Title page

Effects of age and gender on face perception and memory

Title: Use of the Oxford Face Matching Test Reveals an Effect of Ageing On Face Perception But Not Face Memory

Authors: Mirta Stantić[1], Bethan Hearne[1], Caroline Catmur[2], Geoffrey Bird[1, 3]

Institutions:

[1]Department of Experimental Psychology, University of Oxford, New Radcliffe House, Walton St, Oxford OX2 6BW, United Kingdom

[2]Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, United Kingdom

[3]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom

Correspondence:

Mirta Stantic – mirta.stantic@psy.ox.ac.uk
Geoffrey Bird – geoff.bird@psy.ox.ac.uk
Department of Experimental Psychology,
University of Oxford
New Radcliffe House
Walton St
Oxford OX2 6BW
United Kingdom

Declarations of interest: none.

Contributions Statement

MS: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing; BH: Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization; CC: Conceptualization, Methodology, Writing - Review & Editing, Supervision; GB: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition

Acknowledgments

MS is funded by an ESRC DTP studentship and a Wilfrid Knapp Science Scholarship. GB is supported by the Baily Thomas Charitable Trust.

Effects of age and gender on face perception and memory

Keywords

Oxford Face Matching Test, Face perception, Face memory, Aging

Abstract

Effects of ageing on both face perception and face memory have previously been reported. Previous studies, however, have not controlled for the effects of face perception when assessing face memory, meaning that apparent effects of ageing on face memory may actually be due to effects of ageing on face perception. Here, both face perception and face memory were assessed in a sample of adults ranging in age from 18 to 93, and the effect of age on face memory was assessed after controlling for face perception. Face perception was assessed using both a standard test and the Oxford Face Matching Test (OFMT), deliberately designed to avoid the bias towards younger, neurotypical samples that may be present in other tests. An effect of ageing on face perception was found using both tests, with the unbiased OFMT being more sensitive to the effect of age. Importantly, when controlling for face perception using the OFMT, no effect of age on face memory was found. Indicative scores on the OFMT from a sample of 989 participants are provided, broken down by age and gender.

Highlights

- Effects of age on face perception were found in a large sample of neurotypical adults
- Effects of age on face memory were not found when controlling for face perception
- Indicative scores on the OFMT are provided across a range of ages for each gender

Declarations of interest: none.

It has been widely reported that various aspects of cognition decline with age, including aspects of memory (Naveh-Benjamin et al., 2000; Light, 1991; Bowles et al., 1985) and visual processing (Roberts & Allen, 2016). Of particular interest, given its importance to normal social interaction, is an understanding of face perception and face memory processes in older adults, both of which have been studied extensively. Anecdotally, older adults complain of difficulties in recognizing people and 'matching' a person's face to their name (Schweich et al., 1992). Empirically, findings generally support an age-related decline in both face memory (Maylor, 1990; Maylor & Valentine, 1992; Crook & Larrabee, 1992; Duchaine, Germine, & Nakayama, 2007) and face perception (Duchaine, Germine, & Nakayama, 2007; Shah et al., 2015; Megreya & Bindemann, 2015, Habak, Wilkinson, & Wilson, 2008). Problematically, however, as far as we are aware, no previous study has attempted to examine the age-related decline in face memory after controlling for face perception. This is an issue because the cognitive processes involved in face perception tasks are also required for face memory tasks. Face perception tasks require a perceptual representation of each face to be formed and a judgement made as to whether the two faces are of the same individual. A face memory task, such as naming the face of a famous individual, requires forming a perceptual representation of the stimulus face and a judgement made as to whether that face matches faces of famous individuals stored in memory. Therefore, in order to discover whether age has an effect on face memory independent of its effect on face perception, one must control for face perception. If one does not, an apparent decline in face memory due to ageing may actually result from a decline in face perception, with face memory unaffected.

This discussion highlights the fact that face matching is required when performing both face perception and face memory tasks. Our representations of identities (e.g., "the person in front of me is Mary") are flexible and allow for storage of multiple exemplars, or facial images, of the same person (e.g., "Mary's hair is different than when we last met"). The process of face matching involves deciding whether a particular facial image is sufficiently similar to a facial identity for it to be an image of that identity. This process may be subject to systematic biases or random noise, and either biases or noise may scale with age. Indeed, there has been some suggestion that the decline in performance in older age could be related to an increased rate of false alarms in older adults (Shapiro and Penrod, 1986; Bastin & Van der Lindon, 2003; Bartlett et al., 1989). This would manifest as an increased likelihood of stating that a particular image is an exemplar of a given facial identity when in fact it is not (i.e., a bias towards the ‘match’ response in experimental tasks); or, in day-to-day experience, an increase in mistaking strangers for acquaintances.

Much like other atypical groups (e.g., those with developmental prosopagnosia or autism) who may use different strategies to process faces (Gauthier et al., 2009; Joseph & Tanaka, 2003; Avidan et al., 2011; DeGutis et al., 2012; Palermo et al., 2011), there have also been some reports of older neurotypical adults using different face processing strategies from younger adults. There is evidence for changes in the degree of holistic processing (increase (Konat, Bennett, & Sekuler, 2013) or decrease (Schwarzer, Kretzer, Wimmer, & Jovanovic, 2010; Wiese, Kachel, & Schweinberger, 2013)), or a differential reliance on internal vs external features (Meinhardt-Injac, Persike, & Meinhardt, 2014; Schwarzer, Kretzer, Wimmer, & Jovanovic, 2010) in older adults, but these effects are not always observed (Boutet & Meinhardt-Injac, 2019; Meinhardt-Injac, Boutet, Persike, Meinhardt, & Imhof, 2017; Menihardt-Injac, Persike, & Meinhardt, 2014). Given the possibility of a change in how faces are processed with age, it is of particular importance that a sensitive and unbiased test is used to compare the face perception of older and younger adults. We suggest that using the recently developed Oxford Face Matching Test (OFMT; Stantić et al., 2021a) in an ageing population could assist in answering many of the outstanding questions raised above.

The Oxford Face Matching Test was recently introduced as an unbiased measure of face perception. Pairs of face stimuli are presented to participants who must decide whether the face images are from the same individual, or from different individuals. The test is described as unbiased because the difficulty of the test was calibrated using algorithmic ratings of face similarity, rather than the usual practice of calibration according to neurotypical (typically young student) samples.

To illustrate further, test authors typically aim for average accuracy to be in the range of 75%-80% (when two-alternative-forced-choice tests are used and chance-level performance is 50%). If average performance is approximately 75%, then the test can identify individuals scoring both better and worse than average, and has the best chance of avoiding floor and ceiling effects. This level of average performance is normally achieved using multiple rounds of pilot testing in which different stimuli are trialled in order to achieve the desired accuracy (i.e., both too difficult, and too easy, stimuli are removed). If the participant samples used for stimulus calibration are all of the same type (e.g., all neurotypical, of university age, or from one gender or ethnicity), then test difficulty is unintentionally calibrated to that type of participant. A problem therefore occurs when the performance of two groups is contrasted, one group consisting of the type of individuals that test difficulty was calibrated for, and the other not, particularly when those groups use different strategies to accomplish the same task. It is possible under such a scenario that the two groups would show equivalent performance on the infinite set of stimuli available to researchers, but for group differences to be observed on the specific set of stimuli selected for a particular test.

The OFMT aims to overcome this problem by using ratings of face similarity from three facial recognition algorithms when calibrating stimulus difficulty. Pairs of face images from the same person, or from different people, are rated for their similarity by the algorithms (see Stantić et al., 2021a). When the algorithms all agree, the algorithmic judgements of facial similarity are used to create trials of varying difficulty where very similar facial images from the same people, and very different images from different people, are easy to correctly judge as from the same individual or not. Conversely, very different facial images from the same person, and very similar facial images from different people constitute difficult trials. In addition to standard tests of validity and reliability, the OFMT has been used to demonstrate altered face perception in both those with very poor face memory (developmental prosopagnosics) and those with very good face memory (super recognisers; Stantić et al., 2021a). It was also recently used to investigate face perception in individuals with Autism Spectrum Disorder (Stantić, Ichijo, Catmur, & Bird, 2021b), due to research reporting that autistic individuals are more likely to process faces in a piecemeal, featural fashion, rather than holistically like neurotypical individuals (Gauthier, Klaiman, & Schultz, 2009; Teunisse & de Gelder, 2003; although see e.g., Nishimura, Rutherford, & Maurer, 2008; Brewer, Bird, Grey, & Cook, 2019; Ventura et al., 2018). As noted above, any difference in face processing strategy may lead to standard tests of face processing being biased towards the neurotypical group.

Accordingly, this study reports the face perception and face memory performance of a group of adults in which ages ranged from 18 to 93. Face perception was tested with a standard test, the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010), and the novel unbiased OFMT. Face memory was tested with a standard test of face memory, the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). Self-reported face difficulties were recorded using the Twenty-Item Prosopagnosia Index (PI-20; Shah et al., 2015). Analyses investigated the effect of age on face perception and face memory, and on face memory after controlling for face perception.

Additionally, the study had secondary aims of further validating the OFMT against the GFMT and CFMT in a large, age-diverse sample, and examining the effect of gender on OFMT performance. With respect to the effect of gender, previous research consistently indicates a small but significant difference in performance such that females tend to be better than males on both tasks of face memory and face perception (Lott et al., 2006; Duchaine & Nakayama, 2006; Bowles et al., 2009; Duchaine et al., 2007). Finally, indicative OFMT scores organised by gender and age for almost one thousand participants are provided in order to aid interpretation of OFMT scores in future studies.

Methods

Participants

Participants were recruited through a combination of recruitment platforms, including social media, the recruitment platform Prolific, and participant databases. Participants were excluded when not meeting the study inclusion criteria (over 18 years old, no current psychiatric diagnosis, no psychotropic medication currently or in the last 6 months), for failing the attention checks on the OFMT (80% accuracy on attention checks was required), or for being potentially prosopagnosic (CFMT score lower than 42 along with PI-20 score higher than 72). Note that participant ethnicity was not recorded, and that this is a limitation of the study due to the existence of the 'other-race effect', the finding of superior ability to recognise faces of one's own ethnicity, compared to faces of another ethnicity (see Meissner & Brigham, 2001 for review).

607 participants completed the task online, after which 16 participants were excluded for previously defined attention check failures and 8 additional participants were excluded for being potentially prosopagnosic (scoring above 72 on the PI-20 and below 42 on the CFMT). In the final sample there were 355 females (*M(age)* = 47.4, *SD(age)* = 17.6, range = 18-83), 224 males (*M(age)* = 53.8, *SD(age)* = 15.6, range = 19-93), and 4 participants who would rather not say (*M(age)* = 37.0, *SD(age)* = 19.1, range = 26-65). Although age was used as a continuous variable across all analyses, stratified recruitment was used to ensure at least 90

participants in each decade-wide (X0-X9) age bracket up to 70 years old, with those over 70 treated as one group for the purposes of recruitment. After exclusions, all reported analyses include 90 participants in the 18-29 age group, 82 participants in the 30-39 age group, 88 participants in the 40-49 age group, 149 participants in the 50-59 age group, 87 participants in the 60-69 age group, and 87 participants in the 70+ age group. Gender distribution was not equalised across age brackets and the proportion of males varied between 15% and 56%.

Tasks

The Oxford Face Matching Test (OFMT) (Stantić et al., 2021a)

On each trial of the OFMT participants have to indicate whether two faces presented for 1600ms are of the same individual or different individuals, as well as how similar the faces are on a scale from 1 (very dissimilar) to 100 (very similar). There is no time limit to respond. Faces presented in the test are male and female, all Caucasian, and range in age from 18 to 70. Faces are presented in frontal view, on a blank background, with the hair, jawline, and ears visible. All faces are presented in greyscale. Images were obtained from a number of publicly available databases as well as authors' personal databases. 200 trials are presented in a random order across four blocks of 50 trials. Scores can range from 0 to 200. OFMT scores were previously shown to be reliable (correlation between tests r = .75 over a 2-weeks delay; Stantić et al., 2021a). See Figure 1A for a sample trial.

The Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006)

The CFMT is a test of face memory in which participants are initially required to learn six target faces from different viewpoints (left, frontal, right). Following the learning phase, participants complete 72 test trials in which they are required to identify the learned identity among two distractors. Test trials get progressively more difficult with the addition of visual white noise. All of the faces are of young Caucasian men. Stimuli were obtained from the test authors' database and faces are presented in greyscale and cropped to exclude external facial features. Scores can range from 0 to 72. CFMT scores were previously shown to be reliable (r = .68-.92; Bate et al., 2020, Stantić et al., 2021a; Wilmer et al., 2010; McKone et al., 2011). See Figure 1C for a sample trial.

The Glasgow Face Matching Test (GFMT) (Burton, White, & McNeill, 2010)

On each trial of the GFMT participants have to indicate whether two faces presented simultaneously are of the same individual or different individuals. There is no time limit to view the faces or to respond. Face presented in the test are of young Caucasian men (*M(age)* = 22.9, *SD(age)* = 6.7) and women (*M(age)* = 23.2, *SD(age)* = 7.0). Greyscale faces are presented in frontal view, on a blank background, with hair, jawline, and ears visible. Images were obtained from the test authors' database. The 40 trials in the short version of the test are presented in a random order. Scores can range from 0 to 40. GFMT scores were previously shown to be reliable (*r* =.77, Stantić et al., 2021a). See Figure 1B for a sample trial.

20-Item Prosopagnosia Index (PI-20) (Gray, Bird, & Cook, 2017)

The PI-20 is a self-report questionnaire used to identify difficulties in face recognition. It has previously been found to correlate moderately well with established tests of face perception and memory (Gray, Bird, & Cook, 2017), particularly at the extremes of face recognition ability (Bobak, Mileva, & Hancock, 2019). Participants rate their agreement with 20 statements describing their face recognition ability on a five-point Likert scale. Scores can range from 20 to 100, with lower scores indicating lower reported difficulties and higher scores indicating higher reported difficulties with face perception in everyday life. PI-20 scores were previously shown to be reliable (*r* = .89; Stantić et al., 2021a).

Visual acuity chart (Hetherington, 1954)

At the end of the study, participants were shown a standard Snellen visual acuity chart and asked to input the last line they could read without zooming in or getting closer to the screen. An exclusion criterion was defined to remove data from any participants whose scores were not within the 25-75th percentile norms for their age (none met this criterion). A significant correlation between age and visual acuity was observed (r = .40, p < .001), however inclusion of visual acuity as a control variable did not alter the pattern of significance in any analysis reported below.

Procedure

Participants completed the tasks in a randomized order after providing their demographic information and before the visual acuity chart. They were able to take breaks between tasks, as well as during tasks where provided. Completion of all tasks took approximately 45 minutes on average. Participants were able to adjust the size of stimuli to account for differences in screen size, such that stimuli could be presented at the same relative visual angle for all participants. The procedure was approved by the Central University Research Ethics Committee at the University of Oxford under approval R71495/RE001. No part of the procedure or analyses were preregistered. A sample trial for each task is shown in Figure 1.

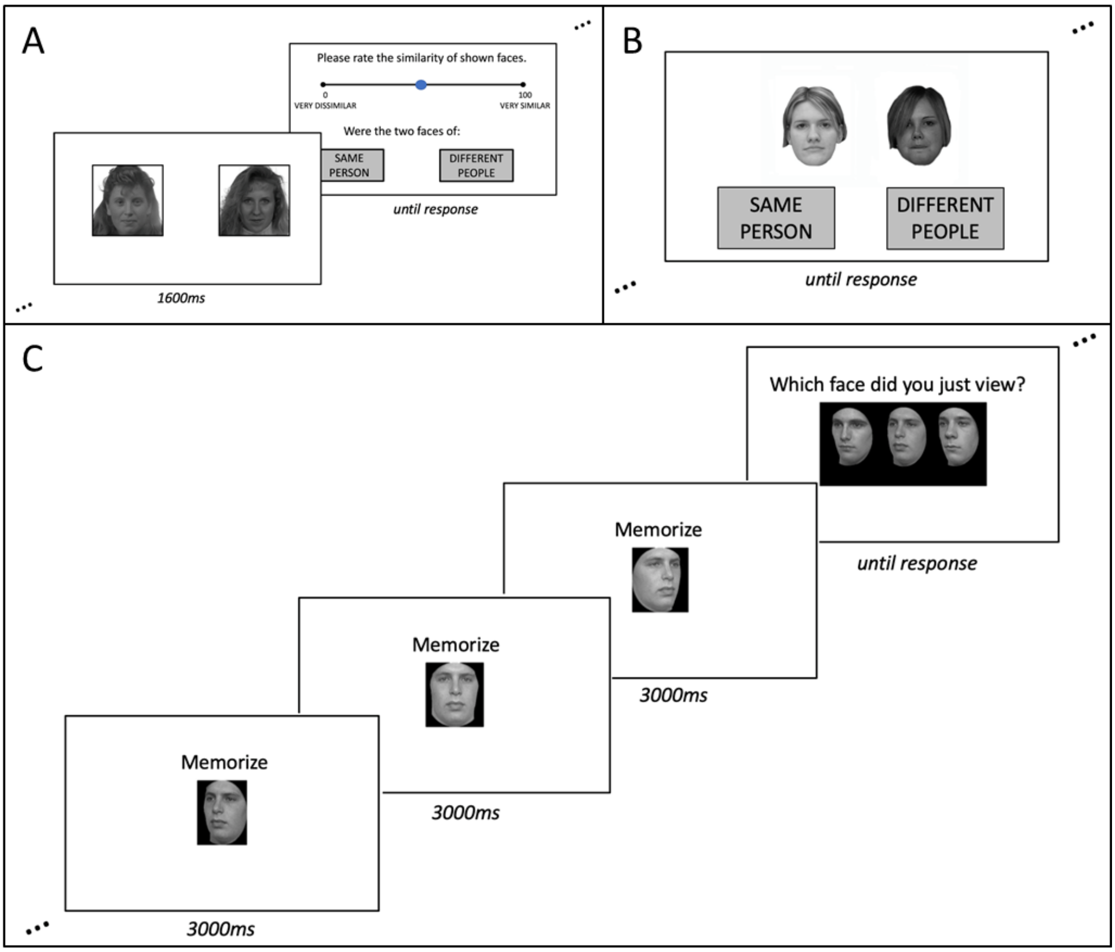

Figure 1. Sample trials of the A) Oxford Face Matching Test, B) Glasgow Face Matching Test, and C) Cambridge Face Memory Test.

Results

Mean, standard deviation of the mean, and the range, of scores obtained on each test separated by gender is reported in Table 1.

In order to determine the effect on age on face processing a hierarchical multiple regression was conducted for each test which included linear, quadratic and cubic effects of age on task performance. The best fitting model is reported for each face processing measure.

Effect of Age on Face Perception

OFMT: The effect of age on OFMT performance was best explained by a regression model containing both linear and quadratic effects of age. A linear model showed age significantly predicted performance (β = -.419, t = -11.11, p < .001, R^2 = .175). The addition of the quadratic effect significantly increased the amount of variance explained (quadratic effect: β = -.116, t = -3.08, p = .002, R^2 = .189, R^2_{change} = .014), but the addition of the cubic effect did not (cubic effect: β = .050, t = .59, p = .553, R^2 = .189, R^2_{change} = .00). The model containing both linear and quadratic effects therefore showed that age was a significant predictor of OFMT performance (F(2, 579) = 67.33, p < .001) and this model explained 18.9% of the variance in performance, see Figure 2, upper left panel. Age was significantly correlated with the rate of false alarms (r(582) = .29, p < .001) and misses (r(582) = .17, p < .001) on the OFMT, with the correlation between age and false alarms significantly greater than the correlation between age and misses (z = 2.31, p = .02).

GFMT: The effect of age on GFMT performance was also best explained by a regression model containing both linear and quadratic effects of age. A linear model showed age significantly predicted performance (β = -.212, t = -5.22, p < .001, R^2 = .043). The addition of the quadratic effect significantly increased the amount of variance in performance explained (quadratic effect: β = -.177, t = -4.42, p < .001, R^2 = .073, R^2_{change} = .031), but the addition of the cubic effect did not (cubic effect: β = -.126, t = -1.39, p = .164, R^2 = .074, R^2_{change} = .003). The model containing both linear and quadratic effects therefore showed that age was a significant predictor of GFMT performance (F(2, 579) = 23.83, p < .001) and this model explained 7.6% of the variance in performance, see Figure 2, upper right panel. Given that GFMT allows for unlimited viewing time, it is possible for the errors to be driven by faster response times (i.e., that older participants show a speed-accuracy trade-off such that they sacrifice accuracy for increased speed). However, the data indicate that the reduced accuracy with age could not be due to a speed-accuracy trade-off, as reaction time increased with age in a linear fashion (β = .30, t = 7.65, p < .001, R^2 = .092). In contrast to the OFMT, age was correlated with the rate of misses (r(582) = .30, p < .001) but not false alarms (r(582) = -.02, *p* = .624) on the GFMT, and the correlation between age and misses was significantly greater than that between age and false alarms (z =-4.863, $p < .001$).

Formal comparison of correlations (OFMT:age and GFMT:age) revealed that the OFMT was more sensitive to the linear effect of age than the GFMT ($z = 3.88$, $p < .001$). However, there was no significant difference in sensitivity between tests when comparing correlations with the quadratic effect of age ($z = 1.4$, $p = .16$).

Effect of Age on Face Memory

The effect of age on face memory was assessed in the same manner as the effect of age on face perception, using CFMT scores as the dependant variable. Results revealed that the effect of age on CFMT performance was also best explained by a regression model containing both linear and quadratic effects of age. A linear model showed age significantly predicted performance ($\beta = -.249$, $t = -6.21$, $p < .001$, $R^2 = .062$). The addition of the quadratic effect significantly increased the amount of variance explained by the model (quadratic effect $\beta = -.131$, $t = -3.28$, $p = .001$, $R^2 = .079$, $R^2_{change} = .017$), but the addition of the cubic effect did not (cubic effect: $\beta = -.104$, $t = -1.15$, $p = .249$, $R^2 = .077$, $R^2_{change} = .002$). The model containing both linear and quadratic effects therefore showed that age was a significant predictor of CFMT performance ($F(2, 579) = 24.96$, $p < .001$), and this model explained 7.9% of the variance in performance, see Figure 2, lower left panel.

To control for the effects of face perception on CFMT scores, and therefore to examine the effect of age on face memory independent of effects of face perception, two further regression analyses were performed. The first regressed CFMT scores on age, GFMT scores (so that age effects on face memory can be identified independent of face perception by holding face perception constant), and the age x GFMT score interaction (allowing for the relationship between face perception and face memory to vary by age). The second regression analysis was identical but included OFMT rather than GFMT scores.

The regression conducted using GFMT scores revealed a significant effect of age ($\beta = -.55$, $t = -2.07$, $p = .04$), indicating an effect of age on face memory even after controlling for face perception; however, in the analysis conducted using the OFMT the effect of age on face memory was not significant after accounting for face perception ($\beta = -.65$, $t = -1.54$, $p = .12$). This is likely due to the OFMT's greater sensitivity in detecting the effect of age on face perception, and also its increased sensitivity to individual differences in the extremely good range of performance (Stantić et al., 2021a). Indeed, the latter finding was replicated in the current data where 16 participants scored perfectly on the GFMT with a further 65 only making 1 or 2 mistakes, compared to 0 participants making these many errors on the OFMT. The problem with ceiling effects on the GFMT means that sensitivity is lost for good performers. Note that the pattern of significance was not changed by the inclusion of gender in either regression, and neither gender (OFMT: $\beta = .03$, $t = 0.81$, $p = .42$; GFMT: $\beta = .03$, $t = 0.79$, $p = .43$) nor the age x face perception interaction (OFMT: $\beta = .57$, $t = 1.48$, $p = .14$; GFMT: $\beta = .39$, $t = 1.46$, $p = .43$) was significant in either regression. The pattern of significance was also identical when the age x face perception interaction term was not included in the original regressions.

Lack of evidence against the null hypothesis (that face memory is unaffected by age after accounting for face perception using the OFMT) provided by a p-value above .05 is not evidence for the null hypothesis. In order to contrast evidence for the null and alternative hypotheses one needs to calculate a Bayes Factor. Accordingly, two Bayesian correlations were conducted using JASP (JASP Team, 2020, Version 0.14.1). The first correlated age with CFMT scores after OFMT scores had been regressed out (i.e., the residuals from a regression in which CFMT scores were the dependent variable and OFMT scores the independent variable), and the second correlated age with CFMT scores after GFMT scores had been regressed out. In the absence of a more informed prior, JASP defaults were used to calculate Bayes Factors (BF_{10}; indicating the ratio of evidence for the alternative hypothesis over the null hypothesis). Bayes Factors supported the conclusions drawn from the frequentist analysis, with a BF_{10} of 0.108 providing moderate-to-strong evidence *against* a relationship between age and OFMT-corrected CFMT scores, and a BF_{10} of 603.412 providing strong evidence *for* a relationship between age and GFMT-corrected CFMT scores[i].

Effect of Age on Self-Reported Difficulties with Face Recognition

A hierarchical multiple regression revealed that PI-20 scores were best explained by a linear effect of age ($\beta = .153$, $t = 3.73$, $p < .001$, $R^2 = .023$), see Figure 2, lower right panel.

Effect of Gender – All Measures

See Table 1 for gender-specific means on all measures. A small, but significant effect of gender was observed for OFMT performance ($t(577) = 3.76$, $p < .001$, $d = 0.32$), whereby females scored higher than males. There was no interaction between age and gender ($F(61, 449) = 1.03$, $p = .41$). The female advantage was also significant for the GFMT $t(577) = 3.24$, $p = .001$, $d = 0.27$), CFMT ($t(577) = 2.45$, $p = .014$, $d = 0.21$), and PI-20 ($t(577) = -2.11$, $p = .035$, $d = 0.18$). There were no interactions between age and gender for any of the other tests: GFMT ($F(61, 449) = 1.23$, $p = .12$); CFMT ($F(61, 449) = .91$, $p = .68$); or PI-20 ($F(61, 449) = 1.04$, $p = .41$).

Table 1. Descriptive statistics for all 583 participants, showing average performance, standard deviation, and range of scores for all face perception tasks (Oxford Face Matching Test, Glasgow Face Matching Test, Cambridge Face Memory Test and 20-Item Prosopagnosia Index) split by gender. The somewhat lower CFMT scores observed here compared to previous studies are likely driven by the increased representation of participants over 60 years of age, who score below previously reported mean scores (M=46.78).

		M±SD		**Range [min-max]**	
OFMT		153.80±13.06		110-190	
	Male		151.29±12.97		110-181
	Female		155.42±12.77		121-190
	Other		170.00±18.08		151-187
GFMT		32.92±4.48		18-40	
	Male		32.15±4.75		18-40
	Female		33.38±4.25		20-40
	Other		36.67±3.51		33-40
CFMT		50.43±10.40		15-72	
	Male		49.05±10.28		27-72
	Female		51.22±10.36		15-72
	Other		60.33±13.20		46-72
PI-20		45.05±12.65		21-94	
	Male		46.44±12.76		24-84
	Female		44.16±12.57		21-94
	Other		47.00±5.29		43-53

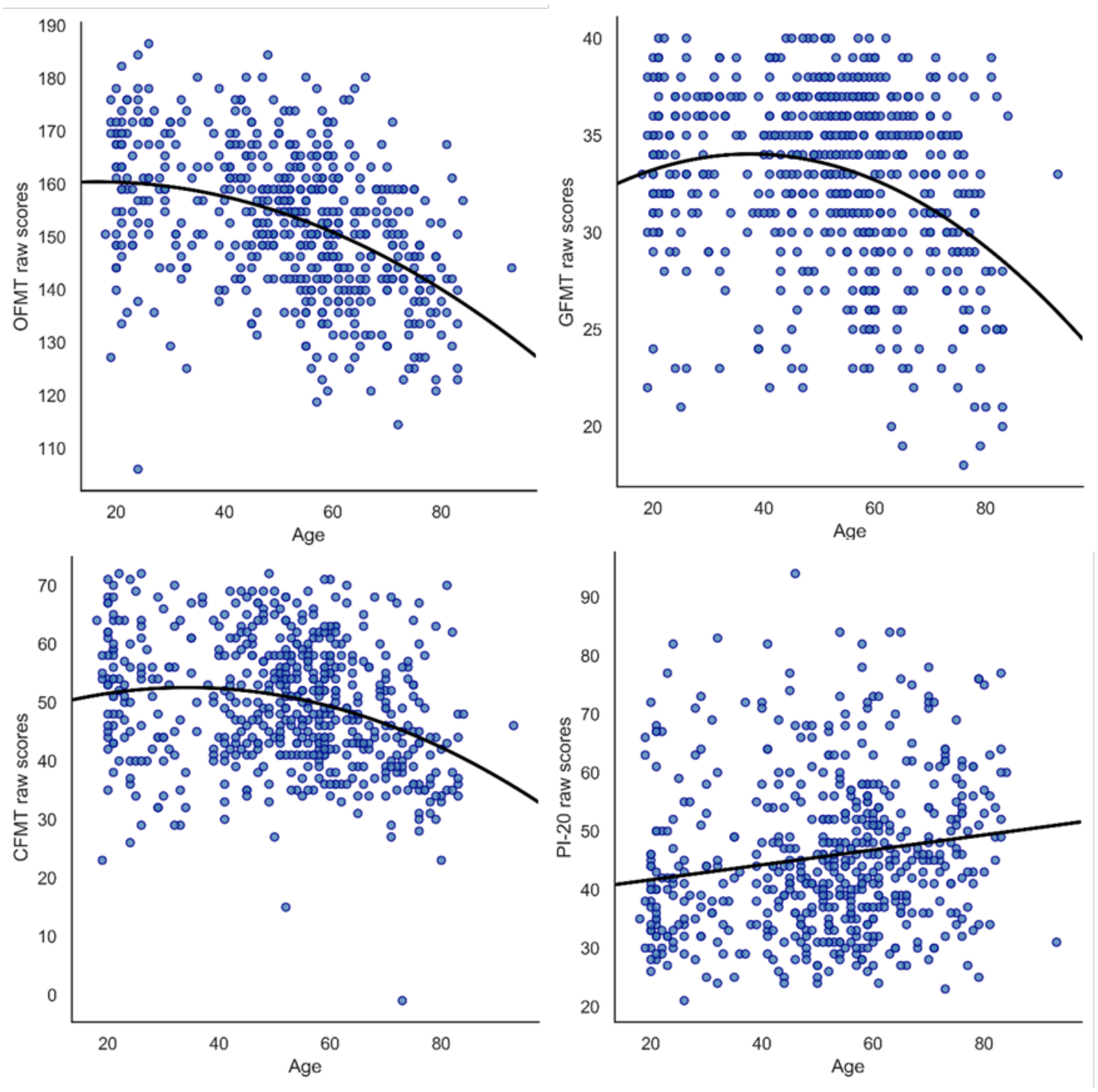

Figure 2. Performance on all face tasks shown as a function of age with the best fit line for the model that explains most variance. Upper left: Oxford Face Matching Test (OFMT) performance shown as a function of age with the best fit line for the quadratic regression model; Upper right: Glasgow Face Matching Test (GFMT) performance shown as a function of age with the best fit line for the quadratic regression model; Lower left: Cambridge Face Memory Test (CFMT) performance shown as a function of age with the best fit line for the quadratic regression model. Note that the relationship between age and CFMT performance does not hold after accounting for the effect of age on face perception using the OFMT; Lower right: 20-Item Prosopagnosia Index (PI-20) scores shown as a function of age with the best fit line for the linear regression model, note that higher scores on the PI-20 indicate a higher incidence of self-reported face difficulties in everyday life.

Relationships Between OFMT and Other Face Processing Measures

OFMT scores showed moderate-to-strong positive relationships with GFMT (r = .45, p < .001) and CFMT (r =.50, p < .001) scores, as previously found in the initial validation of the OFMT (Stantić et al., 2021a). Correlations between other tests, as well as between other tests and OFMT scores on match and mismatch trials separately, are shown in Table 2.

Table 2. Relationships between all face perception measures included in the battery for all 583 participants. Two asterisks denote significance at the level of p < .001. CFMT = Cambridge Face Memory Test, PI-20 = 20-Item Prosopagnosia Index, GFMT = Glasgow Face Matching Test, OFMT = Oxford Face Matching Test, Match = accuracy on same/match trials of the OFMT; Mismatch = accuracy on different/mismatch trials of the OFMT.

	M	***SD***	**CFMT**	**PI-20**	**GFMT**	**OFMT**	**OFMT match trials**	**OFMT mismatch trials**
CFMT	50.43	10.40	-					
PI-20	45.05	12.65	-.26**	-				
GFMT	32.92	4.48	.41**	-.28**	-			
OFMT	153.90	13.05	.50**	-.28**	.45**	-		
Match	71.95	13.72	.30**	-.19**	.20**	.60**	-	
Mismatch	81.50	11.98	.19**	-.09*	.27**	.39**	-.49**	-

OFMT Indicative Scores by Age and Gender

The sample described above was combined with previously-collected data from 406 neurotypical adults (276 female, 130 male) making a combined sample of 989 individuals (631 women, 354 men, 4 other/would rather not say). Mean (and SD) OFMT scores from this sample are reported below organised by age, and further subdivided by gender (Table 3).

Table 3. Distribution of scores on the OFMT by age and gender. Age brackets are defined to be inclusive (e.g., participants in the age bracket 30-39 include both 30-year-olds and 39-year-olds.) Participants who defined their gender as 'other' or 'would rather not say' are included in the average age scores but not the gender-specific scores.

		n	**OFMT** ***M±SD***	**OFMT Range**
18-29		409	157.12±13.23	107-187
	Male	101	150.75±13.85	107-179
	Female	305	159.10±12.26	107-187
30-39		113	154.18±14.71	104-180
	Male	42	148.83±15.83	104-179
	Female	70	157.00±12.90	119-180
40-49		119	156.45±12.25	126-190
	Male	64	153.97±11.94	126-181
	Female	55	159.33±12.08	128-190
50-59		160	153.01±12.37	119-181
	Male	53	152.94±10.70	122-174
	Female	106	153.07±13.22	119-181
60-69		91	148.63±12.53	120-180
	Male	43	148.35±12.42	120-176
	Female	47	149.28±12.58	128-180
70+		93	142.92±11.24	115-171
	Male	45	141.76±12.18	115-171
	Female	48	144.02±10.29	123-168

Discussion

This study aimed to test the effect of ageing on face perception using both a standard test (GFMT) and a novel test (OFMT), which relies on unbiased algorithmic judgments to determine difficulty of items. The effect of ageing on face memory was also tested using the CFMT. Crucially, the effect of age on face memory was assessed both before, and after, accounting for any effects on face perception. The study also assessed the relationship between the OFMT and the other face measures, and the effect of gender on all measures.

With respect to face perception, both the OFMT and GFMT revealed a linear and a quadratic relationship with age whereby the effect of age on face perception increased as age increased (i.e., the rate of age-related decline in face perception ability increased as age increased). The OFMT proved to be the more sensitive measure. The increased sensitivity of the OFMT likely stems from at least three factors: its larger number of trials compared to the GFMT, the absence of a celling effect for superior performers, as well as its use of algorithmic similarities to determine the difficulty of individual items. The increased sensitivity to the effects of age for the OFMT compared to the GFMT proved crucial when the effect of age on face memory was investigated. Both an analysis of CFMT scores without accounting for face perception, and when accounting for face perception using the GFMT, suggested that ageing impacted face memory. However, after controlling for face perception using the more sensitive OFMT, no effect of ageing on face memory was identified using frequentist or Bayesian analyses. To our knowledge, this is the first study to examine the impact of ageing on face memory independent of face perception, and so this result is in need of replication. However, if reliable, this result suggests interventions to improve face recognition in older adults should primarily be targeting face perception.

As outlined in the Introduction, the matching, or decision-making aspect, of these tasks is also potentially susceptible to effects of ageing. There have been previous suggestions that the rate of false alarms increases in age (Shapiro & Penrod, 1986; Bastin & Van der Lindon, 2003; Bartlett et al., 1989), and this effect was partially supported in these data. Age was significantly associated with an increased rate of both false alarms and misses on the OFMT, but the correlation between age and false alarms was stronger than between age and misses. On the GFMT, however, the correlation with age was significant only for misses and not false alarms, with the correlation between age and misses significantly stronger than that for age and false alarms.

With respect to OFMT validation, as expected, correlations between OFMT scores and the GFMT and CFMT were moderate-to-strong. Data from a combined sample of nearly one thousand participants were used to provide indicative scores for various age groups, and for males and females within those age groups, on the OFMT. Furthermore, a small but significant effect of gender was observed for OFMT scores (and all other tests) indicating a female advantage. McKelvie (1981) proposed that gender differences in face processing are the result of different interests, suggesting that men are less interested in faces than women and therefore worse at processing them. However, research on a range of cognitive tasks tends to find a slight female advantage in performance (Van Exel et al., 2001), a result interpreted as suggesting that females try harder on experimental tests than males. Thus, females are either slightly better at face processing than males, and/or put more effort into taking part in experimental studies in general.

It is worth noting a potential limitation with the study relating to the age of the stimulus faces. Stimulus faces in the GFMT are on average approximately 23 years old, and the CFMT includes faces of approximately the same average age. Stimulus faces in the OFMT are from a wider range of ages, resulting in an older average age of approximately 45 years. It has previously been shown that, similar to the other-race effect (Meissner & Bringham, 2001), an equivalent age-effect exists such that older adults recognize faces of other older adults better than those of younger adults (Lamont, Stewart-Williams, & Podd, 2005, Kuefner et al., 2008). Therefore, it is possible that the relatively young age of stimulus faces in the GFMT and CFMT explain some of the age effects on those tasks, such that as the age of the participant gets further from the age of the stimulus faces, performance decreases. This cannot account for the age effect on the OFMT however, as the age of the stimulus faces corresponds to the mean age of participants (meaning that performance would be impaired at both the younger and older ends of the age spectrum).

A more substantial limitation of this study is that the specificity of any effects of ageing with respect to stimulus class cannot be established, i.e., it cannot be ascertained whether the impaired face perception seen in older adults is specific to faces, or whether similar effects would be observed for comparable non-face stimuli. There are certainly effects of ageing on visual perception in general, including on colour perception, temporal resolution, visual acuity, motion perception, and a loss of high spatial frequencies (see Roberts & Allen, 2016, for a review), making it plausible that age-related decline in face perception is the result of more general effects of ageing on visual perception. In addition, due to the length of the testing session, no other measures of cognitive decline were included in this study, which may have identified any abnormal cognitive decline in the older participants not captured by a simple visual acuity test. It is, of course, likely that general cognitive decline associated with normal ageing affects both face matching and face memory, as well as a number of other cognitive capacities.

Further limitations of the study are shared with all studies that adopt a similar design. Most obvious is the fact that cross-sectional studies cannot separate true effects of age from cohort effects. Also, due to the online nature of our study, the test of visual acuity had to be conducted remotely. This meant it was not possible to control testing conditions as much as in the lab. Furthermore, it is possible that older people have less access to technology than younger individuals. Older individuals do not tend to use computers as much as younger individuals, and tend to perceive their use as less relevant to their everyday lives (Selwyn, 2004) and be more anxious about, or less interested in, using computers (Laguna and Babcock, 1997; Van Deursen & Helsper, 2015). Consequently, the online nature of this study may have served as more of a selection criterion in our older participants such that only older adults who are more confident with technology took part. To avoid this as much as possible, extensive support was offered in setting up the experiment, and no feedback was received to indicate that the use of technology had been an obstacle to taking part, or that participants struggled to understand the instructions for any of the tasks.

In conclusion, using an unbiased and sensitive test of face perception, a significant reduction in face perception with age was observed, but this was not accompanied by an effect of age on face memory. Instead, results suggested that effects of age on face memory were explained by effects on face perception. Distributions of scores on the OFMT broken down by age and gender are provided in order to facilitate comparisons in future studies of social cognition across the lifespan.

References

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

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, *17*(1), 14.

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly journal of experimental psychology*, *72*(4), 872-881.

Boutet, I., & Meinhardt-Injac, B. (2019). Age differences in face processing: the role of perceptual degradation and holistic processing. *The Journals of Gerontology: Series B*, 74(6), 933-942.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Ellie Wilson, C., & 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*(5), 423-455.

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of gerontology*, *40*(1), 71-77.

Brewer, R., Bird, G., Gray, K. L., & Cook, R. (2019). Face perception in autism spectrum disorder: Modulation of holistic processing by facial emotion. *Cognition*, 193, 104016.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. *Behavior, aging and the nervous system*, *114*, 49.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286-291.

Crook III, T. H., & Larrabee, G. J. (1992). Changes in facial recognition memory across the adult life span. *Journal of Gerontology*, *47*(3), P138-P141.

DeGutis, J., Cohan, S., Mercado, R. J., Wilmer, J., & Nakayama, K. (2012). Holistic processing of the mouth but not the eyes in developmental prosopagnosia. Cognitive Neuropsychology, 29(5-6), 419-446.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576-585.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive neuropsychology*, *24*(4), 419-430.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive neuropsychology*, *24*(4), 419-430.

Faubert, J. (2002). Visual perception and aging. *Canadian Journal of Experimental Psychology*, 56(3), 164.

Gauthier, I., Klaiman, C., & Schultz, R. T. (2009). Face composite effects reveal abnormal face processing in Autism spectrum disorders. *Vision Research*, 49(4), 470-478.

Grady, C. L., Maisog, J. M., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., ... & Haxby, J. V. (1994). Age-related changes in cortical blood flow activation during visual processing of faces and location. *Journal of Neuroscience*, *14*(3), 1450-1462.

Gray, K. L., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royal Society Open Science*, *4*(3), 160923.

Habak, C., Wilkinson, F., & Wilson, H. R. (2008). Aging disrupts the neural transformations that link facial identity across views. *Vision research*, *48*(1), 9-15.

JASP Team (2020). JASP (Version 0.14.1)[Computer software].

Joseph, R. M., & Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of child psychology and psychiatry*, 44(4), 529-542

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision research*, 88, 38-46.

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.

Laguna, K., & Babcock, R. L. (1997). Computer anxiety in young and older adults: Implications for human-computer interactions in older populations. *Computers in human behavior*, *13*(3), 317-326.

Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. Memory & cognition, 33(6), 1017-1024.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual review of psychology*, *42*(1), 333-376.

Lott, L. A., Haegerstrom-Portnoy, G., Schneck, M. E., & Brabyn, J. A. (2005). Face recognition in the elderly. *Optometry and Vision Science*, *82*(10), 874-881.

Maylor, E. A. (1990). Recognizing and naming faces: Aging, memory retrieval, and the tip of the tongue state. *Journal of Gerontology*, *45*(6), P215-P226.

Maylor, E. A., & Valentine, T. (1992). Linear and nonlinear effects of aging on categorizing and naming faces. *Psychology and Aging*, *7*(2), 317.

McKelvie, S. J., Standing, L., Jean, D. S., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society*, *31*(5), 447-448.

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5-22.

Meinhardt-Injac, B., Boutet, I., Persike, M., Meinhardt, G., & Imhof, M. (2017). From development to aging: holistic face perception in children, younger and older adults. *Cognition*, 158, 134-146.

Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2014). Holistic processing and reliance on global viewing strategies in older adults' face perception. *Acta psychologica*, 151, 155-163.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3.

Naveh-Benjamin, M., Craik, F. I., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *The Quarterly Journal of Experimental Psychology Section A*, *53*(3), 609-625.

Nishimura, M., Rutherford, M. D., & Maurer, D. (2008). Converging evidence of configural processing of faces in high-functioning adults with autism spectrum disorders. *Visual Cognition*, 16(7), 859-891.

Owsley, C. (2016). Vision and aging. *Annual review of vision science*, 2, 255-271.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, 49(5), 1226-1235.

Roberts, K. L., & Allen, H. A. (2016). Perception and cognition in the ageing brain: a brief review of the short-and long-term links between perceptual and cognitive decline. *Frontiers in aging neuroscience*, 8, 39.

Schwarzer, G., Kretzer, M., Wimmer, D., & Jovanovic, B. (2010). Holistic face processing among school children, younger and older adults. European Journal of Developmental Psychology, 7(4), 511-528.

Schweich, M., Van der Linden, M., Brédart, S., Bruyer, R., Nelles, B., & Schils, J. P. (1992). Daily-life difficulties in person recognition reported by young and elderly subjects. *Applied Cognitive Psychology*, *6*(2), 161-172.

Selwyn, N. (2004). The information aged: A qualitative study of older adults' use of information and communications technology. *Journal of Aging studies*, *18*(4), 369-384.

Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P., & Tyler, L. K. (2007). On the tip-of-the-tongue: neural correlates of increased word-finding failures in normal aging. *Journal of cognitive neuroscience*, *19*(12), 2060-2070.

Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20-item prosopagnosia index (PI20): relationship with the Glasgow face-matching test. *Royal Society Open Science*, *2*(11), 150305.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139.

Stantić, M., Brewer, R., Duchaine, B., Banissy, M. J., Bate, S., Susilo, T., Catmur, C. & Bird, G. (2021a). The Oxford Face Matching Test: A Non-Biased Test of The Full Range of Individual Differences in Face Perception. *Behavior Research Methods.*

Stantić, M., Ichijo, E., Catmur, C., & Bird, G. (2021b) Short Report: Face Memory and Face Perception in Autism. *Autism*.

Teunisse, J. P., & De Gelder, B. (2003). Face processing in adolescents with autistic disorder: The inversion and composite effects. Brain and cognition, 52(3), 285-294.

Van Deursen, A. J., & Helsper, E. J. (2015). A nuanced understanding of Internet use and non-use among the elderly. *European journal of communication*, 30(2), 171-187.

Van Exel, E., Gussekloo, J., De Craen, A. J. M., Bootsma-Van Der Wiel, A., Houx, P., Knook, D. L., & Westendorp, R. G. J. (2001). Cognitive function in the oldest old: women perform better than men. *Journal of Neurology, Neurosurgery & Psychiatry*, *71*(1), 29-32.

Ventura, P., Carmo, J. C., Souza, C., Martins, F., Leite, I., Pinho, S., Barahona-Correa, B., & Filipe, C. N. (2018). Holistic processing of faces is intact in adults with autism spectrum disorder. *Visual Cognition*, 26(1), 13-24.

Wiese, H., Kachel, U., & Schweinberger, S. R. (2013). Holistic face processing of own-and other-age faces in young and older adults: ERP evidence from the composite face task. *Neuroimage*, 74, 306-317.

---

[i] An anonymous reviewer suggested the use of the following thresholds for the PI-20 and CFMT - 65 and 44, respectively. Using these thresholds (which resulted in the removal of 25 participants) the pattern of significance was as reported above for the OFMT. The GFMT Bayesian results were also as reported above, although the Frequentist results mirrored those for the OFMT, supporting the claim that face memory deficits in older adults can be accounted for by effects of ageing on face perception.