Optimizing the RMET to measure bias not performance differences

Our Internet survey yielded a robust difference among neurodiverse and neurotypical persons in the interpretation of eyes, write Gerit Pfuhl and Leif Ekblad.

BY: Gerit Pfuhl and Leif Ekblad

Understanding and expressing emotions comes naturally to humans, and emotion processing plays a fundamental role in explaining the behavior of others (Olsson & Ochsner, 2007). Expression and recognition of emotions, regardless of the modality used, facilitates normal social interaction. Accordingly, differences in emotion processing, by either expressing them or interpreting them differently, can lead to misunderstandings and social withdrawal. Indeed, training emotion recognition improves social skills (e.g., Rice et al., 2015). The group most studied consists of people with an autism spectrum diagnosis (ASD). These individuals are, by definition (DSM-V, American Psychiatric Association, 2013), characterized as having difficulties in using, sharing and responding to emotions. Furthermore, persons diagnosed with ASD are a) impaired in the ability to communicate, b) deficient in understanding social interactions, and c) likely to show repetitive behaviors or restricted interests (DSM-V, American Psychiatric Association, 2013). This is a definition by three core traits.

However, Happé, Ronald, and Plomin (2006) reported that children on the autistic spectrum and controls do not fall into two distinct clusters. Rather, autism is merely the extreme end of a unimodal distribution along each of the three core traits, implying that the dividing line is arbitrary. Not only that, but the three core traits are only weakly correlated, suggesting that the definition of autism that requires the presence of all three traits is less driven by data than by the coincidence of all three traits being conspicuous. This suggests that a) some people who are not diagnosed with autism also have difficulties in emotion processing, and b) some people with autism may not have severe problems with emotion processing.

Given that autistic traits are a continuum (Bailey, Palferman, Heavey, & Le Couteur, 1998; Baron-Cohen et al., 2001a; Happé et al., 2006; Lundström et al., 2012), it is not ideal to dichotomize participants based on a formal diagnosis that itself rests on three core traits. As we will review below, the inconsistent literature on emotion processing in autism may be due to small sample sizes in the studies, the individual symptom severity along the three core traits for each participant, and a normative view based on “healthy controls.”

Here we first briefly present the neurodiversity concept. We then review briefly the emotion processing literature in autism. Next, we present a study with more than 10,000 participants who provided mental state descriptors for eye stimuli. Through an iterative process, we have derived new mental state descriptors that allow to measure the bias a person has in interpreting these eye stimuli.

The neurodiversity concept

Autistic traits are a quantitative, not a qualitative, difference in the general population (Bailey et al., 1998; Baron-Cohen et al., 2001a; Happé et al., 2006; Lundström et al., 2012). Happé et al. (2006) argued for a continuum in the general population based on questionnaires sent out to parents of twins (Ronald, Happé, & Plomin, 2005). This argument was also supported by another large twin study where parents of Swedish twins were interviewed (Lundström et al., 2012). Relatives of diagnosed persons with autism do show above average systemizing and below average empathizing compared to matched unaffected families (Grove et al. 2013). Further, mild expression of autistic traits can be advantageous. For example, many mathematicians score high on autistic traits (Baron-Cohen et al., 2001a). It is in this vein that some of the people with high-functioning autism (HFA) prefer to define themselves not as ill but rather as neurodiverse (Griffin & Pollak, 2009; Jaarsma & Welin, 2012).

The concept of neurodiversity is not limited to autism. Persons with ADHD, dyslexia, and dyspraxia may also refer to themselves as neurodiverse. Thus, neurodiversity stands for everything but having a disease or disorder and being inferior to others (Kapp et al., 2013). Accordingly, Ekblad (2013) developed a questionnaire that measures neurodiversity (ND) and neurotypicality (NT). Using the terms neurodiverse and neurotypical avoids referencing someone as a person with autism, Asperger syndrome, or healthy, especially since there is no “healthy” within personality psychology. We will use this neurodiverse/neurotypical questionnaire to measure where along the continuum a person lies.

Emotion processing and autistic traits

The basic emotions (i.e., happy, sad, fearful, angry, disgusted, and surprised) are thought to be universally transmitted and understood (Ekman, 2003; but see Russell, 1994). Uljarevic and Hamilton (2013) found only marginal impairments in detecting sadness, anger, surprise, fear, and disgust but no difference for happiness among persons with autism. In a recent review, Nuske, Vivanti, and

Dissanayake (2013) concluded that emotion impairments are neither universal nor specific to autism. This conclusion is not surprising given that deficit in emotion processing is just one of the symptoms of autism and is shared with other mental disorders or vulnerabilities toward them (e.g., schizophrenia, depression) (Leppänen, 2006; Trémeau, 2006).

Similarly, in a review by Harms, Martin, and Wallace (2010), data on facial emotion recognition impairments was found to be inconsistent. Compensatory mechanisms may allow normative performance. Indeed, in the "Reading the Mind in the Eyes" test (RMET) Miu, Pana, and Avram (2012) found no difference in the accuracy, but they did find a longer response time in selecting the best fitting emotion to the eyes among a population of students classified as either high in autistic traits or low. Related results have been found when basic emotions had to be matched to an entire face, as with the Karolinska directed emotional faces task. Here, Sucksmith et al. (2013) found a longer response time in persons with autism but not in their parents and healthy controls. They also found that persons with autism had a lower accuracy than the other two groups, and there was a significant correlation with the Emotional Quotient. A lower accuracy for the ASD group was also found when stimuli were presented only briefly, 250 ms (Walsh, Creighton, & Rutherford, 2016). The authors used complex expressions like arrogant, bored, flirtatious, and thoughtful. However, in another study, Tracy et al. (2011) found no difference among children with and without autism on accuracy or speed in recognizing basic and complex emotions. These contradictory results might be due to individual differences in eye fixation. Kirchner et al. (2011) report that people with ASD who spent more time looking at the eyes performed better on the RMET, but eye fixation time was not correlated with autism severity as measured with the ADI-R.

These data suggest impairments in emotion recognition in people with ASD, but they also indicate that the tasks used are not sensitive enough to measure emotion recognition reliably. This might be due, on the one hand, to the ASD group being heterogeneous, since the standard deviation in the ASD group is often larger than it is in the control group. On the other hand, it might be attributable to measuring performance only instead of including a measure of bias. Comparing people with autism to "healthy controls" is a performance measurement (i.e., the score of the healthy controls serves as a benchmark). That is prominently the case in the RMET. The test was developed by selecting the answer options of non-autistic students as "correct" despite there being no correct answer per se (Baron-Cohen et al., 2001a).

Answering a task is also influenced by cognitive biases. An example is the familiarity bias. The people we see most often are very familiar to us and also perceived as more beautiful (i.e., our own children, parents, spouse) (Peskin & Newell, 2004). Instead of familiarity, the RMET has been analyzed for bias based on the valence of the stimuli. In the meta-analysis of Richman and Unoka (2015), people with major depression had a lower overall accuracy score, mainly driven by errors on positive items (positive valence). This takeaway is in agreement with other behavioral tasks finding a negativity bias or absence of a positivity bias in people with depression (Peckham, McHugh, & Otto, 2010).

Given the inconsistent results on emotion recognition tasks (Harms et al., 2010), we set out to modify the RMET to assess performance and bias along the autism continuum. We expected a stronger discrepancy between the groups by selecting new emotional expressions through an iterative process among humans who varied in the amount of their autistic traits. Further, instead of using a performance benchmark (correct/incorrect), we provide a different scoring validated on over 2,000 people. This scoring classifies responses as either more autistic/neurodiverse (see below) or more neurotypical-like. This score is a bias measure.

Method

Over 10,000 persons participated in four online surveys hosted at the Aspie Quiz website (http://rdos.net/eng) and at Qualtrics (Qualtrics.com). Participants were asked to choose their gender and state of residence, whether they have been diagnosed with AS/HFA/PDD or are suspected of belonging to this group, or to state that they do not have any known developmental disorder. Their birth year was also asked for, but some participants did not select a sensible birth year. Rather, they picked one of the first years in the list that started at 1900. Since diagnosis was not verified by a clinician, we are not using this response as a grouping factor. Instead, we are solely using a person's Aspie Quiz score. This score correlates highly with the Autism Quotient score (Ekblad, 2013). Hence, the experiments primarily used the ND and NT group classification, which a) had well-documented properties by using the Aspie Quiz scoring system, and b) did not rely on a single statement. Approximately 5% to 10% of the participants per task indicated to be diagnosed with autism. About 50% of the participants were from the USA, and the rest were mostly from Europe and Australia.

Compliance with ethical standards

The studies comply with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The studies were constructed in such a way that participants could not be identified (i.e., no IP address was registered). Participants were informed about the objectives of the study, and informed consent was obtained from all participants before taking the test. Research items in the Aspie Quiz could be skipped by selecting the “?” alternative (checked by default).

Materials

The Reading the Mind in the Eyes test (RMET). Baron-Cohen and colleagues (2001a) presented students a selection of cropped faces showing mainly the region around the eyes and displaying various mental states, which they termed the “Reading the Mind in the Eyes test” (RMET). Through an iterative process, they kept those “eyes” that had a clear preferred emotional expression (i.e., over twice as many answers than the other three mental state descriptors received). It is important to stress that the expression seen on the pictures had no correct answer per se. After that, the authors presented this stimulus set to a naïve group of students and adults who were diagnosed with Asperger syndrome. The latter group showed a larger spread in their answers and they less often agreed on the option the student population had agreed upon. We build on this iterative process for selecting better mental state descriptors. For the short named “adult version,” we run a version based on 37 eye stimuli and four mental state descriptors. The “child version” contained 28 eye stimuli with an age-appropriate vocabulary (Baron-Cohen et al., 2001a, 2001b). Twenty-five stimuli are identical between the two versions and all stimuli are from young to old Caucasian adults. The adult version also asked for clarity, rated on a 5-point Likert scale from “not very clear” to “very clear.” For the iterative process (see procedure) we chose the child version, because we wanted to avoid any influence of verbal intelligence and English proficiency on the RMET.

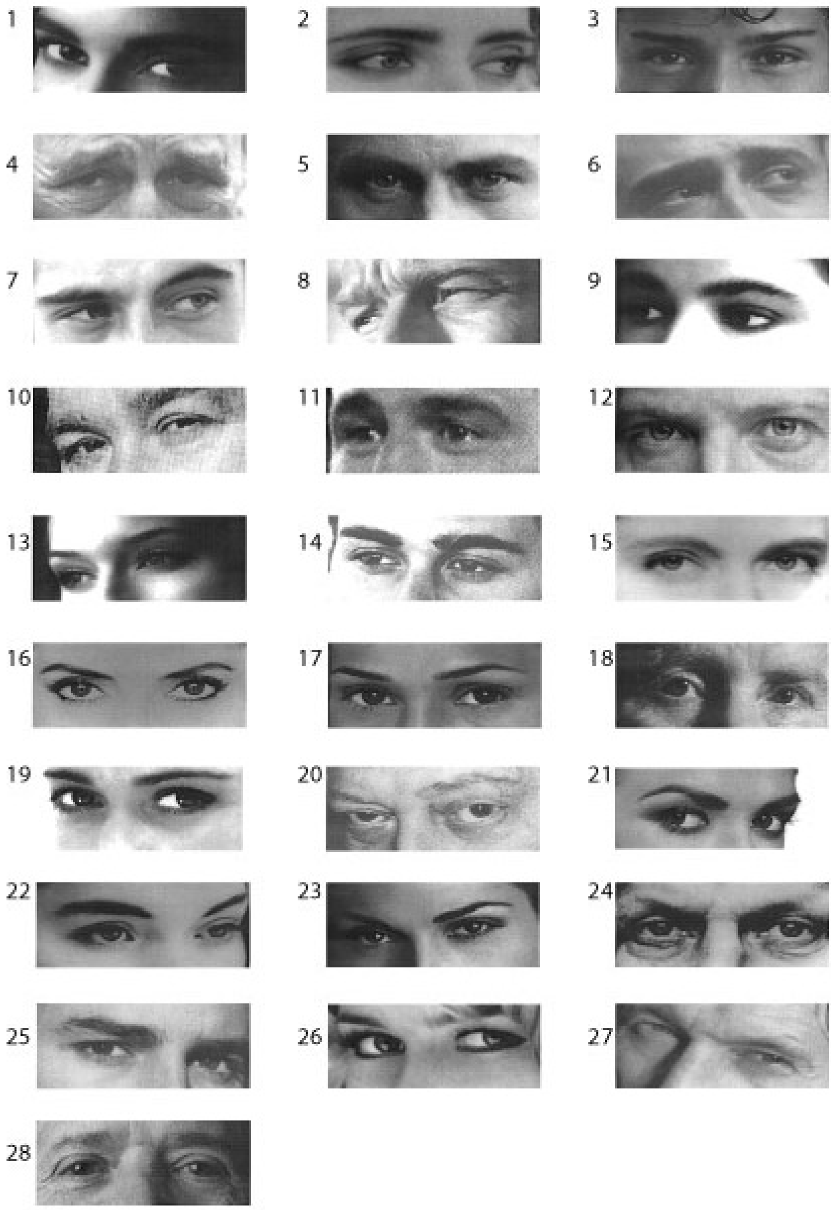

FIGURE 1: 28 stimuli from the child version of the RMET.

The Aspie Quiz. The Aspie Quiz is a questionnaire assessing autistic-like traits on a scale with the answer options: yes, a little, and no. At the time of the study it had 150 items. The Aspie Quiz was validated against Baron-Cohen's Autism Spectrum Quotient and scores correlated 0.83 (Ekblad, 2013).

The primary factor in Aspie Quiz was named neurodiversity factor, and the secondary factor was named neurotypical factor. The neurodiversity factor measures autistic traits whereas the neurotypical factor measures non-autistic traits. For classification of a person into neurotypical and/or neurodiverse, we used the difference between these two scores. A participant with a score difference above or equal to 35 was classified as neurodiverse (ND) (i.e., the autistic-like traits prevail), and someone with a score difference below or equal to 35 was classified as neurotypical (NT). Scores in between were classified as mixed. The cutoff was set to 35 in the investigation phase (Ekblad, 2013) confirming 80% of diagnosed AS/HFA/PDD.[1]

Procedure

The RMET was presented before proceeding to the Aspie Quiz items. This order prevented any bias in performing the task due to one's score along the neurodiverse/neurotypical continuum.

The optimization of the RMET was done in several steps. In the first step, the test was run with the four original emotions from the RMET and an "I don't know" alternative (final version 2; H9, 2013). In the next step, an attempt was made to find the emotions that could best differentiate the neurodiverse and the neurotypical populations from each other. Emotions were collected from previous research in the Aspie Quiz about stims (Pfuhl & Ekblad, 2017) and also included all the original choices, resulting in a list of 85 emotions. First, all the 85 emotions were selectable (final version 2; H10, 2013). Then, 16 emotions, 10 emotions, six emotions, and four emotions (final version 2; H11-H14, 2013) were selected based on which had the largest difference between the populations. The idea was to make the transition to fewer alternatives more gradual so as not to miss out too early on an emotion when alternatives were reduced. Subjects also had the option to respond with "I don't know."

Data analysis

Participants who answered less than 10 items were excluded. We used multinomial logistic regression on the original RMET (Tables 2 and 3 in Appendix), the six alternatives version (Table 4 in Appendix), and the new four alternatives version (Table 5 in Appendix). The ND/NT score was the sole predictor. The factors age and gender were not included in the regression models since they were not of primary interest. We also grouped people based on diagnosis and their ND/NT score for comparison with previous studies. The rate of decline was calculated based on the percentage of "I don't know" responses for each eye expression. Statistical significance is reported with Bonferroni-Holm corrected p-values.

Scores for the RMET were calculated in two ways. First, the mental state descriptor, defined as the correct alternative, was scored as "1," and all other answers as "0" for the child and adult RMET. (See Baron-Cohen et al., 2001a, 2001b.) The average was calculated after excluding "I don't know" answers (i.e., if a participant selected a mental state descriptor in 26 cases, the score was based on the sum divided by 26 to account for that tendency). Guessing would yield 25% correct.

This approach was not possible for the modified versions, which intentionally had no correct mental state descriptor. Instead, a novel method of scoring based on the raw data was used. Based on the multinomial regression analysis, each answer option received a z-score. (See Tables 2–5 in the Appendix.) A positive z-score indicated a preference for this mental state descriptor among NDs, and a negative z-score indicated a preference for that mental state descriptor among NTs. For each participant, we then summed up those z-scores. A participant's total z-score was correlated to his or her ND/NT score.

Results

Demographics

Table 1 gives the number of participants per gender, diagnosis, and Aspie Quiz classification. Age, country of origin, and ancestry differed little between the datasets, so averages for all datasets were calculated. Average age (SD in parenthesis) was 28 (12) years in the ASD group, 30 (12) years in the ND group, 33 (13) years in the mixed group, and 35 (13) years in the NT group. Median age was 24 years in the ASD group, 28 years in the ND group, 30 years in the mixed group, and 32 years in the NT group. Of the participants, 52% were from the USA, 16% from the UK, 7% from Australia, 6% from Canada, and 19% from other countries. 80% of the participants had European ancestry, 5% had Asian, 3% had African, 2% had American-Indian, and 10% didn't give or had mixed or other ancestries.

TABLE 1: Number of participants per gender, diagnosis, and Aspie Quiz classification.

Experiment (dataset)	Total # of participants	Gender	Indicated being diagnosed ASD	ND (score ≥ 35)	Mixed	NT (score ≤ 35)
RMET original adult version (L13) Active 2016-04-24 to 2016-05-08	3,656	male	96	494	477	767
		female	102	923	507	488
RMET original child version (H9) Active 2013-01-13 to 2013-01-20	2,552	male	70	483	364	453
		female	39	446	361	445
RMET 85 options (H10) Active 2013-01-20 to 2013-02-17	7,613	male	240	1,525	1,147	1,260
		female	139	1,676	1,002	1,003
RMET 16 options (H11) Active 2013-02-17 to 2013-02-24	2,039	male	76	355	281	333
		female	40	475	270	325
RMET 10 options (H12) Active 2013-02-24 to 2013-03-02	1,630	male	55	265	211	250
		female	59	380	228	296
RMET 6 options (H13) Active 2013-03-02 to 2013-03-08	1,676	male	61	317	264	257
		female	33	377	203	258
RMET 4 options (H14) Active 2013-03-08 to 2013-03-12 and 2013-08-12 to 2013-08-16	2,516	male	102	456	365	507
		female	44	510	300	378

Performance on the original RMET – adult and child version

Adult version. The average performance of NDs was 66% correct. Persons with a mixture of both traits had on average 72.5% correct. The NT group had 77% correct. People indicating that they had been diagnosed with autism scored lowest with 64% correct. Table 2 (see Appendix) provides the frequency for each of the mental state descriptors among the groups and their z-scores; the mixed group is not presented. Further, we asked all participants to state the clarity of the item. Here too, a group difference was found. The ND/ASD group rated the items as less clear than the NT group, $F(3, 3{,}545) = 68.91$, $p < .0001$, $d = .0117$. The lowest rating was for item 11 with 2.87 (not part of the child version), and item 37 received the highest clarity score of 3.79 across all participants.

Child version. The groups differed significantly in the number of items answered correctly, $F(3, 2{,}439) = 13.00$, $p < .0001$, $d = .063$ with an average percentage of items correct of 69.3% for people indicating being diagnosed with autism, 72.3% of items correct for NDs, 74.1% correct for the mixed group, and 75.3% of items correct for the NT group. Further, the ND group had on average four “I don’t know” responses, whereas the NT group had on average 1.5 “I don’t know” responses. This difference in the number of “I don’t know” responses was statistically significant, $F(3, 2{,}439) = 34.55$, $p < .0001$, $d = .1$. Detailed results are presented in Table 3 (see Appendix).

Scores based on predefined alternatives were calculated for the child and adult RMET. Participants’ performance score (ranging from 0% correct to 100% correct) was used to calculate a correlation to the ND/NT score difference, which resulted in a correlation of $r(2{,}441) = –.14$ for the child RMET and $r(3547) = –.26$ for the adult RMET. Scores based on z-values were calculated and correlated to the ND/NT score difference. This resulted in a correlation of $r(2{,}441) = .29$, $p < .001$, for the child RMET and $r(3{,}547) = .35$, $p < .001$, for the adult RMET. That is, the effect size increased from small to medium by using the z-scores.

Results of the modified RMET

In the new versions emerging from the iterative process, there was no predefined correct answer. Many facial expressions had different semantic labels describing internal states and emotions. Firstly, these versions yielded a similar proportion of “I don’t know” answers between the groups, $p > .05$. Notably, NDs chose a negative connotation (rude, hate, evil, unkind, fake, and plotting) more often compared to NTs, who interpreted the same eye stimuli as much more positive (determined, skeptical, serious, dreamy, and playful). Female eyes were judged as more sexual by NTs, whereas NDs more often chose for the same stimulus staring.

Next, we looked at the mode of each of those mental state descriptors per group. In the six-answer option version, the mode (i.e., the mental state descriptor selected by the majority in each group) was the same for 14 of the 28 items. These mental state descriptors never reached over 50% but were always above chance level of 16.7%. In the new four-answer option version, the most chosen mental state descriptor was the same for all groups in 17 out of the 28 cases, and only once (item 2) did an item reach over 50% in all groups. Tables 4 and 5 provide an overview of the frequency for each mental state descriptor per group (mixed group not shown) for the six-item and four-item version, respectively.

Next, in the new four-answer version, we calculated a new "correct" score based on a) the mental state descriptors the NT group had chosen and b) the mental state descriptors the ND group had chosen. On average, the NT group had 12.9 items "correct," the ASD group had 10.9 items correct, the ND group had 11.2 items correct, and the mixed group had 12.3 items correct. This group difference was significant, $F(3, 2{,}412) = 50.72$, $p < .0001$, $d = .12$. A similar analysis using the ND modes as "correct" yielded 11.4 items correct for the ND group, 12 items correct for the mixed group, 12.1 items correct for the NT group, and 11.1 items correct for the ASD group. This group difference was significant, $F(3, 2{,}414) = 14.28$, $p < .0001$, $d = .07$. Note that 17 items are the same in both scorings.

Again, a better discriminator was the z-score for each participant. Here, the correlation of the z-scores to the ND/NT score yielded far higher coefficients: $r(2{,}414) = .45$ for the six-alternative version and $r(1675) = .41$ for the new four-alternative version (i.e., medium to strong effect sizes).

Discussion

The modified version of the RMET is a sensitive measure for detecting emotion processing differences along the autistic spectrum, and it provided larger effect sizes than using the standard version. The modified RMET is based on a large sample of people scoring high on autistic traits. This is in contrast to the often small sample size of laboratory studies (Kirchner et al., 2011; Walsh et al., 2016; see also Harms et al., 2010). Testing online provides less control than laboratory testing. However, our task had no correct answer per se (cheating was not possible) and was completely voluntary. A recent study used Amazon's Mechanical Turk to see whether non-student samples and doing the test online reproduces findings from laboratory tests (Crump, McDonnell, & Gureckis, 2013). The authors did reproduce the findings from a range of cognitive tasks, despite the necessity to concentrate and watch carefully the stimuli on the screen. They also reported that for certain tasks it was harder to recruit. That is, recruiting is the difficult part not the attention during the task itself.

The mental state descriptors found here reflect how those people interpret the eye stimuli. There is—we have to stress—still no correct answer. However, through the iterative process and use of the summed z-score, the ability of the test to discriminate between persons varying in their degree of autistic traits was markedly improved from a small effect size to a medium effect size. The difference is most pronounced for items where “staring” was an answer option. The ND and NT differed by up to 15%. This was detectable since we let people on the autistic spectrum chose the mental state descriptors, which differed from the selection process of Baron-Cohen et al. (2001a). That is, our iterative process was not biased towards NTs, and it included a gradual process on which people along the entire autistic spectrum participated.

The standard version of the RMET requires that participants answer all items. As such, being unsure forces one to select the best-fitting response despite such a choice being just slightly favored over the other options. We had considered providing a ranking scale first. In other words, participants had to indicate from 1 (highest agreement) to 4 (lowest agreement) the order of the mental state descriptors. However, we traded that consideration off against a single answer option to ensure more participants would score all items. That is, we included the option to ‘opt out’ by selecting “I don’t know.” That is to say, we offered an alternative to guessing (Dolnicar & Grün, 2014).

We found a higher prevalence of “I don’t know” responses among people with autism and the ND group compared to the NT group. We think this tendency is due to perceived difficulty rather than motivation. Firstly, the “I don’t know” responses were not clustered toward the end of a participant’s responses. Secondly, the clarity scoring also indicated that the higher the autistic traits, the less clear the stimuli were perceived. Thirdly, when offering new mental state descriptors, the “I don’t know” responses dropped.

Sawyer, Williamson, and Young (2014) found that people with Asperger’s disorder opted less for not responding compared to controls. When not sure which emotion is displayed, NTs withhold their decision more often than people diagnosed with autism. The difference was more in the metacognitive ability rather than emotion processing per se. Furthermore, persons with autism have been found to be less sensitive to rewards or points (Damiano et al., 2012) because they engaged more in a physical effort task without considering the outcomes. Harms et al. (2010) suggested that previously found inconsistent results in emotion processing in autism are due to the use of compensatory mechanisms. It has been suggested that time pressure may increase the group difference. But here the literature is also mixed.

Walsh et al. (2016) found an impairment in emotion recognition when using very short stimulus exposure and no time limit to respond. Tracy et al. (2011) found similarly fast and accurate responses among children and adolescents with or without a diagnosis of ASD. Here, we imposed no incentives, points, or time pressure. We were genuinely interested in how people perceived the stimuli. Under optimal conditions and given enough cognitive resources, performance of people with autism can be similar to that of NTs (See, for example, Harms et al. 2010.) For example, in the RMET, additional cognitive effort spent, as seen in longer response times, yielded no performance difference (Miu et al., 2012). By offering the option of “I don’t know” we already found a reduced group difference.

Negativity bias

The modified RMET yielded a very specific bias—a negativity bias. There was a preference for negative mental state descriptors among the ND and ASD groups. This might be due to a more pessimistic view of the world. Indeed, there is a high prevalence of depression among people with autism (Ghaziuddin, GHaziuddin, & Greden, 2002). It remains to be seen whether this bias is unique for autism or overlaps with depression or schizophrenia (Leppänen, 2006; Trémeau, 2006).

Tasks that measure bias can tell us more about the processing strategy. Such tasks may also allow testing recent computational models of autism. Van de Cruys et al. (2014) explain the symptoms of autism with a failure of sensory attenuation. That is, there is an overweighting of sensory information and reduced influence of contextual and prior information. This tendency results in large prediction errors and constant surprise. It also requires the need to update and relearn because the experience (prior information) is too weak. But such permanent discrepancy between expectation and sensory information can lead to frustration and the desire for predictability. Can one apply this model of failure in sensory attenuation to the RMET? On the one hand, the RMET is a task of judging mental states without any contextual information provided. On the other hand, everyone has preconceived notions about the human nature in general, a worldview on how nice or nasty people are. This viewpoint is shaped by experience. Neurotypicals who are devoid of depression have a positive worldview and are more optimistic. Hence, they see things more positively (Isaacowitz, 2005; Peters, Vieler, & Lautenbacher, 2016). People with depression, though, do perform worse on the RMET (Richman & Unoka; 2015).

Neurodiverse people may also have a less positive worldview. For people with autism, this negative worldview may come from the often frustrating unpredictability experienced in social situations (i.e., they often err and prefer to avoid such experience subsequently) (van de Cruys et al., 2014). Indeed, we found that more NDs and people with autism than NTs preferred not to rate some of the RMET stimuli.

Social situations are complex situations requiring a lot of cognitive effort, especially if there is a lack of experience. But this is a vicious circle for people with autism. The prediction errors they make due to overweighting of sensory information and too little influence of experience (or prior worldview) is frustrating rather than rewarding (made the correct prediction). Instead of increasing the experience and hence reducing the prediction error, they chose to avoid exposure to social situations. This decision, in turn, ensures a steady arousal when again exposed to social situations. Dalton et al. (2005) found that people with autism were aroused by looking at faces, but this arousal had a negative valence, seen as hypoactivation in the amygdala. Thus, the pressure felt to have to look at a face and judge what it expresses is not preferred. Here, this negative attitude toward looking at faces was measured and quantified. It is not a failure in emotion processing capabilities; it is a bias in choosing more negative attributes, especially staring in the ND group. Responding with staring may express an avoidance behavior (i.e., not long looking at the eyes).

Conclusion

Our Internet survey yielded a robust difference among neurodiverse and neurotypical persons in the interpretation of eyes. This robust difference is based on studying a large population and avoiding dichotomization of subjects into typical developing (healthy) humans and people with autism. That is, the score difference is a continuous measurement of autistic-like traits. The modified RMET distinguishes better between people with autism and people scoring high on neurodiverse and neurotypical traits, respectively. Further, the modified test reduced guessing because we allowed participants not to answer. These small procedural differences, no time pressure, and an ‘opt out’ option yielded important differences in emotion processing among participants. People with autism and neurodiverse people selected staring and other negatively connoted attributes for faces often, whereas neurotypicals showed no negativity bias. This negativity bias may lead to a vicious circle, avoiding social contacts instead of training to become better at social interaction. Since experience plays a crucial role in emotion processing (Jack & Schyns, 2015; Hartshorne & Germine 2015), interactions between neurotypicals and people with autism should be more frequent, and more time and patience should be afforded to them. Providing feedback on one’s performance in emotion recognition has the potential to increase social skills and confidence in social situations (Rice et al., 2015; Sawyer et al., 2014).

References

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders. 5th ed., text rev*. Washington, DC: APA.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. *Journal of Autism and Developmental Disorders*, *28*(5), 369–392. doi:10.1023/A:1026048320785

Baron-Cohen, S., Wheelwright, S., Hill, J. et al. (2001a). 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 Child Psychology Psychiatry*, *42*(2), 241–251. doi:10.1111/1469-7610.00715

Baron-Cohen, S., Wheelwright, S., Scahill, V., Lawson, J., & Spong, A. (2001b). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorders*, *5*, 47–78.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLOS One*, *8*(3), e57410. doi:10.1371/journal.pone.0057410

Dalton K. M., Nacewicz B. M., Johnstone T. et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. doi:10.1038/nn1421

Damiano, C. R., Aloi, J., Treadway, M., Bodfish, J. W., & Dichter, G. S. (2012). Adults with autism spectrum disorders exhibit decreased sensitivity to reward parameters when making effort-based decisions. *Journal of Neurodevelopmental Disorders*, *4*(1), 13. doi:10.1186/1866-1955-4-13

Dolnicar, S., & Grün, B. (2014). Including don’t know answer options in brand image surveys improves data quality. *International Journal of Market Research*, *56*(1), 33–50. doi:10.2501/IJMR-2013-043

Ekblad, L. (2013). Autism, personality, and human diversity: Defining neurodiversity in an iterative process using Aspie Quiz. *SAGE Open*, *3*(3), 1–14. doi:10.1177/2158244013497722

Ekman, P. (2003). *Emotions Revealed: Understanding Faces and Feelings*. London: Weidenfeld & Nicolson.

Ghaziuddin, M., Ghaziuddin, N., & Greden, J. (2002). Depression in persons with autism: Implications for research and clinical care. *Journal of Autism and Developmental Disorders*, *32*(4), 299–306. doi:10.1023/a:1016330802348

Griffin, E., & Pollak, D. (2009). Student experiences of neurodiversity in higher education: Insights from the BRAINHE project. *Dyslexia*, *15*(1), 23–41. doi:10.1002/dys.383

Grove, R., Baillie, A., Allison, C. et al. (2013). Empathizing, systemizing, and autistic traits: Latent structure in individuals with autism, their parents, and the general population controls*. Journal of Abnormal Psychology*, *122*(2), 600–609. doi:10.1037/a0031919

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9(*10), 1218–1220. doi:10.1038/nn1770

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. doi:10.1007/s11065-010-9138-6

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443. doi:10.1177/0956797614567339

Isaacowitz, D. M. (2005). The gaze of the optimist. *Personality and Social Psychology Bulletin*, *31*(3), 407–415. doi:10.1177/0146167204271599

Jaarsma, P., & Welin, S. (2012). Autism as a natural human variation: Reflections on the claims of the neurodiversity movement. *Health Care Analysis, 20*, 20–30. doi:10.1007/s10728-011-0169-9

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, *25*(14), R621–R634. doi:10.1016/j.cub.2015.05.052

Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2013). Deficit, difference, or both? Autism and neurodiversity. *Developmental Psychology*, *49*(1), 59–71. doi:10.1037/a0028353

Kirchner, J. C., Hatri, A., Heekeren, H. R., & Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *Journal of Autism and Developmental Disorders*, *41*(2), 158–167. doi:10.1007/s10803-010-1032-9

Leppänen, J. M. (2006). Emotional information processing in mood disorders: A review of behavioral and neuroimaging findings. *Current Opinion in Psychiatry*, *19*(1), 34–39. doi:10.1097/01.yco.0000191500.46411.00

Lundström, S., Chang, Z., Råstam, M. et al. (2012). Autism spectrum disorder and autistic-like traits. *Archive of General Psychiatry*, *69*(1), 46–52. doi:10.1001/archgenpsychiatry.2011.144

Miu, A. C., Pana, S. E., & Avram, J. (2012). Emotional face processing in neurotypicals with autistic traits: Implications for the broad autism phenotype. *Psychiatry Research*, *198*, 489–494. doi:10.1016/j.psychres.2012.01.024

Nuske, H. J., Vivanti, G., & Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognition & Emotion*, *27*(6), 1042–1061. doi:10.1080/02699931.2012.762900

Olsson, A., & Ochsner, K. N. (2007). The role of social cognition in emotion. *Trends in Cognitive Sciences*, *12*(2), 65–71. doi:10.1016/j.tics.2007.11.010

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety*, *27*(12), 1135–1142. doi:10.1002/da.20755

Peskin, M., & Newell, F. N. (2004). Familiarity breeds attraction: Effects of exposure on the attractiveness of typical and distinctive faces. *Perception, 33*(2), 147–157. doi:10.1068/p5028

Peters, M. L., Vieler, J. S. E., & Lautenbacher, S. (2016). Dispositional and induced optimism lead to attentional preference for faces displaying positive emotions: An eye-tracker study. *Journal of Positive Psychology*, *11*(3), 258–269. doi:10.1080/17439760.2015.1048816

Pfuhl, G., & Ekblad, L. (2017, December 18). Ekblad Pfuhl Autistic self-stimulatory behaviors (stims): Useless repetitive behaviors or nonverbal communication? Retrieved from psyarxiv.com/4xyc2

Rice, L. M., Wall, C. A., Fogel, A., & Shic, F. (2015). Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD. *Journal of Autism and Developmental Disorders*, *45*(7), 2176–2186. doi:10.1007/s10803-015-2380-2

Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: Meta-analysis. *British Journal of Psychiatry*, *207*(6), 483–489.

doi:10.1192/bjp.bp.114.152108

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristics of autism. *Developmental Sciences*, *8*, 444–458. doi:10.1111/j.1467-7687.2005.00433.x

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141.

Sawyer, A. C. P., Williamson, P., & Young, R. (2014). Metacognitive processes in emotion recognition: Are they different in adults with Asperger's Disorder? *Journal of Autism and Developmental Disorders*, *44*(6), 1373–1382. doi:10.1007/s10803-013-1999-0

Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia*, *51*, 98–105. doi:10.1016/j.neuropsychologia.2012.11.013

Tracy, J. L., Robins, R. W., Schriber, R. A., & Solomon, M. (2011). Is emotion recognition impaired in individuals with autism spectrum disorders? *Journal of Autism and Developmental Disorders*, *41*(1), 102–109. doi:10.1007/s10803-010-1030-y

Trémeau, F. (2006). A review of emotion deficits in schizophrenia. *Dialogues in Clinical Neuroscience*, *8*(1), 59–70.

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorder*, *43*, 1517–1526. doi:10.1007/s10803-012-1695-5

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. doi:10.1037/a0037665

Walsh, J. A., Creighton, S. E., & Rutherford, M. D. (2016). Emotion perception or social cognitive complexity: What drives face processing deficits in autism spectrum disorder? *Journal of Autism and Developmental Disorders*, *46*(2), 615–623. doi:10.1007/s10803-015-2606-3

APPENDIX: https://psykologisk.no/wp-content/uploads/2017/12/4_e18_tables_2-5.pdf

Citation

Pfuhl, G., & Ekblad, L. (2017). Optimizing the RMET to measure

bias not performance differences. *Scandinavian Psychologist*, *4*, e18. https://doi.org/10.15714/scandpsychol.4.e18
Abstract

Optimizing the RMET to measure bias not performance differences

Background: Human social interaction and communication is complex. Sending a verbal message is often accompanied by intonations, facial expressions, grimaces, and body postures. Nonverbal signals are potentially open for misinterpretation. One popular test for assessing the interpretation of facial expressions is the “Reading the Mind in the Eyes” Test (RMET). This test has been used to relate Theory of Mind abilities along the autistic spectrum. However, this test was normed on a small sample of students, and answers were coded binary as either correct or wrong.
Methods: We recruited from various forums, blogs, and personal websites over 10,000 people. To assess autistic traits (neurodiversity), we used the Aspie Quiz, which agrees well with the AQ test (Ekblad, 2013). Importantly, we included an “I don’t know” answer option. Further, participants could freely indicate which emotion they read in the eyes. Applying an iterative process, we derived alternative mental state descriptors.
Results and conclusion: This optimized RMET increased the ability to differentiate between people with few or many autistic traits, respectively. By using logistic regression, the test is able to measure difference in bias, not just performance. We found a pronounced negativity bias among people who scored high on many autistic traits. This bias may contribute to a vicious circle of avoiding social interactions.

Keywords: emotion, facial expressions, high-functioning autism, neurodiversity, nonverbal.

Author affiliations: Gerit Pfuhl – Department of Psychology, University of Tromsø – The Artic University of Norway, Tromsø, Norway; Leif Ekblad – Independent researcher, Eslöv, Sweden.

Contact information: Gerit Pfuhl, Department of Psychology, University of Tromsø, 9037 Tromsø, Norway.
Email: gerit.pfuhl@uit.no.

Received: June 8, 2017. **Accepted:** November 12, 2017.
Published: December 22, 2017.

Language: English.

Competing interests: The authors report no conflict of interest. The authors alone are responsible for the contents and writing of this paper.

Acknowledgement: The authors would like to thank R. Biegler for valuable discussions. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

This is a peer-reviewed paper.