Vision Research 199 (2022) 108071

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Seeing faces where there are none: Pareidolia correlates with age but not autism traits

Muhammad Rahman, Jeroen J.A. van Boxtel *

Discipline of Psychology, School of Health Sciences, Faculty of Health, University of Canberra, 2617 BRUCE, ACT, Australia

ARTICLE INFO

Keywords:
Autism spectrum disorder
Autistic-like traits
Face processing
Pareidolia
Age

ABSTRACT

Previous research has found that individuals with autism spectrum disorder experience difficulties when visually processing face stimuli compared to developmentally typical individuals. Whether, in the typically-developing population, face detection depends on autism-like traits (ALTs) is less clear. In this report, we aimed to develop an experimental design that is more sensitive to any individual differences in face detection than previous reports. We employed pareidolia, that is, cases where non-face stimuli are perceived to be faces, assuming this is more difficult than detection of 'real' faces, decreasing changes of ceiling performance. We also show multiple faces per trial, allowing for a more graded assessment of face detection ability. Participants were 263 individuals aged between 18 and 82 years of age. Pareidolia was investigated in two online experiments, with different types of stimuli: objects that could be perceived as faces (i.e., embedded faces task) and Mooney faces (Mooney face task). In the latter condition, we also investigated the face inversion effect. We found that neither detection ability or the inversion effect depended on ALTs. We did find a dependence of age for both measures, and a complex dependence on gender for Mooney faces. Our data suggest that face detection (and specifically pareidolia) does not depend on ALTs, but does depend on the age of the observer. The dependence on age appears to be different between the two experiments, suggesting that the underlying mechanisms necessary for face detection in our two experiments mature and decline at different rates.

1. Introduction

Face perception is a crucial part of social perception. Information transmitted from the face assists our ability to recognise familiar faces, and to infer the emotional dispositions of others. A number of disorders (e.g., autism spectrum disorders, and schizophrenia) feature deficits in facial perception, impacting social cognition, and interpreting social and emotional cues. Compared to typical developing individuals, individuals with Autism Spectrum Disorders (ASD) experience difficulties when visually processing face stimuli (Blair, Frith, Smith, Abell, & Cipolotti, 2002; Faja, Webb, Merkle, Aylward, & Dawson, 2009; Scherf, Behrmann, Minshew, & Luna, 2008; Wallace, Coleman, & Bailey, 2008). Specifically, this impairment includes, but is not limited to, a deficit in face memory (Boucher & Lewis, 1992) and processing facial expressions (Ashwin, Baron-Cohen, Wheelwright, O'Riordan, & Bullmore, 2007). Individuals with ASD experience difficulties in face memory tasks, even when there is minimal memory demand (e.g. simultaneous versus consecutive presentation of stimuli (Scherf et al., 2008; Wallace et al., 2008)). Furthermore, individuals with ASD demonstrate deficits when visually processing a specific component of the face, namely the eyes, and often place more focus on the lower part of faces than typically developed individuals (Klin, Jones, Schultz, Volkmar, & Cohen, 2002). Perhaps as a consequence of that, they appear less able to process facial expressions (Ashwin et al., 2007). Moreover, individuals with ASD exhibit a priming effect for single face parts, which typically developed individuals do not display (Lahaie et al., 2006). Thus, individuals with ASD appear to have a bias to process faces in terms of their component parts, with the exception of eyes, rather than globally. However, this does not necessarily imply that individuals with ASD have impaired global processing strategies. Instead, it may be that people with ASD have excellent local perceptual processing abilities (Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). Alternatively, differences in perception could be due to differences in the coordination of global and local processing at a higher (cognitive/attentive) level (Happé & Frith, 2006; Iarocci, Burack, Shore, Mottron, & Enns, 2006), rather than superior or deficient abilities of the two types of processing per se.

Face processing impairments are not limited to individuals diagnosed with ASD. Individuals with autism-like traits (ALTs) display

---

* Corresponding author at: University of Canberra, Building 12, 11 Kirinari Street, Bruce VIC, 2617, Australia.
E-mail address: Jeroen.vanboxtel@canberra.edu.au (J.J.A. van Boxtel).

https://doi.org/10.1016/j.visres.2022.108071
Received 18 August 2021; Received in revised form 28 April 2022; Accepted 9 May 2022

deficits when processing faces, or other social stimuli (e.g., English, Maybery, & Visser, 2017; Halliday, MacDonald, Scherf, & Tanaka, 2014; van Boxtel, Peng, Su, & Lu, 2016). Moreover, they also experience deficits in general emotion recognition (McKenzie et al., 2018). These findings parallel the processing delays and deficits often observed in individuals with ASD (Stavropoulos, Viktorinova, Naples, Foss-Feig, & McPartland, 2018).

Nevertheless, typically-developing individuals are generally very sensitive to face stimuli, so sensitive in fact that they often perceive faces in non-face stimuli, e.g., in clouds. This effect is called pareidolia, and leads to face-activity in the cortex (Liu et al., 2014; Wardle, Taubert, Teichmann, & Baker, 2020). Studying pareidolia can reveal important aspects of face perception because it may help reveal how the brain constructs face representations, even when no actual face parts are present. Past research has shown that individuals with ASD will not spontaneously report, or orient towards, a pareidolic face as much as TD individuals (Guillon et al., 2016; Ryan, Stafford, & King, 2016), but they can see it if asked to look for them (Akechi, Kikuchi, Tojo, Osanai, & Hasegawa, 2014; Guillon et al., 2016). Interestingly, as opposed to the effects in ASD, no clear effects of ALTs have been reported in pareidolia (Verhallen et al., 2017). The distinction between spontaneously reported faces in non-face stimuli, and actively searching for (and finding) a face in non-face stimuli may be very important. The former could be seen as a form of pure pareidolia, while the latter is perhaps more a “difficult face-detection task”. We do not make that distinction in this report, but do try to highlight differences in tasks when they appear relevant.

One of the possible causes for an absence of an effect in studies looking at ALTs is that the stimuli usually employed in these studies consisted of just one face per presentation. This set-up may be insensitive to small effects of ALTs, especially if performance is near ceiling. If a difference exists in the ability to detect faces, it is more likely to be picked up in displays with an increased number of such faces. Therefore, in the current study we decided to study displays with a variable number of pareidolic faces (either all embedded within a single image, or presented simultaneously, but separately from each other), and investigate how well individuals with a varying level of ALTs can detect these faces, without knowing how many faces are presented. Pareidolia varies widely between individuals (Zhou & Meng, 2020), and is therefore more suitable to study individual differences in face perception, compared to using real face stimuli.

Face pareidolia has been investigated with different types of stimuli, which may or may not rely on different underlying mechanisms. The first one is a case where faces are perceived due to combining different objects, part of objects, or shadows together into a composite face. Examples of these types of stimuli are the fruit faces by Arcimboldo[1], or seeing a face in the clouds. We will refer to these types of stimuli as *embedded faces*, EF, in reference to the embedded figures test (Happé, 2013). The second type of stimuli are *Mooney faces* (Mooney, 1957), MF, in which photos are converted in dichromatic stimuli, by thresholding the luminance at a certain level. We will investigate the ability to detect multiple pareidolic faces for both types of stimuli.

Face processing consists of various stages, ranging from detection of the presence of a face, and recognising the age, gender or emotion of the face, to identifying the individual. These processes are probably (partly) dissociable (e.g., Robertson, Jenkins, & Burton, 2017). In this report, we will focus on detection, as it is arguably the most basic function of face processing, and, we opine, is best suited for an initial investigation of our novel approach. The non-faces we use may also not be ideal stimuli to study sex, gender, and identity perception, while this is less of a problem for the more general function of face detection.

Finally, one well-established effect in face perception is the inversion effect. The face inversion effect is an archetypal face processing task where inverted faces are characteristically more difficult to perceive than upright faces (Farah, Tanaka, & Drain, 1995). Interestingly, this inversion effect is larger for faces, but it does occur for object categories (Yin, 1969). Research has shown that individuals with ASD display a typical face inversion effect for normal faces (Lahaie et al., 2006; Scherf et al., 2008). Although an inversion effect has been shown for Mooney faces (George, Jemel, Fiori, & Renault, 1997; Kanwisher, Tong, & Nakayama, 1998; Schwiedrzik, Melloni, & Schurger, 2018), and embedded faces (Pavlova, Romagnano, Fallgatter, & Sokolov, 2020), we have found no research looking at whether the Mooney-face inversion effect depends on ALTs. Two prior studies that looked at the dependence of inversion effects of real face images on ALTs, found that it was weaker in individuals with more ALTs (Laycock, Wood, Wright, Crewther, & Goodale, 2019; Wyer, Martin, Pickup, & Macrae, 2012).

The current study hypothesised that there is a negative relationship between ALT and the ability to detect faces, such that individuals with high scores of ALTs will detect fewer faces than those with low scores of ALTs, especially for trials with many faces. It was also hypothesised that the face inversion effect will be demonstrated by participants in the current study, such that participants will make more errors when stimuli are presented in an inverted fashion compared to when presented in an upright fashion. We also investigated whether the inversion effect was larger in people with higher levels of ALTs.

2. Method

2.1. Participants

An a priori power analysis using G*Power version 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted to determine the minimum sample sizes to test the study hypotheses. Results indicated the required sample sizes to achieve 95% power for detecting medium effects, at a significance criterion of $\alpha = 0.05$, were $N = 105$ for a multiple regression with one predictor. We collected data from 263 individuals. After removing incomplete datasets, 224 participants remained ($M_{age} = 44.74$, $SD_{age} = 17.15$, range: 18–82; 16 participants did not provide their age; 84 males, 140 females). Convenience sampling was used to recruit participants via two means: (1) a study advertisement was placed in the University of Canberra’s first-year psychology participant pool. Students obtained credit for their unit by participation in the study; and (2) a community sample was recruited via study advertisements posted on Facebook and Qualtrics, targeting adults in Australia. These participants had the opportunity to win one of three $50 gift vouchers for their participation in the study.

The inclusion criteria for the current study were: (1) aged over 18 years; (2) current residence in Australia; and (3) no significant visual impairments. There were no study exclusion criteria.

The study advertisements invited participants to participate in an anonymous online survey hosted by Qualtrics, Version 2020 (Qualtrics, Provo, UT, USA), which took approximately 45 min to complete. All participants provided informed consent before proceeding with the study. The study was approved by the University of Canberra’s Human Research Ethics Committee (HREC–1968).

2.2. Materials

The study utilized an online questionnaire that consisted of, first, demographic information (age, gender), then two face detection tasks, and then the autism quotient (AQ) questionnaire. Autism spectrum traits were measured using the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The AQ is a self-administered instrument that measures the degree to which an adult with normal intelligence has autism-like traits (ALTs).

Face detection accuracy. Face detection accuracy was measured using two different tasks. One of the tasks involved finding and counting the number of embedded faces in a range of images. Fig. 1a contains an

---

[1] See examples on: https://www.wikiart.org/en/giuseppe-arcimboldo.

Fig. 1. Example items. (a) Example from the embedded face task, Violettes du 20 Mars 1815, by Jean-Dominique-Étienne Canu, with three embedded faces (top-left, top-right, and middle of the image). (b) Example item from the Mooney face task. The Mooney faces were taken from the dataset generated by Schwiedrzik et al. (2018). They were arranged in a 3 × 3 grid. In this upright example, 2 faces are present: one in the top left, and one in the middle right.

example item; there were 25 items. The test images were predominantly taken from various sources on the Internet, with remaining images developed using photo editing software. Each item image contained between zero to 13 human faces. The other task involved counting the number of Mooney faces in a 3x3 grid of Mooney stimuli. There could be 0 to 9 faces. The Mooney faces were taken from Schwiedrzik et al. (2018). There were again 25 items, 8 of which inverted. Fig. 1b contains an example item from the MF task. As this task was conducted online, we were unable to control the visual presentation of the images (e.g. size, brightness).

Linear mixed models were run using Matlab R2019a. The associated figures were created using Jamovi 1.1.19.0.

2.3. Procedure

The survey was open between July and October 2019. Participants who consented to participate were first asked to provide demographic information. They were then presented with the EF and MF tasks. The EF and MF tasks were counterbalanced such that half of the participants commenced with the EF task and the other half with the MF task. The order of the images within each task were randomised to minimise order effects. The images from each task were displayed until the participant answered the question; response time was not recorded. Participants reported the number of human faces they could see in each image before proceeding to the next image. In the EF task the number was keyed in, and in the MF task it was picked from an ordered list of radio-buttons (0–9). A practice trial with one example image from each task was run prior to completion of the actual face detection task. Following the completion of the EF and MF tasks, participants were asked to complete the AQ, which was also presented in Qualtrics.

2.4. Data screening

Prior to analysis, raw data was screened to identify unusual patterns of responses, duplications, missing data, and univariate outliers. Of the 263 participants who consented to participate in the study, one case had no recorded data, and was removed. A further 38 participants had missing values across all study variables and were subsequently removed from the dataset. Univariate outliers were identified by transforming raw scores of each item into standardized z-scores. Further, investigation found 35 cases in which individual responses to an item had z-scores > 3.29 ($p < .001$). These scores were identified as univariate outliers and the item was excluded from the analysis for that individual. We observed that the Mooney face task showed several individuals who responded with the same answer on most trials. We arbitrarily set a limit to 10 identical answers out of the 25 stimuli. Individuals with > 10 identical answers were removed from the analysis. This removed 12 individuals from the Mooney task, most of these individuals reported 9 faces on most trials.

2.5. Item selection for the EF and MF tasks

An item analysis was conducted on both the EF and MF tasks. Pursuant to the recommendations of Cohen, Swerdlik, and Phillips (1996) and Carpenter, Balsis, Otilingam, Hanson, and Gatz (2009), item analysis consisted of examining item difficulty indices, item discrimination indices, and item reliability. A more lenient acceptable range was adopted for the indexes due to the expected difficulty in scoring 100% in the face detection tasks. For instance, if there are nine faces in an image and a participant reports eight, then it will be an incorrect answer, however, it is quite close to the correct answer. Therefore, we chose to broaden the acceptance criteria, which we decided before analysing the data.

Item difficulty index. The item difficulty index represents the proportion of the total number of test takers who answered an item correctly (Carpenter et al., 2009). The lower the item difficulty index, the harder the item and vice versa (Cohen et al., 1996). Items answered correctly or incorrectly by a high percentage of people are unlikely to discriminate among test takers and are therefore candidates for deletion. A difficulty index of 0.95 indicates that most people answered the item correctly (i.e., 95% of the sample), and thus the item provides little useful discriminative power. We considered scores between 0.15 and 0.85 acceptable item difficulty indices. EF items had a difficulty index between 0.0045 and 0.9911, and 10 items were within the acceptable range. MF items had a difficulty index between 0.3393 and 0.8973, and 23 items were within the acceptable range. Items that were not within range were excluded from our analysis.

Item discrimination index. The item discrimination index was used to identify items that were the least effective at differentiating between the high and low scorers on the EF and MF tasks. To calculate the item discrimination indices, high scorers (top 27%) and low scorers (bottom 27%) on each of the 25 items of the EF and MF tasks were identified. Then, the percentage of participants in each group (i.e., high and low scorers) with correct responses was calculated for each item (Carpenter et al., 2009). The low scorers' percentage was subtracted from the high scorers' percentage to calculate a discrimination index for each item. A positive difference of 0.25 and above was used to indicate adequate item discrimination in the present analysis. A total of 12 items on the EF task and 21 items on MF task met this criterion. Fig. 2 displays a graph of the items that were retained with item difficulty on one axis and discriminability on the other axis.

Item reliability. All items, including those deemed problematic from the item difficulty and discrimination analysis, were further examined for internal consistency using Cronbach's alpha. Items were excluded if there was a negative correlation with the overall test score and alpha improved with removal. No items needed to be removed based on item reliability, indicating good internal consistency.

In total, 9 items from EF task, and 21 items from MF task (of which 6 were inverted) were retained because they met acceptable criteria for item difficulty index, item discrimination index and item reliability scores.

2.6. *Cross validation analysis*

The main analysis was done on all data. To gauge the robustness of the analysis, we performed a cross-validation analysis, whereby item selection (based on the item analyses) and the subsequent statistical testing were performed using different participants.

For the cross-validation, we split the data into two equal groups: half of the participants were used to do the item analysis, and the other half were used to perform the statistical tests on (i.e., the test group). This random selection, and subsequent analysis, were repeated 100 times (with different random assignments to item analysis and test groups), creating 100 different item analyses, and 100 statistical analyses using those different item analyses. The main interest is whether the cross-validated statistical analyses return, overall, similar results to the main analysis. This is presented in the results section.

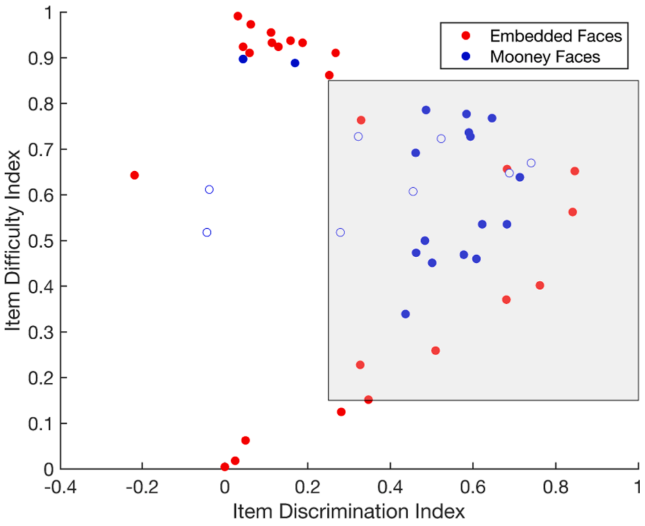

Fig. 2. Items from the Embedded Face task and Mooney Face task plotted with the Item Difficulty Index (y-axis) and Item Discrimination Index (x-axis). The shaded area represents the items within the acceptable range, and those that were retained in the final scales. Open symbols represent inverted items.

However, we also obtain an idea of how robust the item selection itself is. The robustness of the item selection is presented in Fig. 3. Overall, most items that were selected based on the complete dataset, were also selected in the vast majority of the cross-validations, even though the number of participants was only half that of the overall analysis. This effect was more robust for the MF dataset, than for the EF dataset.

For each statistical analysis during the cross-validation we fitted many linear mixed models to the data of the participants in the test group, including almost all possible combinations of main effects and interactions. To reduce the total number of models that were compared, we included a subset of all possible factors and interactions. This selection was based on the fit to the complete dataset, and included the factors and interactions that were present in $> 20\%$ of well-fitting models (operationally defined as models that were within 2 AIC points from the best model). Models with all possible combinations of these factors and interactions were fitted to the data in each cross-validation (i.e., 64 models per cross-validation in the EF set, and 512 models in the MF set), and compared using the AIC measure. We kept the model with the lowest AIC for each cross-validation.

3. Results

3.1. *Embedded face task*

The reported number of faces depended on the presented number of faces in the embedded face task (Fig. 4). To assess the influence of various factors, we performed a massive linear mixed-models comparison. The dependent variable was the reported number of faces (RepFaces), and the factors were the presented number of pareidolic faces (CorrectFaces; coded as continuous), age, gender, and AQ, as well as most interactions. To reduce the number of possible models, we only included two-way interactions that included the term CorrectFaces, but we did include all three-way and four-way interactions. A random intercept and a random slope (dependence on the CorrectFaces) were included for each subject (ID). Models were compared with the Aikaike Information Criterion (AIC). The best model was RepFaces $\sim$ 1 + CorrectFaces*Age + (1 + CorrectFaces |ID), however, this model was singular. We therefore decided to test models without the random intercepts, but keeping the random slopes. This is warranted, because in our tasks it makes sense that at zero presented faces, people report seeing zero faces. The best model was very similar RepFaces $\sim$ -1 + CorrectFaces *Age + (CorrectFaces $-$ 1|ID). This model's AIC was even lower than the singular model, and a significant effect of presented number of pareidolic faces ($t(1863) = 14.11$, $p < 0.00001$) and age ($t(1863) = -8.32$, $p < 0.00001$), and a significant interaction ($t(1863) = 6.43$, $p < 0.00001$). The higher the age, the lower the reported number of faces. Adding gender as a factor to the last model (without interactions) showed a non-significant log-likelihood ratio test ($X^2(1) = 0.093$, $p = 0.76$), and an AIC score 2 units higher, suggesting moderate evidence for the model without the factor of gender. Including interactions resulted in the same qualitative pattern (but an AIC 7 points higher). Including AQ with or without interactions resulted in similarly worse models compared to the best model.

The subsequent cross-validation analysis (see methods) revealed that in 65% of the cross-validations the best model included the main effects of Age, CorrectFaces, *and* their interaction (as in the model discussed above). When this was not the case, the interaction was missing. All models included age and CorrectFaces as main effects. Only, 52% of models include AQ, and 14% of models include Gender (as main effect *or* interaction, but rarely both). This analysis suggests that our analysis based on the whole data set is robust, and not dependent on using the same participants in the item selection, and the statistical analysis.

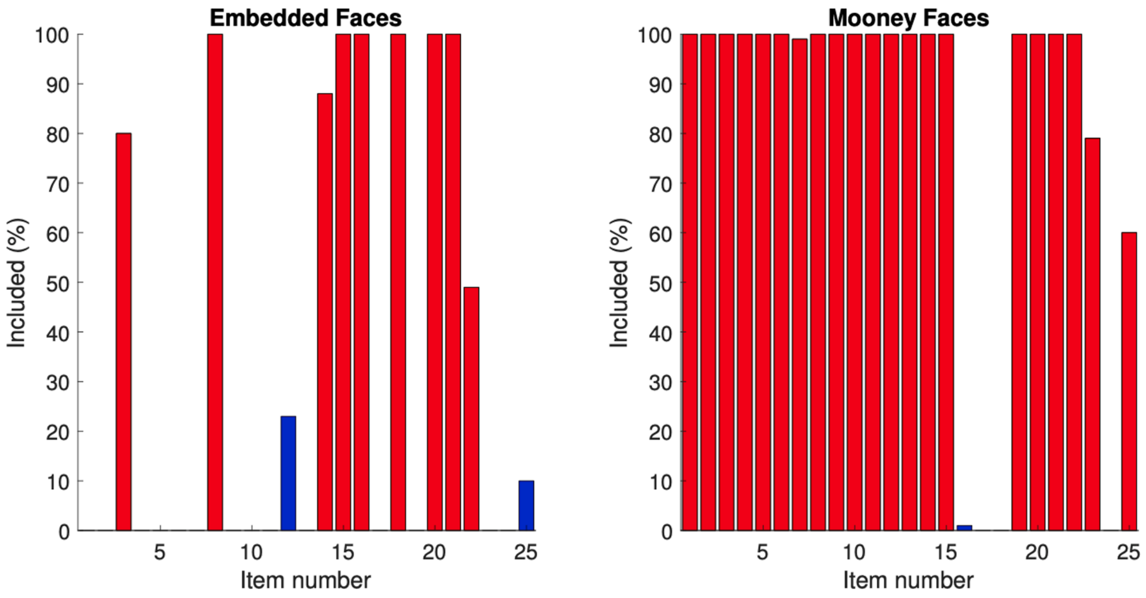

Fig. 3. Cross-validation item selection results for the embedded faces, and Mooney faces. The plots show the percentage of cross-validations in which each item was included. Red bars are those that were included when item selection was done on the whole data set, blue bars are items that were not included when analysing the whole dataset.

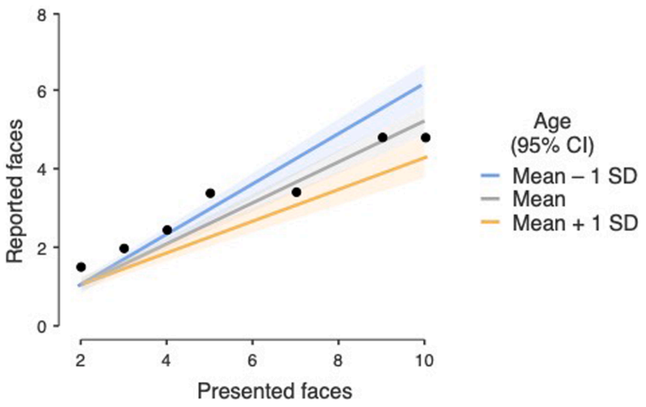

Fig. 4. Plot of mean number of reported faces, dependent on presented faces, in the Embedded Face task, including the dependence on the covariate age. The data points are the mean item scores over individuals. Note that regressions were not made directly through these data, but through individual items, per individual.

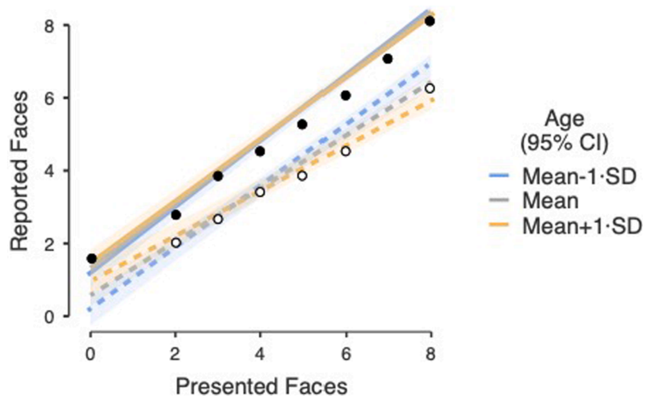

Fig. 5. Plot of mean number of reported faces, dependent on presented faces, in the Mooney Face task. The dependence on the covariate age is also shown. Dashed lines, and open circles are for inverted items. The data points are the mean item scores over individuals. Note that regressions were not made directly through these data, but through individual items, per individual.

3.2. Mooney face task

The reported number of faces depended on the presented number of faces in the Mooney face task (Fig. 5). As in the embedded face task, we performed a massive comparison of linear mixed models with the factors presented number of pareidolic faces (CorrectFaces; coded as continuous), age, gender, AQ, and inversion were included as factors/covariates. A random intercept and a random slope (dependence on the number of presented faces), and a random effect of the inversion (Inversion) were included for each subject (ID). To decrease the total number of possible interactions, we only included terms up to (and including) three-way interactions. The best model was RepFaces ~ 1 + Age* CorrectFaces + Age*Inverted + Gender:Inverted + Age: CorrectFaces:Inverted + Gender: CorrectFaces:Inverted + (1 + CorrectFaces + Inverted | ID). We confirmed that, indeed, the addition of Inversion as a random factor was supported, by performing a log-likelihood ratio test comparing the models with and without Inversion, using the restricted maximum likelihood (REML) fit method ($X^2(3) = 141.52$, $p < .00001$). This best model showed a significant effect of presented number of pareidolic faces ($t(4086) = 24.9$, $p < .00001$), and Inversion ($t(4086) = 4.31$, $p < .00001$), but not Age ($t(4086) = -1.76$, $p = .078$). However, age had a significant interaction with CorrectFaces ($t(4086) = -2.01$, $p = .044$), with Inversion ($t(4086) = 2.21$, $p = .012$), and a three-way interaction with CorrectFaces, and Inversion ($t(4086) = 8.62$, $p < .00001$). The effect of age was stronger on the inverted displays than on the upright displays, and only evident when CorrectFaces was small. Gender and Inversion had a significant interaction, $t(4086) = 3.59$, $p < .0005$. There was also a significant three-way interaction between gender, CorrectFaces, and Inversion, $t(4086) = 3.06$, $p = .0022$.

In the cross validation, we found that the model described above was the best model in 10% of the cross-validations, but was within 2 AIC points in 44% of cross-validations. Another model was best in 26% of cross-validations, and was identical to the above model, but excluded the main effect of age, and the interaction between age and CorrectFaces. This model was within 2AIC points of the beset model in 58% of cross validations.

In terms of the relative impact of age, gender and AQ, 100% of models had age as main effect or interaction (32% as a main effect), 40% of models had AQ as either a main effect of interaction, and 87% of models had Gender as an interaction. This suggests that, as in the EF dataset, AQ did not appear to be quite as influential as age. Gender did appear to be influential in the MF dataset, while it was not in the EF dataset.

3.3. The effect of age on accuracy

The previous analyses indicated that there was a significant effect of age on pareidolia. The effects were not anticipated, and we therefore performed an additional post-hoc analysis. We looked at the overall accuracy of face detection, depending on age. A response was considered accurate if the response was within ± 1 from the actual presented number of faces. We then calculated the overall accuracy over all items, per individual. Fig. 6 shows the mean accuracy dependent on age for both the embedded face task, and the Mooney face task. This relationship was significant for both embedded faces ($b = -0.22$, $t(206) = -4.85$, $p < .001$, $R^2 = 0.10$), and Mooney faces ($b = -0.17$, $t(194) = -2.32$, $p = .02$, $R^2 = 0.027$). For the Mooney face data, this relationship was mostly carried by the inverted displays (even though they were fewer in number), $b = -0.33$, $t(194) = -3.45$, $p = .0007$, $R^2 = 0.06$, and not the upright displays, $b = -0.010$, $t(194) = -1.09$, $p = .28$, $R^2 = 0.006$.

A previous report analysed Mooney face detection and categorisation accuracy dependent on age group (Carbon, Gruter, & Gruter, 2013). To make a visual comparison possible with this previous report, we display our data per age category (15–24 y, 25–34 y, etc.) as well. There appears to be a gradual decrease in accuracy dependent on age for the EF task. The MF task shows evidence of an increase in accuracy in early adulthood, followed by a plateau, and a subsequent decrease in accuracy after about 65 years of age. This pattern was more pronounced in the inverted data.

4. Discussion

The current study aimed to investigate whether face perception in non-face stimuli (pareidolia) differs amongst individuals with low and high ALTs. We found no such dependence in either the embedded face task or the Mooney face task. The inversion effect was also not stronger in individuals with high levels of ALT. We did find a dependence on age, and complex interactions with gender for the Mooney face task.

4.1. Face pareidolia in ASD, and individuals with varying levels of ALTs

Face perception appears to be decreased in ASD (Weigelt, Koldewyn, & Kanwisher, 2012). Consistent with this observation, face pareidolia is also reported to be affected in ASD. For example, children with ASD are less likely to report pareidolia when not prompted to search for a face (Ryan et al., 2016), and individuals without ASD have a stronger preference to look at upright faces compared to inverted faces than individuals with ASD, whether the faces are real (Pelphrey et al., 2002) or face-like objects (i.e., pareidolia; Guillon et al., 2016).

Pareidolia in Mooney faces is decreased in ASD in some reports (e.g., Naumann, Senftleben, Santhosh, McPartland, & Webb, 2018; Sun et al., 2012), but not others (Tavares, Mouga, Oliveira, & Castelo-Branco, 2016). Similar equivocal effects have been reported for embedded face tasks (Akechi et al., 2014). In the typically-developing population, increased ALTs have been associated with decreased face processing by some (Halliday et al., 2014; Stavropoulos et al., 2018). However, our results do not indicate pareidolia depends on the level of ALTs, which is consistent with other previous reports (e.g., Van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2018; Verhallen et al., 2014).

How to explain these differences? The impairment for face perception in ASD appears to be mostly limited to certain tasks, e.g., involving memory, and emotion (Weigelt et al., 2012). Alternatively, task differences could exist between recognition, detection and categorization tasks (Carbon et al., 2013). Indeed, task differences have been shown to modulate the impact of ALTs in other social stimuli, such as biological motion tasks (van Boxtel et al., 2016).

Overall the literature appears to show quite strong effects of ASD on face processing, but this is much less so for ALTs. It is known within the ASD literature that individuals with high functioning ASD do show less impairments in face perception than other individuals with ASD (Weigelt et al., 2012). This suggests that any differences in face processing dependent on ALTs may be too small to be detected in most experiments, because even individuals with ASD who are high-functioning show less impairment. Even though our experimental design was set up to be more sensitive than previous designs (by showing more than one face at a time), we did not find a significant impact of ALTs.

Possibly, our design was still too insensitive, because individuals could take as much time as they wanted on each trial. One possible approach to make tasks more sensitive is using shorter presentation durations, because higher levels of ALT have been associated with inefficient face perception, specifically in terms of processing delays

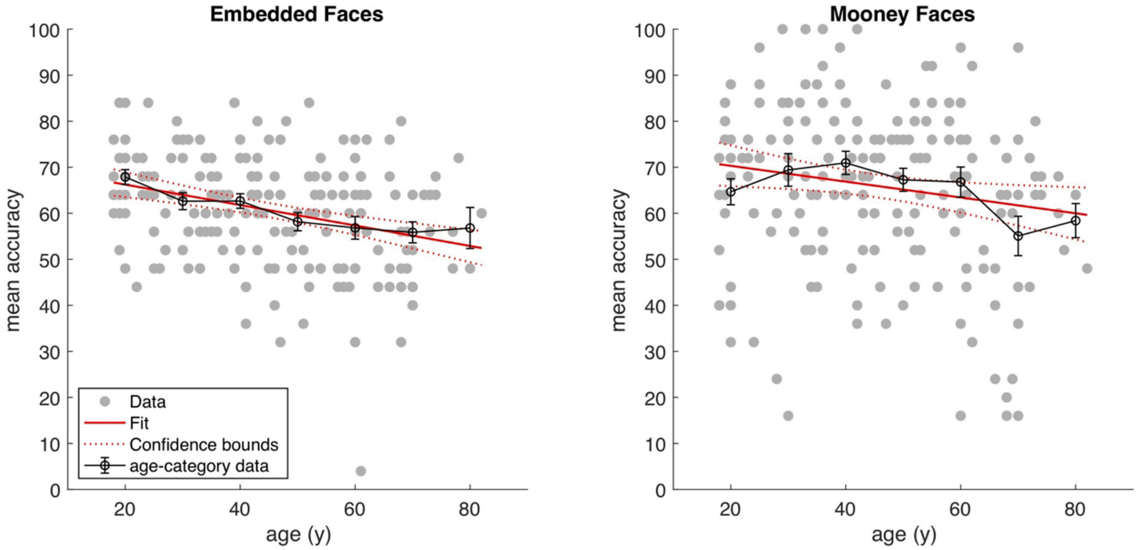

Fig. 6. Accuracy on the embedded face task (left) and Mooney face task (right), dependent on age. Regression lines and confidence bounds are drawn in red, and mean and s.e.m. and drawn age categories (15–24, 25–34, etc).

(Stavropoulos et al., 2018; Van der Hallen, Evers, Brewaeys, Van den Noortgate, & Wagemans, 2015)[2].

4.2. Inversion effect and ALT

The current study found a significant difference in participants' processing of inverted and upright faces, with performance being superior when detecting upright faces. This is a well-known inversion effect for faces (Faja et al., 2009; Lahaie et al., 2006; Scherf et al., 2008), but here shown for pareidolia (see also, George et al., 1997; Kanwisher et al., 1998; Schwiedrzik et al., 2018).

Past research looking at the inversion effect dependent on ALTs, showed that increased ALTs are associated with decreased inversion effects (Laycock et al., 2019). Our results are not consistent with this finding. However, Laycock et al. (2019) only found the significant decrease in an eye-movement measure, not in behavioral data, suggesting that the dependent measure may determine whether a modulating effect of ALTs will be found.

We did find interactions between the inversion effect and age, suggesting that the inversion effect is different at different ages. A similar finding, but with real (not Mooney) faces, was reported by Germine, Duchaine, and Nakayama (2011). We will discuss the age effects in more detail below.

4.3. Age-dependent face perception

Our results show a decrease in face detection with increased age of the observer. This is consistent with previous investigations (Boutet, Taler, & Collin, 2015; Carbon et al., 2013; Havard & Memon, 2009; Lamont, Stewart-Williams, & Podd, 2005), although those studies mostly focused on face recognition. In a large-scale study with over 60,000 individuals, it was shown that optimal performance was reached by people in their early thirties, and both younger and older individuals showed decreased performance (Germine et al., 2011).

There is much less research on tasks other than recognition. Face categorization tasks (for gender, and age) also show an age dependence (Carbon et al., 2013), showing an increase in the first 15 years of life, followed by a long plateau, and then a decrease after about 65 years (which reaches significance only in the oldest age group, 81–88 years of age). The same study found that face detection also showed a dependence on age, but with a decrease in performance setting in later, around 75 years in age (but not reaching significance for any age-group comparison). The study used Mooney faces, consistent with our approach. Our results are largely consistent with these findings. However, we do find a significant decrease in the Mooney face detection task. Additionally, we find a significant age effect in the embedded face detection task. In our analyses, age was a continuous variable, and we used regression analyses. To qualitatively compare our results to the results of Carbon et al. (2013), we visualised the data based on age groups (see Fig. 6). Consistent with Carbon et al. (2013), our Mooney face task showed an increase in detection accuracy early in life, a subsequent long plateau, and a decrease later in life. However, the embedded face task did not show this pattern. Instead accuracy in the embedded face task appears to decrease gradually.

Face processing depends on several processes, which may depend on age in different ways (Boutet et al., 2015). The difference in the modulating effects of age in our two face detection tasks suggests that face detection is supported by different mechanisms in both cases. For example, Mooney face perception is thought to depend on the perception of closure (Mooney, 1957), and the perception of closure is inferior in elderly individuals (Basowitz & Korchin, 1957). The perception of embedded faces may depend on the balance between perceiving a face or the objects that make up the face. There may therefore be an influence of perceptual inhibition involved, which is decreased in the elderly (Jennings, Mendelson, Redfern, & Nebes, 2011). Other potential differences involve the dependence on featural versus configural information, the processing of which changes with age (Mondloch, Le Grand, & Maurer, 2002; Murray, Halberstadt, & Ruffman, 2010).

Although we cannot determine, based on our data, what the differences in face processing are that underlie the different age dependencies in our two tasks, our study does suggest that they differ in their dependence on age (in a cross-sectional study design). Most studies investigate age dependence by comparing a young and an old group. While useful to show a general effect of age, the advantage of using a broad range of ages, as in our report, is that it can reveal these different trajectories dependent on age.

4.4. Gender-dependent face perception

We did not anticipate gender to have a strong impact on our pareidolia findings. Past findings have been somewhat equivocal, in that some research found that females are better than males in "fruit faces" (Pavlova, Scheffler, & Sokolov, 2015), real faces (Goldstein & Chance, 1971; Lewin & Herlitz, 2002; Loven, Svard, Ebner, Herlitz, & Fischer, 2014), and embedded faces (Pavlova et al., 2020; but only for inverted faces), while other work showed that males are better at Mooney faces (Foreman, 1991; Silverstein et al., 2021; Verhallen et al., 2014). Our results are to some degree consistent with that. We found no significant gender effect in the EF dataset, and found that males did better in the MF dataset. However, the gender effect in the MF dataset was quite complex, with males scoring slightly higher (being more accurate) than females on average, but only for inverted faces, and when the actual number of faces was low, explaining both the significant three-way and two-way interaction. Given that we only found significant gender differences for inverted faces (similar to Pavlova et al., 2020 for embedded faces), it is likely that previously-reported gender findings may be stronger when faces are inverted.

The literature as a whole appears to show opposite gender effects for real faces (where females outperform males), and Mooney faces (where males outperform females). This suggests that the Mooney face task and normal face recognition (and possibly also EF recognition) rely on partly different mechanisms. These mechanisms need not be face-specific, and could rely on differences in general configural information processing, as suggested above, or figure-ground segregation.

5. Conclusion

In contrast to our predictions and past research, we did not find that levels of ALTs predicted face detection abilities. In addition, the present study found that, while participants were overall superior at detecting upright faces compared to inverted faces, this face inversion effect was not more pronounced among individuals with high (vs. low) levels of ALTs. One potential explanation for these findings is that individuals with ALT do not have deficits in face detection, although it is possible that our untimed trials allowed sufficient time for individuals with slower face processing capabilities to achieve typical levels of performance.

Our findings did show dependencies on age and, to some degree, gender. Older individuals are less accurate at the face detection tasks for pareidolic faces. Our findings suggest a different dependent on age for our two detection tasks, which we tentatively link to their different dependencies on different perceptual processes (e.g. closure perception, perceptual inhibition), which are possible not face-specific, but do

---

[2] We did not record response times to individual items, but we do have access to the overall duration of the whole experiment (which includes the AQ questionnaire). Therefore, we can separate fast from slow responders. Overall, the models of the 10% fastest responders (and 50% fastest responders) were very similar to the models reported in this report. The main difference being that for the MF task (but not for the EF) there is a 3-way interaction between CorrectFaces, AQ, and Inversion.

influence face processing.

CRediT authorship contribution statement

Muhammad Rahman: Conceptualization, Data curation, Writing – original draft. **Jeroen J.A. van Boxtel:** Conceptualization, Methodology, Writing – review & editing, Software, Visualization, Supervision.

References

Akechi, H., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2014). Neural and behavioural responses to face-likeness of objects in adolescents with autism spectrum disorder. *Scientific Reports, 4*(1), 1–7.

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., & Bullmore, E. T. (2007). Differential activation of the amygdala and the 'social brain' during fearful face-processing in Asperger Syndrome. *Neuropsychologia, 45*(1), 2–14. https://doi.org/10.1016/j.neuropsychologia.2006.04.014

Basowitz, H., & Korchin, S. J. (1957). Age differences in the perception of closure. *The Journal of Abnormal and Social Psychology, 54*(1), 93.

Blair, R. J., Frith, U., Smith, N., Abell, F., & Cipolotti, L. (2002). Fractionation of visual memory: Agency detection and its impairment in autism. *Neuropsychologia, 40*(1), 108–118. https://doi.org/10.1016/s0028-3932(01)00069-0

Boucher, J., & Lewis, V. (1992). Unfamiliar face recognition in relatively able autistic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*(5), 843–859. https://doi.org/10.1111/j.1469-7610.1992.tb01960.x

Boutet, I., Taler, V., & Collin, C. A. (2015). On the particular vulnerability of face recognition to aging: A review of three hypotheses. *Frontiers in Psychology, 6*, 1139.

Carbon, C. C., Gruter, M., & Gruter, T. (2013). Age-dependent face detection and face categorization performance. *PloS One, 8*(10), Article e79164. https://doi.org/10.1371/journal.pone.0079164

Carpenter, B. D., Balsis, S., Otilingam, P. G., Hanson, P. K., & Gatz, M. (2009). The Alzheimer's Disease Knowledge Scale: Development and psychometric properties. *The Gerontologist, 49*(2), 236–247.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.

English, M. C., Maybery, M. T., & Visser, T. A. (2017). Threatening faces fail to guide attention for adults with autistic-like traits. *Autism Research, 10*(2), 311–320. https://doi.org/10.1002/aur.1658

Faja, S., Webb, S. J., Merkle, K., Aylward, E., & Dawson, G. (2009). Brief report: Face configuration accuracy and processing speed among adults with high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(3), 532–538.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Foreman, N. (1991). Correlates of performance on the Gollin and Mooney tests of visual closure. *Journal of General Psychology, 118*(1), 13–20. https://doi.org/10.1080/00221309.1991.9711129

George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: A spatio-temporal ERP study. *Neuroreport, 8*(6), 1417–1423. https://doi.org/10.1097/00001756-199704140-00019

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201–210.

Goldstein, A. G., & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Perception and Psychophysics, 9*(2), 237–241.

Guillon, Q., Rogé, B., Afzali, M. H., Baduel, S., Kruck, J., & Hadjikhani, N. (2016). Intact perception but abnormal orientation towards face-like objects in young children with ASD. *Scientific Reports, 6*, 22119.

Halliday, D. W., MacDonald, S. W., Scherf, K. S., & Tanaka, J. W. (2014). A reciprocal model of face recognition and autistic traits: Evidence from an individual differences perspective. *PloS One, 9*(5), Article e94013. https://doi.org/10.1371/journal.pone.0094013

Happé, F. (2013). *Embedded figures test (EFT). Encyclopedia of autism spectrum disorders*. New York: Springer.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Havard, C., & Memon, A. (2009). The influence of face age on identification from a video line-up: A comparison between older and younger adults. *Memory, 17*(8), 847–859.

Iarocci, G., Burack, J. A., Shore, D. I., Mottron, L., & Enns, J. T. (2006). Global-local visual processing in high functioning children with autism: Structural vs. implicit task biases. *Journal of Autism and Developmental Disorders, 36*(1), 117–129. https://doi.org/10.1007/s10803-005-0045-2

Jennings, J. R., Mendelson, D. N., Redfern, M. S., & Nebes, R. D. (2011). Detecting age differences in inhibition processes with a test of perceptual and motor inhibition. *Experimental Aging Research, 37*(2), 179.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*(1), B1–B11. https://doi.org/10.1016/s0010-0277(98)00035-3

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809–816.

Lahaie, A., Mottron, L., Arguin, M., Berthiaume, C., Jemel, B., & Saumier, D. (2006). Face perception in high-functioning autistic adults: Evidence for superior processing of face parts, not for a configural face-processing deficit. *Neuropsychology, 20*(1), 30.

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

Laycock, R., Wood, K., Wright, A., Crewther, S. G., & Goodale, M. A. (2019). Saccade Latency Provides Evidence for Reduced Face Inversion Effects With Higher Autism Traits. *Frontiers in Human Neuroscience, 13*, 470. https://doi.org/10.3389/fnhum.2019.00470

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition–women's faces make the difference. *Brain and Cognition, 50*(1), 121–128. https://doi.org/10.1016/s0278-2626(02)00016-7

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex, 53*, 60–77.

Loven, J., Svard, J., Ebner, N. C., Herlitz, A., & Fischer, H. (2014). Face gender modulates women's brain activity during face encoding. *Social Cognitive and Affective Neuroscience, 9*(7), 1000–1005. https://doi.org/10.1093/scan/nst073

McKenzie, K., Murray, A. L., Wilkinson, A., Murray, G. C., Metcalfe, D., O'Donnell, M., & McCarty, K. (2018). The relations between processing style, autistic-like traits, and emotion recognition in individuals with and without Autism Spectrum Disorder. *Personality and Individual Differences, 120*, 1–6.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*(5), 553–566.

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue canadienne de psychologie, 11*(4), 219.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Murray, J. E., Halberstadt, J., & Ruffman, T. (2010). The face of aging: Sensitivity to facial feature relations changes with age. *Psychology and Aging, 25*(4), 846.

Naumann, S., Senftleben, U., Santhosh, M., McPartland, J., & Webb, S. J. (2018). Neurophysiological correlates of holistic face processing in adolescents with and without autism spectrum disorder. *Journal of Neurodevelopmental Disorders, 10*(1), 27.

Pavlova, M. A., Romagnano, V., Fallgatter, A. J., & Sokolov, A. N. (2020). Face pareidolia in the brain: Impact of gender and orientation. *PloS One, 15*(12), Article e0244516. https://doi.org/10.1371/journal.pone.0244516

Pavlova, M. A., Scheffler, K., & Sokolov, A. N. (2015). Face-n-Food: Gender Differences in Tuning to Faces. *PloS One, 10*(7), Article e0130363. https://doi.org/10.1371/journal.pone.0130363

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*(4), 249–261. https://doi.org/10.1023/a:1016374617369

Robertson, D. J., Jenkins, R., & Burton, A. M. (2017). Face detection dissociates from face identification. *Visual Cognition, 25*(7–8), 740–748.

Ryan, C., Stafford, M., & King, R. J. (2016). Brief Report: Seeing the man in the moon: Do children with autism perceive pareidolic faces? A pilot study. *Journal of Autism and Developmental Disorders, 46*(12), 3838–3843.

Scherf, K. S., Behrmann, M., Minshew, N., & Luna, B. (2008). Atypical development of face and greeble recognition in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49*(8), 838–847. https://doi.org/10.1111/j.1469-7610.2008.01903.x

Schwiedrzik, C. M., Melloni, L., & Schurger, A. (2018). Mooney face stimuli for visual perception research. *PloS One, 13*(7), Article e0200106.

Silverstein, S. M., Thompson, J. L., Gold, J. M., Schiffman, J., Waltz, J. A., Williams, T. F., … Strauss, G. P. (2021). Increased face detection responses on the mooney faces test in people at clinical high risk for psychosis. *npj Schizophrenia, 7*(1), 1–7.

Stavropoulos, K. K., Viktorinova, M., Naples, A., Foss-Feig, J., & McPartland, J. C. (2018). Autistic traits modulate conscious and nonconscious face perception. *Social Neuroscience, 13*(1), 40–51.

Sun, L., Grutzner, C., Bolte, S., Wibral, M., Tozman, T., Schlitt, S., … Uhlhaas, P. J. (2012). Impaired gamma-band activity during perceptual organization in adults with autism spectrum disorders: Evidence for dysfunctional network activity in frontal-posterior cortices. *Journal of Neuroscience, 32*(28), 9563–9573. https://doi.org/10.1523/JNEUROSCI.1073-12.2012

Tavares, P. P., Mouga, S. S., Oliveira, G. G., & Castelo-Branco, M. (2016). Preserved face inversion effects in adults with autism spectrum disorder: An event-related potential study. *Neuroreport, 27*(8), 587–592. https://doi.org/10.1097/WNR.0000000000000576

van Boxtel, J. J., Peng, Y., Su, J., & Lu, H. (2016). Individual differences in high-level biological motion tasks correlate with autistic traits. *Vision Research, 141*, 136–144. https://doi.org/10.1016/j.visres.2016.11.005

Van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The use of prior knowledge for perceptual inference is preserved in ASD. *Clinical psychological science, 6*(3), 382–393.

Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., & Wagemans, J. (2015). Global processing takes time: A meta-analysis on local-global visual processing in ASD. *Psychological Bulletin, 141*(3), 549–573. https://doi.org/10.1037/bul0000004

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. (2014). An online version of the Mooney Face Test: Phenotypic and genetic associations. *Neuropsychologia, 63*, 19–25.

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. (2017). General and specific factors in the processing of faces. *Vision Research, 141*, 217–227.

Wallace, S., Coleman, M., & Bailey, A. (2008). Face and object processing in autism spectrum disorders. *Autism Research, 1*(1), 43–51. https://doi.org/10.1002/aur.7

Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I. (2020). Rapid and dynamic processing of face pareidolia in the human brain. *Nature communications, 11*(1), 1–14.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews, 36*(3), 1060–1084. https://doi.org/10.1016/j.neubiorev.2011.12.008

Wyer, N. A., Martin, D., Pickup, T., & Macrae, C. N. (2012). Individual differences in (non-visual) processing style predict the face inversion effect. *Cogn Sci, 36*(2), 373–384. https://doi.org/10.1111/j.1551-6709.2011.01224.x

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.

Zhou, L.-F., & Meng, M. (2020). Do you see the “face”? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology, 14*, Article e2.