchaine & Nakayama, 2006).

Our results clearly failed to support the prediction of unusually high FRA performance in either group of portrait artists. In Study 1, after eight months of extensive art training, including a substantial component of portrait training, art students were no better than a normative comparison group at recognizing faces. A strength of our Study 1 was that we collected baseline data such that we could

compare performance pre- and post-training. Though we saw a small improvement in CFMT scores from time 1 to time 2, the same improvement was seen in the normative sample (Wilmer et al., 2010); thus, the improvement was no greater than would be expected from a simple practice effect. The lack of baseline differences between the art students and normative controls on the CFMT suggested also that people who are motivated to enroll in art courses do not differ in FRA from the norm. This lack of baseline difference thus argues against what we call the "Chuck Close" hypothesis (named after the famous portrait artist who is severely impaired at FRA): the idea that a fascination with recording faces on canvas may often result from a difficulty with recording faces in one's own mind.

We also failed to find any impact of art training on face perception as measured by the CFPT (Duchaine et al., 2007). Furthermore, the art students demonstrated a large inversion effect at both time points. These findings stand in contrast to those of White et al. (2015), who found that expert forensic examiners were less disrupted by inversion than control participants, and Zhou et al. (2012), who found evidence of increased reliance on featural encoding (via a smaller composite effect) among art students. It is unlikely that we failed to detect such an effect due to inadequate statistical power as the effect size of the interaction term for inversion by time was close to zero. Rather, we must conclude that a year of portraiture training is insufficient to improve face processing skills, or to mitigate the impact of inversion on face processing. However, lacking a comparison sample, we were unable to determine whether our art students performed similarly on the CFPT to the general

population. It is possible that they showed smaller inversion effects than are typical at both time points. Further data would be required to answer that question.

What if the experiences of art students, exceptional as they are, were *still* not enough to impact FRA? To address this question, in Study 2, we recruited a substantial cohort of 28 professional portrait artists. Despite their exceptional expertise and experience, we were unable to detect a significant advantage in FRA relative to a large, web-recruited normative sample. In contrast to their unexceptional FRA, however, these same artists outperformed the normative sample at recognizing abstract art, a result that is important for at least three reasons. First, it suggests that the (uniformly high) reliability of our tests was sufficient to detect a small to medium effect in this sample. Second, it argues against explanations based on exceptional memory (or attentiveness) in the web-recruited normative sample relative to our artist sample. Third, it suggests that art experience may be associated with superior performance in at least one aspect of visual memory. It is important to note, however, that the effect size associated with this group difference in abstract art recognition was small. Furthermore, within the artist sample, standardized z scores for abstract art recognition and face recognition were not significantly different. It would, therefore, be premature to conclude that artists show superior art recognition, or that their memory art outstrips their memory for other important classes of visual stimuli, such as faces. Rather, it would be valuable to replicate this finding in a larger sample of portrait artists to gain a more precise estimate of effect size.

In addition to their superior abstract art memory, the artists recognized words normally, ruling out the possibility that artists exhibit a global memory deficit that is

specifically counteracted in the visual realm by their art experience. In sum, the findings of Experiments 1 and 2 converge to demonstrate a remarkably ordinary level of FRA performance in two groups with exceptional face-related training, experience, and expertise.

There are some important limitations of our research. First, these studies were necessarily correlational, which limits the causal inferences that may be drawn. A more extended longitudinal study could provide additional insight into how intensive, unusual experiences impact FRA, and would potentially allow the arrow of cause and effect (where any benefits are found) to be established. Second, though we used the most well-validated and reliable measure of FRA currently available (the CFMT; Duchaine & Nakayama, 2006), a larger battery of face recognition tests could more conclusively rule out a difference in any aspect of FRA between portrait experts and comparison samples. Indeed, research with both super-recognizers and prosopagnosics has revealed substantial heterogeneity in profiles of performance across tasks that tap into different aspects of face processing and recognition (e.g., Bobak, Bennetts, et al., 2016; Le Grand et al., 2006). It is possible that our portrait artists would have out-performed control participants on tests that made different demands on the face processing system. Zhou et al (2012), for example, found that art students showed smaller holistic processing effects than control participants in a face composite task. It is possible, therefore, that artists would excel in tasks in which featural processing strategies more strongly benefit performance. Finally, it is possible that training and experience in portraiture – though a highly intensive face processing experience – is not the ideal training for enhancing FRA specifically.

Consider that portrait artists generally refer to live models or photographs when generating their portraits. Perhaps portraiture taxes – and thereby trains – face perception better than face memory. While this may be the case, face perception ability does correlate highly with FRA (e.g. r=0.60 in Bowles et al, 2009). Such a high correlation suggests that face perception and face recognition are only partially distinct. Moreover, it seems likely that face perception provides a direct enough input to face recognition that improving face perception would naturally improve FRA. Of course, we found no evidence for the enhancement of *either* FRA *or* face perception, suggesting that training that could reasonably be expected to train up one or both in fact trained neither. It remains possible that some further optimized and/or even more intensive training regimen could enhance FRA. Yet for any individual to benefit from such a regimen, its costs to them in terms of time spent would need to justify the benefits they gain from it.

Is the limited plasticity implied by these results, in combination with the broader literature, cause for despair? We argue that they need not be. On the contrary, such results can be used to inspire creative compensatory strategies and justify common sense accommodations: these may be low-tech, such as name-tags, picture directories, extra time budgeted for rote learning of faces, or education of friends and colleagues; or they may be high-tech, such as the face recognition software that is now beginning to make its way into consumer computers and mobile devices. From the perspective of employers, implementation of institution-wide or individual-tailored accommodations could be seen as a relatively low-cost way of attracting, retaining, and optimally utilizing talented employees who may vary in their level of

FRA. From the perspective of public policy, the potential value of face recognition software as an accommodation could be seen as a counterpoint to the privacy concerns that currently limit its development and dissemination. Finally, a clear understanding of the apparent limits of FRA's plasticity could empower institutions and individuals to view with a healthy level of skepticism any product they may encounter that glibly claims to improve one's FRA. We therefore view evidence for limited plasticity in FRA as a call to action rather a cause for despair.

In conclusion, we suggest that the present work provides a strong test of the potential for plasticity in FRA, and that the most parsimonious interpretation of our findings in the context of extant literature is that FRA is remarkably non-plastic; that is, highly stable and resilient to change. Is this stability surprising? We suggest that stability in FRA is actually not so surprising if one considers the pervasiveness of faces in our everyday lives. This pervasiveness provides an exceptional regimen of FRA training that plausibly pushes each individual's FRA to its limits.

References

Avidan, G., Tanzer, M., & Berhmann, M. (2011). Impaired holistic processing in
congenital prosopagnosia. *Neuropsychologia, 49,* 2541-2552. doi:
10.1016/j.neuropsychologia.2011.05.002

Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments:
A critical review and future directions. *Frontiers in Human Neuroscience, 8,*
Article 491. doi: 10.3389/fnhum.2014.00491

Bate, S., Cook, S. J., Duchaine, B. C., Tree, J. J., Burns, E. J. & Hodgson, T. L.
(2014) Intranasal inhalation of oxytocin improves face processing
developmental prosopagnosia. *Cortex*, 50, (1), 55-63. doi:
10.1016/j.cortex.2013.08.006

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-
depth cognitive examination of individuals with superior face recognition
skills. *Cortex, 82,* 48-62. doi: 10.1016/j.cortex.2016.05.003

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem:
Evidence of enhanced face matching in individuals with extraordinary face
recognition skills. *PLoS One, 11,* e0148148. doi:
10.1371/journal.pone.0148148

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2015). Super-recognisers in action:
Evidence from face-matching and face memory tasks. *Applied Cognitive
Psychology, 30,* 81-91. doi: 10.1002/acp.3170

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., … Yovel, G.
(2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-

stimulus ethnic match on the Cambridge Face Memory Test and Cambridge

Face Perception Test. *Cognitive Neuropsychology, 26,* 423-455. doi:

10.1080/02643290903343149

Brunsdon, R., Coltheart, M., Nickels, L., & Joy, P. (2006). Developmental

prosopagnosia: A case analysis and treatment study. *Cognitive*

Neuropsychology, 23(6), 822-840. doi: 10.1080/02643290500441841

Choo, S.-J., Wilmer, J., Herzmann, G., McGugin, R. W., Fiset, D., … Gauthier, I.

(2015). Item response theory analyses of the Cambridge Face Memory Test.

Psychological Assessment, 27, 552-566. doi: 10.1037/pas0000068

Crawford, J. R., & Howell, D. C. (1998). Comparing and individual's test score

against norms derived from small samples. *The Clinical Neuropsychologist,*

12, 482-486. doi: 10.1076/clin.12.4.482.7241

Dalrymple, K. A., Corrow, S., Yonas, A., & Duchaine, B. (2012). Developmental

prosopagnosia in childhood. *Cognitive neuropsychology*, 29(5-6), 393-418.

doi: 10.1080/02643294.2012.722547

DeGutis, J. M., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional

plasticity in ventral temporal cortex following cognitive rehabilitation of a

congenital prosopagnosic. *Journal of Cognitive Neuroscience*, *19*, 1790-1802.

doi: 10.1162/jocn.2007.19.11.1790

DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing

improvements in prosopagnosia: successes and failures over the last 50 years.

Frontiers in human neuroscience, *8*: 561. doi: 10.3389/fnhum.2014.00561

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face processing in developmental prosopagnosia. *Brain*, awu062. doi: 10.1093/brain/awu062

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods, 44,* 587-605. doi: 10.3758/s13428-011-0160-2

Dolzycka, D., Herzmann, G., Sommer, W., & Wilhelm, O. (2014). Can training enhance face cognition abilities in middle-aged adults? *PLoS One, 9,* e90249. doi: 10.1371/journal.pone.0090249

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

Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social, Cognitive and Affective Neuroscience, 2,* 104-113. doi: 10.1093/scan/nsm003

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191. doi: 10.3758/BF03193146

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141,* 2-18. doi: 10.1037/a0024338

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition, 21,* 1224-1247. doi: 10.1080/13506285.2013.824530

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition, 61,* 139-158. doi: 10.1016/j.bandc.2005.11.005

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6,* 1108-1112. doi: 10.1038/nn1121

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15,* 762-768. doi: 10.1111/j.0956-7976.2004.00753.x

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255-260. doi: 10.1016/S1364-6613(02)01903-4

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & Zhou (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception, 39,* 1065-1085. doi: 10.1068/p6608

Palermo, R., Rossion, R., Rhodes, G., Laguesse, R., Tez, T., … McKone, E. (in

press). Do people have insight into their face recognition abilities? *The*

Quarterly Journal of Experimental Psychology. doi:

10.1080/17470218.2016.1161058

Plomin, R., DeFries, J. C., Knopik, V. S., and Neiderhiser, J. M. (2013). *Behavioral*

Genetics. 6th Edn. New York: Worth.

Rhodes, G., Jeffery, L., Taylor, L., & Ewing, L. (2013). Autistic traits are linked to

reduced adaptive coding of face identity and selectively poorer face

recognition in men but not women. *Neuropsycholigia, 51,* 2702-2708. doi:

10.1016/j.neuropsychologia.2013.08.016

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face

recognition. *Psychological Science, 22,* 464-471. doi:

10.1177/0956797611401753

Rivolta, D., Schmalzl, L., Coltheart, M., & Palermo, R. (2010). Semantic information

can facilitate covert face recognition in congenital prosopagnosia. *Journal of*

Clinical and Experimental Neuropsychology, 32, 1002-1016. doi:

10.1080/13803391003662710

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with

extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16,*

252-257. doi: 10.3758/PBR.16.2.252

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training

of familiar face recognition and visual scan paths for faces in a child with

congenital prosopagnosia. *Cognitive Neuropsychology*, 25(5), 704-729. doi:

    10.1080/02643290802299350

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item

    prosopagnosia index (PI20): A self-report instrument for identifying

    developmental prosopagnosia. *Royal Society Open Science, 2,* 140343. doi:

    10.1098/rsos.140343

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition.

    Proceedings of the National Academy of Sciences, *112,* 12887-12892. doi:

    10.1073/pnas.1421881112

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual

    expertise in forensic facial image comparison. *Proceedings of the Royal

    Society B, 282*, 20151292. doi: 10.1098/rspb.2015.1292

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., … Duchaine,

    B. (2010a). Human face recognition ability is specific and highly heritable.

    Proceedings of the National Academy of Sciences, 107, 5238-5241. doi:

    10.1073/pnas.0913053107

Wilmer, J. B., Germine, L., Loken, E., Guo, X. M., Chatterjee, G., … Duchaine, B.

    (2010b). Response to Thomas: Is human face recognition entirely genetic?

    Proceedings of the National Academy of Sciences, 107, letter 24. doi:

    10.1073/pnas.1004299107

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama,

    K. (2012). Capturing specific abilities as a window into human individuality:

The example of face recognition. *Cognitive Neuropsychology, 29,* 360-392.

doi: 10.1080/02643294.2012.753433

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008).

Psychosocial consequences of developmental prosopagnosia: A problem of

recognition. *Journal of Psychosomatic Research, 65,* 445-451. doi:

10.1016/j.jpsychores.2008.03.013

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences

reveal about face processing? *Frontiers in Human Neuroscience, 8,* 562. doi:

10.3389/fnhum.2014.00562

Zhou, G., Cheng, Z, Zhang, X., & Wong, A., C.-N. (2012). Smaller holistic

processing of faces associated with drawing experience. *Psychonomic Bulletin

& Review, 19,* 157-162. doi: 10.3758/s13423-011-0174-x

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., ... Liu, J. (2010). Heritability of

the specific cognitive ability of face perception. *Current Biology*, 20, 137-142.

doi: 10.1016/j.cub.2009.11.067

Footnotes

[1]To further control for sex and age effects, we compared the artist sample to a stratified sample of comparison cases ($n = 301$). We were successful in matching gender (% male: artist sample, 21.4%, comparison sample 18.3%, χ^2(1, N = 329) = 0.17, p = .62. Though we reduced the age difference between samples substantially, t(327) = 2.51, p = .013, d = 0.50, 95% CI [0.11, 0.88], we were unable to eliminate the difference entirely due to the small number of adults over 60 in the comparison sample. The mean age of the stratified comparison sample was 35.75 (SD = 11.91). We repeated each of the Mann-Whitney U tests on the sex- and age-controlled z scores using the stratified sample. All results were consistent with the analysis of the full sample: CFMT: U = 3558, U_{CRIT} = 3291.97 p = .16, r = .078; AAMT: U = 3253, U_{CRIT} = 3291.97, p = .041, r = .112; VPMT: U = 4182.5, U_{CRIT} = 3291.97, p = .90, r = .007.