

“Reading the Mind in the Eyes” in Autistic Adults is Modulated by Valence and Difficulty: An InFoR Study

Matias Baltazar , Marie-Maude Geoffray, Christopher Chatham, Manuel Bouvard, Axelle Martinez Teruel, David Monnet, Isabelle Scheid, Eleonora Murzi, Sandrine Couffin-Cadiergues, Daniel Umbricht, Lorraine Murtagh, Richard Delorme, Myriam Ly Le-Moal, Marion Leboyer, and Anouck Amestoy

Autism spectrum disorders (ASD) are heterogeneous and complex neurodevelopmental conditions that urgently need reliable and sensitive measures to inform diagnosis properly. The Reading the Mind in the Eyes Task (or Eyes Test from now on) is widely used for this purpose. A recent study showed that subcategories of items of the children version of the Eyes Test could be especially discriminative to distinguish ASD and control children. Here, we analyzed the performance on the Eyes Test of 30 high functioning (IQ > 70) adults with ASD and 29 controls from the InFoR cohort multicentric study, using a Generalized Linear Mixed Model. We found that valence and difficulty modulate the performance on the Eyes Test, with easy and positive items being the most discriminative to distinguish ASD and controls. In particular, we suggest this result might be actionable to discriminate ASD patients from controls in subgroups where their overall scores show less difference with controls. We propose for future research the computation of two additional indexes when using the Eyes Test: the first focusing on the easy and positive items (applying a threshold of 70% of correct responses for these items, above which people are at very low risk of having ASD) and the second focusing on the performance gain from difficult to easy items (with a progression of less than 15% showing high risk of having ASD). Our findings open the possibility for a major change in how the Eyes Test is used to inform diagnosis in ASD. *Autism Res* 2021, 14: 380–388. © 2020 The Authors. *Autism Research* published by International Society for Autism Research and Wiley Periodicals LLC.

Lay Summary: The Eyes Test is used worldwide to inform autism spectrum disorders (ASD) diagnosis. We show here that ASD and neurotypical adults show the most difference in performance on *subgroups* of items: ASD adults do not improve as expected when comparing easy and difficult items, and they do not show an improvement for items displaying a positive feeling. We advise clinicians to focus on these comparisons to increase the property of the test to distinguish people with ASD from neurotypical adults.

Keywords: autism spectrum disorders; Reading the Mind in the Eyes; valence; difficulty; Generalized Linear Mixed Model

Introduction

Autism spectrum disorders (ASD) are neurodevelopmental disorders characterized by impaired social interactions and repetitive and restrictive behavior and interests [American Psychiatric Association, 2013]. Several theories have been formulated to explain the source of communication and social reciprocity alterations in autism, suggesting it could arise from poor cognitive or perceptual integration (i.e., having weak “central coherence” [Frith & Happé, 1994]), difficulties

in making sense of social signals (i.e., being “socially blind”) [Baron-Cohen, 2009]), or missing basic mechanisms that otherwise lead typical individuals to be spontaneously attracted to social signals (i.e., lacking “social motivation,” [Chevallier, Kohls, Troiani, Brodtkin, & Schultz, 2012]. But despite these differences in primary causal mechanisms, they all hypothesize that deficits in perception or integration of nonverbal signals are present in ASD. Indeed, a strong body of research has established the presence of difficulties in processing others’ nonverbal social cues and signals in autism,

From the Centre d’Evaluation et Diagnostic de l’Autisme (CEDA), Centre Hospitalier Le Vinatier, Bron, France (M.B., M.-M.G.); Roche Innovation Center, F-Hoffmann La Roche Ltd, Basel, Switzerland (C.C., D.U., L.M.); Université de Bordeaux, Bordeaux, France (M.B., A.A.); Pôle Universitaire de Psychiatrie de l’Enfant et de l’Adolescent, Centre Hospitalier Charles-Perrens, Bordeaux, France (M.B., A.A.); Centre National Pour La Recherche Scientifique, INCIA UMR 5287, Bordeaux, France (M.B., A.A.); Fondation FondaMental, Créteil, France (M.B., A.M.T., I.S., E.M., R.D., M.L., A.A.); UNIACT, NeuroSpin, CEA, Université Paris-Saclay, Paris, France (A.M.T.); Centre Hospitalier de Versailles CHV, Hôpital Mignot, Pôle de Psychiatrie Adulte et Infanto-Juvenile, Centre Expert TSAsdi, Le Chesnay, France (D.M.); INSERM, U955, Institut Mondor de Recherches Biomédicales, IMRB, Laboratoire de Neuro-Psychiatrie translationnelle, Créteil, France (S.C.-C., M.L.); Institut Pasteur, Paris, France (R.D.); Psychiatry and Addictology Department, Université Paris Est Créteil, AP-HP, DMU ADAPT, Mondor University Hospital, Créteil, France (R.D.); Institut Roche, Boulogne-Billancourt, France (M.L.L.-M.)

Received March 30, 2020; accepted for publication August 22, 2020

Address for correspondence and reprints: Matias Baltazar, Centre d’Evaluation et de Diagnostic de l’Autisme (CEDA), Centre Hospitalier Le Vinatier, Bât 504B, 95 Boulevard Pinel, 69 677 Bron Cedex, France. E-mail: matias.baltazar.p8@gmail.com

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published online 15 September 2020 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.2390

© 2020 The Authors. *Autism Research* published by International Society for Autism Research and Wiley Periodicals LLC.

including facial expressions [Uljarevic & Hamilton, 2013] and gaze [Senju & Johnson, 2009].

These features of the syndrome led to the development of tools to assess the perception and understanding of social cues, in order to guide diagnosis and interventions. Among these, the Reading the Mind in the Eyes Task (or Eyes Test from now on) is of particular relevance, due to its ability to discriminate between typical adults and adults with autism without intellectual disability (i.e., high-functioning autistic adults; [Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Penuelas-Calvo, Sareen, Sevilla-Llewellyn-Jones, & Fernandez-Berrocal, 2019]).

The Eyes Test contains 36 pictures of the eye area of male and female actors (in equal number). For each picture, participants have to pick the best complex mental state word out of the four displayed, while all words share the same emotional valence. The same team also designed a children version, where words were simplified to better suit the verbal level of young children [Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001].

Interestingly, some researchers have found that the ability to recognize facial expressions might be differentially impaired in autism, depending on the nature of the emotion and the valence that is displayed [Humphreys, Minshew, Leonard, & Behrmann, 2007]. For example, some publications point toward deficits in recognizing negative emotions such as fear or anger [Shanok, Jones, & Lucas, 2019]. The results of a recent meta-analysis in this domain are less clear, showing an overall deficit in facial emotional recognition in individuals with ASD, independently of IQ or age, but with limited evidence supporting a modulation by emotional valence (except for a trend showing that fear might be less accurately recognized than happiness in ASD patients) [Uljarevic & Hamilton, 2013]. It is of note that this meta-analysis was made on tasks presenting so-called basic expressions [Ekman, 1993], reflecting the content of most studies examining emotion recognition in autism. The question we ask here is how could emotional valence affect the performance at the Eyes Test, a tool that includes very complex mental states such as “Pensiveness,” “Hostility” or “Playfulness”?

One study tested this very idea in children with autism [Baribeau et al., 2015]. Authors used the results from a previous independent publication [Koizumi & Takagishi, 2014] to parcel the items of the Eyes Test (Children version) in three valence categories (positive, neutral, and negative). They found that positive items were the most difficult to recognize out of all three categories for children with autism. They also found that positive stimuli discriminated better between patients and controls, while the overall difference across all items was not significant after controlling for IQ, sex, and age. These results suggest that the Eyes Test overall score might not be sensitive enough to differentiate individuals with ASD from

controls, whereas the subscore for positive items might be. This is crucial information when considering the fact that clinicians have to use the most reliable and sensitive measures in order to properly inform diagnosis and treatment. Note that to the best of our knowledge, no psychometric linking study has been done to establish equivalence between children and adult versions of the test, so one should be cautious when applying conclusions from children data to adults and vice versa. But since both versions were designed with the same purpose in mind (i.e., measuring sociocognitive abilities based on verbal labeling of facial expressions), it seems reasonable to believe that results based on the children version are worthy to inform hypotheses to be tested on adults.

In the present study, we propose to apply a parcellation of items of the Eyes Test (adult version) by valence in accordance with previously published data [Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005] and to test individuals with ASD and controls comparable in terms of age, IQ, sex ratio, and educational level. We expected that participants with autism would perform worse than controls when considering all items. Based on children data, we expected that positive stimuli would induce higher differences between ASD patients and controls, compared to neutral and negative stimuli [Baribeau et al., 2015]. We also explored the influence of the difficulty of items [Baribeau et al., 2015; Lombardo et al., 2016].

Methods

Participants

A total of 59 participants (29 controls and 30 ASD participants) were included in the study (see Table 1). They were all part of the InFoR cohort. InFoR is a multicentric French longitudinal study, promoted by Fondation FondaMental, INSERM, and the Roche Institute. This cohort is constituted of a total of 117 individuals with ASD, including various age groups and levels of intellectual functioning, and 57 controls. Participants were followed during a 2-year period and undertook various assessments at each yearly visit (clinical, neuropsychological, and biological) allowing for deep phenotyping of the cohort. Only adults from the cohort were included in the present study, using data from the initial visit only.

All adult participants were recruited by the Expert Center for Adults with ASD, which is part of the Adult Psychiatry Department of Mondor Hospital (AP-HP, Creteil, France). All patients were adults seeking a diagnosis in a specialized center for the diagnosis of ASD. They all met diagnostic criteria for ASD [American Psychiatric Association, 2013]. ASD diagnosis was established by trained clinicians, based on a clinical interview and on the use of gold standard tools for assessing autistic

Table 1. Participant Characteristics

| Diagnosis | Sample | | Sex | | IQ | | Age | | Education | | |
|-----------|------------|-----------------|-----------------|-----|-----|--------------|-----|-----|--------------|-----|-----|
| | Total n | Female n (%) | Mean (SD) | Min | Max | Mean (SD) | Min | Max | Mean (SD) | Min | Max |
| Control | 29 | 8 (27.6) | 108.0 (16.9) | 78 | 143 | 29.0 (8.4) | 19 | 56 | 15.1 (2.3) | 12 | 20 |
| ASD | 30 | 8 (26.7) | 106.5 (21.2) | 70 | 146 | 29.3 (8.4) | 18 | 52 | 14.5 (2.9) | 8 | 22 |

| ADI-R dimensions | | Mean (SD) |
|--------------------|--|--------------|
| Social interaction | | 15.9 (6.5) |
| Communication | | 9.5 (5.6) |
| RRB | | 4.1 (1.9) |

| ADOS dimensions | | Mean (SD) |
|--------------------|--|--------------|
| Communication | | 3.8 (1.5) |
| Social Interaction | | 7.1 (3.2) |
| RRB | | 0.9 (1.1) |

Abbreviations: ASD, autism spectrum disorder; Max, maximum; Min, minimum; RRB, restricted and repetitive behavior.

features (i.e., the Autism Diagnostic Interview-Revised (ADI-R) [Lord, Rutter, & Le Couteur, 1994] and the Autism Diagnostic Observation Schedule (ADOS) [Lord et al., 2000]). Control participants were recruited through advertisement in the local press, at the local university and through Fondation FondaMental social network that promotes participation on research on mental health. Controls had no DSM-IV Axis-I psychiatric disorder and no first-degree family history of schizophrenia, schizoaffective, autism or bipolar disorder (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision; American Psychiatric Association, 2000). IQ was estimated using the full version of the fourth edition of the Wechsler Adult Intelligence Scale, at the exclusion of two patients who underwent the third edition (one with the full version and one with the abbreviated version). All participants were categorized as “high functioning,” meaning that their overall IQ was superior or equal to 70. Control and ASD participants were not different regarding age, sex ratio, IQ, or years of education (all *P* values superior to 0.37, see Table 1 for descriptive statistics), which is important given previous evidence of their impact on Eyes Test performance [Baker, Peterson, Pulos, & Kirkland, 2014; Kirkland, Peterson, Baker, Miller, & Pulos, 2013].

The study was part of the InFoR project, sponsored by INSERM (French National Institute for Health, Clinical Trial C07-33), through a collaboration with the Roche Institute for Research and Translational Medicine. It was granted approval by local Ethics Committee or “Comité de Protection des Personnes” on November 14, 2008,

authorized by the French authorities (ANSM B80738-70 on August 11, 2008) and registered in a public trials registry (NCT02628808). All study participants gave their informed written consent to participation, in line with French ethical guidelines.

Experimental Design and Procedure

The Eyes Test. A validated French version of the Eyes Test was used [Prevost et al., 2014]. The test contains 36 pictures depicting the eye region of actors displaying different facial expressions. Participants had to identify the mental state associated with the picture by selecting the most appropriate adjective out of the four displayed around the picture.

Items were subdivided into negative (*n* = 12), neutral (*n* = 16), and positive items (*n* = 8) following the work of Harkness et al. [2005].

Difficulty was determined by dichotomizing stimuli according to their associated percentages of correct answers in the norming sample (French adult version) [Prevost et al., 2014]. All stimuli with a percentage of correct recognition in Prevost et al.’s work superior to 70% (the median was 71%) were classified as Easy (*n* = 19) while the others were classified as Difficult (*n* = 17). This median split methodology is similar to previous dichotomizations of the Eyes Test items [Domes, Heinrichs, Michel, Berger, & Herpertz, 2007; Feeser et al., 2015; Radke & de Bruijn, 2015; Woolley et al., 2014].

At this point, we built a contingency table to check the repartition of items between Valence and Difficulty

Table 2. Distribution of Items per Valence and Difficulty Categories

| Valence | Difficulty | | Total (items) |
|---------------|--------------------------------------|---------------------------------------|---------------|
| | Easy (item no./total) | Hard (item no./total) | |
| Negative | 5, 14, 22, 36/4 items | 2, 11, 17, 23, 26, 27, 34, 35/8 items | 12 |
| Neutral | 3, 9, 12, 15, 18, 24, 28, 32/8 items | 4, 7, 8, 10, 13, 19, 29, 33/8 items | 16 |
| Positive | 1, 6, 16, 20, 21, 25, 30/7 items | 31/1 item | 8 |
| Total (items) | 19 | 17 | 36 |

factors (see Table 2). This table shows a mostly even distribution when considering each factor alone. But it shows an unbalanced distribution of items when crossing the two factors. Indeed, the Difficult \times Positive cell contains only one item, compared to four to eight items in other cells. That is why we focused our analyses on assessing the effect of each factor alone, without assessing their interaction.

Procedure. The Eyes Test was proposed among several other evaluations in the context of the InFoR project (data not reported here, except for IQ, ADOS, and ADI-R data). Stimuli from the Eyes Test were printed on paper and presented one by one. Participants had unlimited time to provide their answer.

Statistical Analysis

All analyses were performed using the R software [R Core Team, 2019]. We computed a Generalized Linear Mixed Model using the *glmer* command of the *lme4* package

[Bates, Mächler, Bolker, & Walker, 2015] trying to predict performance at each Eyes Test item (1 = success and 0 = failure) by the following fixed factors: Valence (3 levels: Neutral, Negative, Positive), Difficulty (2 levels: Easy, Difficult) and Diagnostic Group (2 levels: Controls, Patients). We used a binomial link function and a maximum likelihood estimation method (Laplace Approximation). The interactions between Diagnostic Group and Valence and between Diagnostic Group and Difficulty were also tested. Due to the unbalanced distribution of items among the factorial design, we did not explore interactions between Valence and Difficulty, nor between Group, Valence and Difficulty. Participants were entered in the model as a random factor. The α value was set at 0.05 for all statistical tests. Significant effects were then explored by computing pairwise least square means contrasts, using the *multcomp* package [Hothorn, Bretz, & Westfall, 2008] and applying a Bonferroni correction for multiple comparisons to protect against Type I error. All reported *P* values are corrected for multiple comparisons throughout the text, tables and figures. Data were plotted using *ggplot2* [Wickham, 2016].

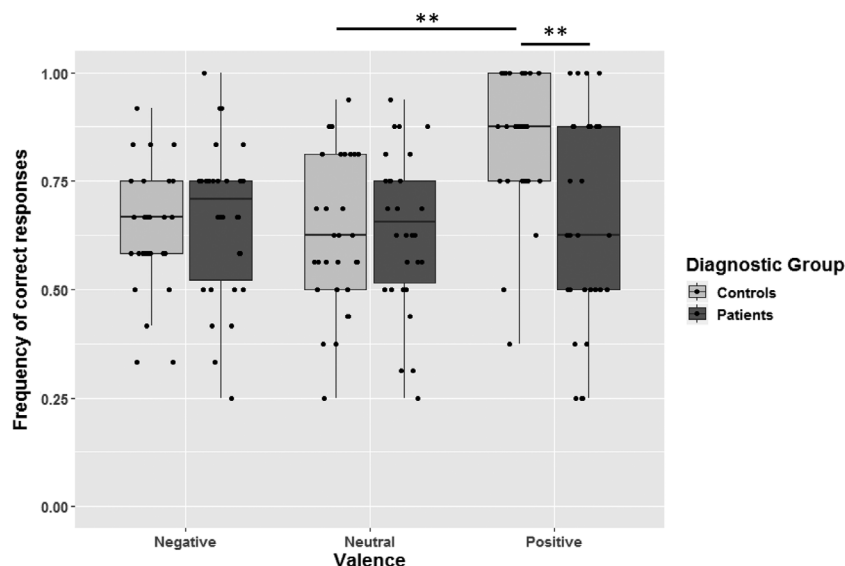


Figure 1. Eyes Test performance as a function of diagnostic group and valence of the items. *Note.* Individual data points are superimposed on boxplots. Performance was quantified as the frequency of correct responses, in order to obtain comparable scores for each valence category even though they contained different numbers of items. $**P < 0.01$.

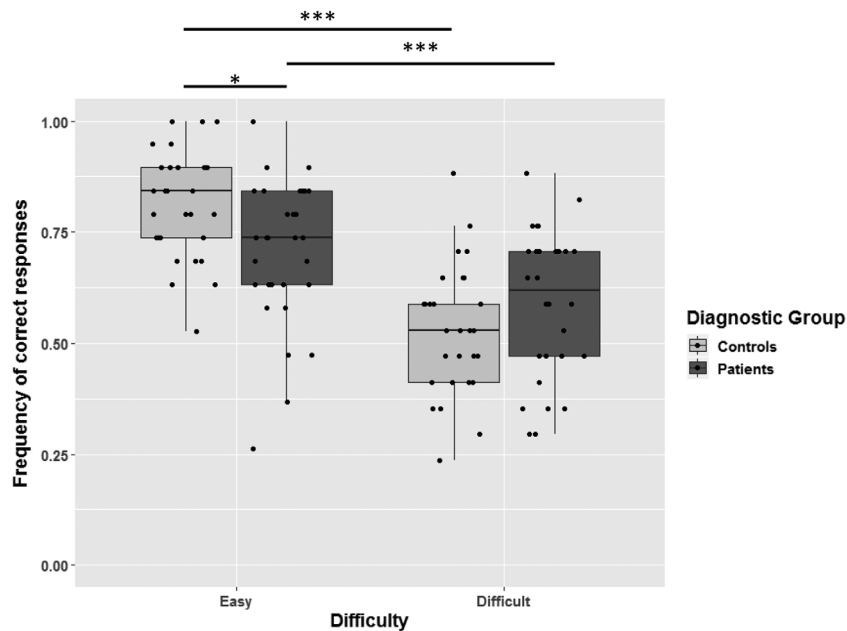


Figure 2. Eyes Test performance as a function of diagnostic group and difficulty of the items. *Note.* Individual data points are superimposed on boxplots. Performance was quantified as the frequency of correct responses, in order to obtain comparable scores for each difficulty category even though they contained different numbers of items. *** $P < 0.001$; * $P < 0.05$.

A power analysis was conducted using the G*Power software (Version 3.1) [Faul, Erdfelder, Lang, & Buchner, 2007] and established that a total sample size of 50 participants would be needed to identify an effect of Group on overall Eyes Test performance. For this computation, we set the power parameter to 80% and derived the expected effect size from the Baron-Cohen and colleagues' paper presenting the revised version of the Eyes Test [Baron-Cohen, Wheelwright, Hill, et al., 2001].

Results

Descriptive statistics are displayed in Figures 1 and 2. We first assessed overdispersion in the Generalized Linear Mixed Model by using the overdispersion function by Bolker [2020]. Results indicated that data were not overdispersed (overdispersion ratio = 0.96; $\chi^2_{2115} = 2038.20$; $P = 0.88$). Results of the GLMM showed that Eyes Test scores significantly differed as a function of Valence ($\chi^2_2 = 6.01$; $P = 0.049$) and Difficulty ($\chi^2_1 = 82.28$; $P < 0.001$). The effect of Group was not significant ($\chi^2_1 = 0.02$; $P = 0.89$). The interaction between Group and Valence ($\chi^2_2 = 12.66$; $P = 0.002$) and between Group and Difficulty ($\chi^2_2 = 7.44$; $P = 0.006$) was significant.

Regarding the Group \times Valence interaction, multiple comparisons (number of comparisons = 15) revealed that Positive items led to significant differences between Groups ($Z = 3.56$; $P = 0.008$), while no significant

differences were found for Negative ($Z = 0.05$; $P = 1$) and Neutral items ($Z = -0.14$; $P = 1$). Among Controls, only the difference between Positive and Neutral items was significant ($Z = -3.67$; $P = 0.016$). Among Patients, pairwise comparisons between Valence categories did not yield any significant results.

Multiple comparisons revealed that the Group \times Difficulty interaction (number of comparisons = 6) was characterized by a significant difference between Groups for Easy items ($Z = 3.19$; $P = 0.030$), while the Group difference was not significant for Difficult items ($Z = 0.12$; $P = 1$). In concordance with the main effect of Difficulty, pairwise comparisons of Easy and Difficult items were both significant among Controls ($Z = 8.11$; $P < 0.001$) and Patients ($Z = 4.64$; $P < 0.001$). A careful examination of the Z values indicates that this difference was more pronounced for controls.

Discussion

In the present study, we first replicated the effect of the valence of stimuli on the Eyes Test performance observed by Baribeau and colleagues in a sample of children [2015]. Specifically, positive stimuli led to better recognition scores in the controls group compared to neutral and negative ones, while no such difference was observed in individuals diagnosed with ASD. Secondly, we showed that easy items better discriminate between patients and controls, compared to difficult items, for which similar

performance was found, a result that is also similar to what was observed in children by Baribeau et al. [2015].

It is of note that even though the effects were similar for adults (in the present study) and children (in Baribeau and colleagues' study, 2015), our effect was driven by control participants performing best for positive items, while the children effect was driven by individuals with ASD performing worse on easy items. Baribeau and colleagues proposed that due to the fact that younger children rely more on the mouth region to identify happiness on faces [Evers, Kerkhof, Steyaert, Noens, & Wagemans, 2014], children with a neurodevelopmental disorder such as ASD could show a relative immaturity in this aspect compared to age matched controls, which means that they rely more on the mouth region to identify positive facial expressions. This could explain their lower performance when identifying positively valenced items at the Eyes Test, since only the eye region is presented. Regarding adults, we propose that the rewarding nature of positively valenced faces, which is observed in neurotypical participants [Wang, Hahn, DeBruine, & Jones, 2016], could be associated with increased motivation, and thus higher resource allocation leading to higher performance. Evidence in the literature indicates that such a mechanism could be dysfunctional in ASD patients [Dichter, Richey, Rittenberg, Sabatino, & Bodfish, 2012]. Other evidence even suggest that displays of social affiliation, such as social touch [Peled-Avron & Shamay-Tsoory, 2017] or eye contact [Madipakkam, Rothkirch, Dziobek, & Sterzer, 2017; Trevisan, Roberts, Lin, & Birmingham, 2017], could be aversive for individuals with ASD. This could further explain why ASD participants did not show the improvement for positive stimuli in the present study, since such stimuli are associated with markers of sociality or cooperativeness (e.g., "Playful," "Friendly," or "Flirtatious"). Note that our effect could also be due to the fact that positive items were also easier, and that group differences were only observed for easy items in our adult sample (see below for a more thorough discussion on the matter).

In the present study, high-functioning adults with autism seemed to be equally good as control participants at identifying complex mental states based on facial expressions from the eye region, while they seem to have difficulty in decoding easier and simpler stimuli. This could reflect the existence of two processes when decoding facial expressions: one intuitive, used for simpler expressions, for which ASD individuals could be particularly impaired; and one "cognitive," based on conscious reasoning about the association of certain facial features and the state of mind they are supposed to reflect, for which high functioning ASD individuals could be equally good as controls. This interpretation is reminiscent of "dual systems" theories [Kahneman, 2011], and is further in line with data showing that high-

functioning ASD individuals display reduced intuitive reasoning and greater deductive reasoning in cognitive tasks [Brosnan, Ashwin, & Lewton, 2017; Brosnan, Lewton, & Ashwin, 2016]. Another explanation could be that due to an overall impairment in emotional facial expression decoding in ASD [Uljarevic & Hamilton, 2013], modulation of performance due to the difficulty of the items is less pronounced in this population, with maximum difference between patients and controls on easier items. We advocate in favor of this latter, more parsimonious hypothesis, but other studies are needed to better understand this effect (e.g., by asking participants to report their reasoning or lack thereof when solving items, by measuring reaction times to evaluate the cognitive cost of solving items, etc.).

The valence parcellation has been criticized in several studies, which found evidence in favor of a one dimensional structure of the Eyes Test [Carey & Cassels, 2013; see also Preti, Vellante, & Petretto, 2017; Vellante et al., 2013 for data on the Italian version of the test]. This brings higher validity to the hypothesis that the results obtained in the present study could be mainly attributable to the difficulty, and not the valence of the items. This also calls for nuanced interpretations in several studies investigating the effect of valence on Eyes Test performance [Fertuck et al., 2009; Frick et al., 2012; Harkness et al., 2005; Preissler, Dziobek, Ritter, Heekeren, & Roepke, 2010; Preti et al., 2017; Richman & Unoka, 2015; Savage & Lenzenweger, 2018; Schilling et al., 2012; Unoka, Fogd, Seres, Kéri, & Csukly, 2015; Wolkenstein, Schönenberg, Schirm, & Hautzinger, 2011]. For example, the meta-analysis by Richman and Unoka [2015] indicates that patients with Major Depressive Disorders are impaired on the Eyes Test, with even more deficit for positive items, attributing this pattern to their being less able to decode rewarding positive facial expressions. But this could in fact reflect a more trivial process where easy items could be more informative in discriminating participants who are impaired (e.g., patients with major depression) from participants with no impairment in overall facial emotion recognition, while difficult items might lead to floor effects in both groups. To summarize this point, assessing the respective contribution of valence and difficulty independently in the adult Eyes Test might be problematic in some versions of the test.

It is of note that the Eyes Test overall score was not discriminative in distinguishing patients and controls in our study. This was not due to our study being insufficiently powered in this regard, as proved by the power analysis we conducted (see Methods section). This result is in contradiction with a recent meta-analysis showing an overall significant difference between ASD patients and controls (including children and adult data [Penuelas-Calvo et al., 2019]). A recent study by Lombardo and colleagues

identified several subgroups based on their performance on the Eyes Test, using an unsupervised clustering approach on large samples of more than 600 patients and 200 controls dispatched in discovery and replication samples. Even though subgroups encompassing the majority of ASD patients did perform significantly worse compared to control participants, some subgroups of ASD and control participants did not differ between one another. As a tentative explanation for the null effect in our samples for the overall score, we propose that our ASD and control samples might be particularly enriched with individuals from these clusters showing weaker group differences.

Contrary to overall score, the easy items allow to distinguish patients and controls in our study, and the same result was found for positive items. This is crucial information since clinicians have to use the most reliable and sensitive measures in order to properly inform diagnosis and treatment. Instead of the Eyes Test overall score, we propose to use the performance gain from difficult to easy items, with participants progressing less than a threshold of 0.15 being at high risk of having ASD (Sensitivity = 70.0%; Specificity = 82.8%). Only considering the subgroup of easy and positive items could also be useful, with participants with frequency of correct responses of more than 0.70 on these items being at a very low risk of having ASD, due to the high specificity of this threshold (Sensitivity = 46.7%; Specificity = 93.1%).

Our study has several limitations. First, even though our sample size was enough to be able to detect overall Eyes Test score differences, it was still quite small regarding the interactions between diagnostic group and valence or difficulty. We used a restrictive Bonferroni correction to protect against Type I errors, but Type II errors might still have occurred, thus requiring future replication studies with bigger sample sizes. This is especially true for the thresholds that we propose, which should be validated in independent and broader samples before being used in clinical settings. Also, our results might be generalizable to French populations only, with the French version of the Eyes Test. Second, we did not include a reaction time measure in our protocol. Such measure could be used to better discriminate between patients and controls. It could also be very informative about the underlying strategies deployed by patients and controls when performing the task, with shorter reaction times reflecting automatic strategies and longer reaction times reflecting more “cognitive” and controlled processes. Third, we did not include among controls a group with a psychiatric or neurodevelopmental condition other than ASD, so the assessment of the specificity of our measure is limited. For example, it is possible that patients with major depression could show the same pattern of results as our ASD sample (see [Richman & Unoka, 2015] for evidence in favor of strong recognition deficit of positive or

easy items of the Eyes Test in patients with major depression and Baribeau et al. [2015] for similar evidence in patients with ADHD).

Conclusions

In the present article, we show that the valence and difficulty of the items modulate the performance of adults with and without autism in the Eyes Test, with easy or positive items having the most discriminative power for sorting out individuals. In particular, we suggest this result might be actionable to discriminate ASD patients from controls in subgroups where their overall scores show less difference with controls. We propose for future research the computation of two additional indexes when using the Eyes Test: the first focusing on the easy and positive items (applying a threshold of 70% of correct responses, above which people are at very low risk of having ASD) and the second focusing on the performance gain from difficult to easy items (with a progression of less than 15% showing high risk of having ASD). These cutoffs have to be tested in broader independent samples to better assess their sensitivity and specificity. Our work opens the possibility for a major change in how the Eyes Test is used to inform diagnosis in ASD.

Acknowledgments

This work has been co-ordinated by Fondation FondaMental (www.fondation-fondamental.org) and achieved thanks to the following organisms and establishments: INSERM, AP-HP, CHU Bordeaux, Hôpital Charles Perrens, Robert Debré, and Henri Mondor's CIC. It was financially supported in part by the Institut Roche, in part by the Investissements d'Avenir program managed by the ANR under reference ANR-11-IDEX-0004-02.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: Dsm-5* (5th ed.). Philadelphia, PA: American Psychiatric Publishing.
- Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and “Reading the Mind in the Eyes”. *Intelligence*, 44, 78–92. <https://doi.org/10.1016/j.intell.2014.03.001>
- Baribeau, D. A., Doyle-Thomas, K. A. R., Dupuis, A., Iaboni, A., Crosbie, J., McGinn, H., ... Anagnostou, E. (2015). Examining and comparing social perception abilities across childhood-onset neurodevelopmental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(6), 479–486. e1. <https://doi.org/10.1016/j.jaac.2015.03.016>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: A study with normal adults, and adults with Asperger syndrome

- or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(2), 241–251.
- Baron-Cohen, S. (2009). Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, 1156, 68–80. <https://doi.org/10.1111/j.1749-6632.2009.04467.x>
- Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, 5(1), 47–78.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <http://dx.doi.org/10.18637/jss.v067.i01>
- Bolker, B. (2020). GLMM FAQ. Retrieved from <https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#overdispersion>
- Brosnan, M., Ashwin, C., & Lewton, M. (2017). Brief report: intuitive and reflective reasoning in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 47(8), 2595–2601. <https://doi.org/10.1007/s10803-017-3131-3>
- Brosnan, M., Lewton, M., & Ashwin, C. (2016). Reasoning on the autism spectrum: A dual process theory account. *Journal of Autism and Developmental Disorders*, 46(6), 2115–2125. <https://doi.org/10.1007/s10803-016-2742-4>
- Carey, J. M., & Cassels, T. G. (2013). Comparing two forms of a childhood perspective-taking measure using CFA and IRT. *Psychological Assessment*, 25(3), 879–892. <https://doi.org/10.1037/a0032641>
- Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16(4), 231–239. <https://doi.org/10.1016/j.tics.2012.02.007>
- Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., & Bodfish, J. W. (2012). Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders*, 42(2), 147–160. <https://doi.org/10.1007/s10803-011-1221-1>
- Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves “mind-reading” in humans. *Biological Psychiatry*, 61(6), 731–733. <https://doi.org/10.1016/j.biopsych.2006.07.015>
- Ekman, P. (1993). Facial expression and emotion. *The American Psychologist*, 48(4), 384–392.
- Evers, K., Kerkhof, I., Steyaert, J., Noens, I., & Wagemans, J. (2014). No differences in emotion recognition strategies in children with autism spectrum disorder: Evidence from hybrid faces. *Autism Research and Treatment*, 2014, 1–8. <https://doi.org/10.1155/2014/345878>
- 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. <https://doi.org/10.3758/BF03193146>
- Feeser, M., Fan, Y., Weigand, A., Hahn, A., Gärtner, M., Böker, H., ... Bajbouj, M. (2015). Oxytocin improves mentalizing—Pronounced effects for individuals with attenuated ability to empathize. *Psychoneuroendocrinology*, 53, 223–232. <https://doi.org/10.1016/j.psyneuen.2014.12.015>
- Fertuck, E. A., Jekal, A., Song, I., Wyman, B., Morris, M. C., Wilson, S. T., ... Stanley, B. (2009). Enhanced “Reading the Mind in the Eyes” in borderline personality disorder compared to healthy controls. *Psychological Medicine*, 39(12), 1979–1988. <https://doi.org/10.1017/S003329170900600X>
- Frick, C., Lang, S., Kotchoubey, B., Sieswerda, S., Dinu-Biringer, R., Berger, M., ... Barnow, S. (2012). Hypersensitivity in borderline personality disorder during mindreading. *PLoS One*, 7(8), e41650. <https://doi.org/10.1371/journal.pone.0041650>
- Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, 50(1), 115–132. [https://doi.org/10.1016/0010-0277\(94\)90024-8](https://doi.org/10.1016/0010-0277(94)90024-8)
- Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition & Emotion*, 19(7), 999–1025.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(3), 346–363.
- Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, 45(4), 685–695. <https://doi.org/10.1016/j.neuropsychologia.2006.08.003>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Macmillan.
- Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in “Reading the Mind in the Eyes Test”. *North American Journal of Psychology*, 15(1), 121.
- Koizumi, M., & Takagishi, H. (2014). The relationship between child maltreatment and emotion recognition. *PLoS One*, 9(1), e86093. <https://doi.org/10.1371/journal.pone.0086093>
- Lombardo, M. V., Lai, M.-C., Auyeung, B., Holt, R. J., Allison, C., Smith, P., ... Baron-Cohen, S. (2016). Unsupervised data-driven stratification of mentalizing heterogeneity in autism. *Scientific Reports*, 6, 35333. <https://doi.org/10.1038/srep35333>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685.
- Madipakkam, A. R., Rothkirch, M., Dziobek, I., & Sterzer, P. (2017). Unconscious avoidance of eye contact in autism spectrum disorder. *Scientific Reports*, 7(1), 13378. <https://doi.org/10.1038/s41598-017-13945-5>
- Peled-Avron, L., & Shamay-Tsoory, S. G. (2017). Don’t touch me ! Autistic traits modulate early and late ERP components during visual perception of social touch. *Autism Research: Official Journal of the International Society for Autism Research*, 10(6), 1141–1154. <https://doi.org/10.1002/aur.1762>
- Penuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fernandez-Berrocal, P. (2019). The “Reading the Mind in the

- Eyes" test in autism-spectrum disorders comparison with healthy controls: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, 49(3), 1048–1061. <https://doi.org/10.1007/s10803-018-3814-4>
- Preissler, S., Dziobek, I., Ritter, K., Heekeren, H. R., & Roepke, S. (2010). Social cognition in borderline personality disorder: Evidence for disturbed recognition of the emotions, thoughts, and intentions of others. *Frontiers in Behavioral Neuroscience*, 4, 182. <https://doi.org/10.3389/fnbeh.2010.00182>
- Preti, A., Vellante, M., & Petretto, D. R. (2017). The psychometric properties of the "Reading the Mind in the Eyes" Test: An item response theory (IRT) analysis. *Cognitive Neuropsychiatry*, 22(3), 233–253. <https://doi.org/10.1080/13546805.2017.1300091>
- Prevost, M., Carrier, M.-E., Chowne, G., Zelkowitz, P., Joseph, L., & Gold, I. (2014). The Reading the Mind in the Eyes test: Validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cognitive Neuropsychiatry*, 19(3), 189–204.
- R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Radke, S., & de Bruijn, E. R. A. (2015). Does oxytocin affect mind-reading? A replication study. *Psychoneuroendocrinology*, 60, 75–81. <https://doi.org/10.1016/j.psyneuen.2015.06.006>
- Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: Meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, 207(6), 483–489. <https://doi.org/10.1192/bjp.bp.114.152108>
- Savage, M., & Lenzenweger, M. F. (2018). The impact of social exclusion on "Reading the Mind in the Eyes" performance in relation to borderline personality disorder features. *Journal of Personality Disorders*, 32(1), 109–130. https://doi.org/10.1521/pedi_2017_31_293
- Schilling, L., Wingenfeld, K., Löwe, B., Moritz, S., Terfehr, K., Köther, U., & Spitzer, C. (2012). Normal mind-reading capacity but higher response confidence in borderline personality disorder patients. *Psychiatry and Clinical Neurosciences*, 66(4), 322–327. <https://doi.org/10.1111/j.1440-1819.2012.02334.x>
- Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience and Biobehavioral Reviews*, 33(8), 1204–1214. <https://doi.org/10.1016/j.neubiorev.2009.06.001>
- Shanok, N. A., Jones, N. A., & Lucas, N. N. (2019). The nature of facial emotion recognition impairments in children on the autism spectrum. *Child Psychiatry and Human Development*, 50, 661–667. <https://doi.org/10.1007/s10578-019-00870-z>
- Trevisan, D. A., Roberts, N., Lin, C., & Birmingham, E. (2017). How do adults and teens with self-declared Autism Spectrum Disorder experience eye contact? A qualitative analysis of first-hand accounts. *PLoS One*, 12(11), e0188446. <https://doi.org/10.1371/journal.pone.0188446>
- Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43(7), 1517–1526. <https://doi.org/10.1007/s10803-012-1695-5>
- Unoka, Z. S., Fogd, D., Seres, I., Kéri, S., & Csukly, G. (2015). Early maladaptive schema-related impairment and co-occurring current major depressive episode-related enhancement of mental state decoding ability in borderline personality disorder. *Journal of Personality Disorders*, 29(2), 145–162. https://doi.org/10.1521/pedi_2014_28_146
- Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, 18(4), 326–354. <https://doi.org/10.1080/13546805.2012.721728>
- Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). The motivational salience of faces is related to both their valence and dominance. *PLoS One*, 11(8), e0161114. <https://doi.org/10.1371/journal.pone.0161114>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer. <https://www.springer.com/gp/book/9780387981413>
- Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, 132(1–2), 104–111. <https://doi.org/10.1016/j.jad.2011.02.010>
- Woolley, J. D., Chuang, B., Lam, O., Lai, W., O'Donovan, A., Rankin, K. P., ... Vinogradov, S. (2014). Oxytocin administration enhances controlled social cognition in patients with schizophrenia. *Psychoneuroendocrinology*, 47, 116–125. <https://doi.org/10.1016/j.psyneuen.2014.04.024>