scientific reports

OPEN

Multifaceted empathy differences in children and adults with autism

Jennifer M. Quinde-Zlibut[1,2✉], Zachary J. Williams[1,2,3,4], Madison Gerdes[5], Lisa E. Mash[6], Brynna H. Heflin[7] & Carissa Cascio[1,2,8,9]

Although empathy impairments have been reported in autistic individuals, there is no clear consensus on how emotional valence influences this multidimensional process. In this study, we use the Multifaceted Empathy Test for juveniles (MET-J) to interrogate emotional and cognitive empathy in 184 participants (ages 8–59 years, 83 autistic) under the robust Bayesian inference framework. Group comparisons demonstrate previously unreported interaction effects between: (1) valence and autism diagnosis in predictions of emotional resonance, and (2) valence and age group in predictions of arousal to images portraying positive and negative facial expressions. These results extend previous studies using the MET by examining differential effects of emotional valence in a large sample of autistic children and adults with average or above-average intelligence. We report impaired cognitive empathy in autism, and subtle differences in emotional empathy characterized by less distinction between emotional resonance to positive vs. negative facial expressions in autism compared to neurotypicals. Reduced emotional differentiation between positive and negative affect in others could be a mechanism for diminished social reciprocity that poses a universal challenge for people with autism. These component- and valence- specific findings are of clinical relevance for the development and implementation of target-specific social interventions in autism.

Autistic individuals have long been erroneously presumed to lack the ability and/or desire to relate to others[1]. Currently, a clinical diagnosis of autism is characterized by marked social and communication deficits, along with restrictive and repetitive behaviors[2]. Extant theories on the origins of these autistic traits focus on genes[3], brain activity[4], and psychological features[5]. The 'Theory of Mind'(TOM) account of autism[5], proposed that autism is characterized by a core deficit in the understanding of others' intentions, knowledge, and emotions. The field of autism research has largely moved beyond such "core deficit" models[6–8] in favor of more holistic models that address the complexities of interactions between individuals with and without autism. However, the TOM theory set the tone for much of the early empathy research in the autistic population. Considered a social 'glue' that allows for successful human relationships, empathy is an essential component of social interactions with which individuals on the autism spectrum reportedly struggle[9–11].

Various studies investigating empathy differences in autism have reported conflicting results, including intact empathic physiological responses[12,13], but reduced or dampened self-reported empathy[13,14]. Notably, empathy research has been slow to develop a well-defined operationalized definition of "empathy" or its various components, and there remains no global consensus on best practices for the measurement of this construct[15,16]. Thus, early autism studies employing differing approaches to measure this inherently complex neuropsychological concept often reported findings that were not always based on the same empathy subconstructs[11,17].

Empathy has traditionally been considered to include the distinct constructs of **emotional empathy**, the ability to *share* another person's feelings and **cognitive empathy**, the capacity to *understand* such feelings[18,19]. A third, now increasingly differentiated component, **empathic concern,** has recently gained consideration as part of the multifaceted nature of empathy[20–22]. Empathic concern is characterized by greater *others*-oriented relational mentalities and encompasses emotional sentiments (e.g. sympathy and compassion) towards someone else's experience[11]. Though the distinction between emotional empathy and empathic concern is subtle, emotional

¹Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA. ²Frist Center for Autism and Innovation, Vanderbilt University, Nashville, TN, USA. ³Medical Scientist Training Program, Vanderbilt University School of Medicine, Nashville, TN, USA. ⁴Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. ⁵Graduate Program in Criminology and Justice Policy, Northeastern University, Boston, MA, USA. ⁶San Diego Joint Doctoral Program in Clinical Psychology, San Diego State University/University of California, San Diego, CA, USA. ⁷Graduate Program in Clinical Psychology, Florida International University, Miami, FL, USA. ⁸Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. ⁹Vanderbilt Kennedy Center, Vanderbilt University, Nashville, TN, USA. ✉email: Jennifer.m.quinde@vanderbilt.edu

empathy describes *sharing* another's feelings, which involves *self*-orientation (i.e., "mirroring" emotion), while empathic concern does not require having the same feeling but being aware of and concerned about another's feeling In fact, this capacity for self-other distinction is considered to be crucial and integral to the empathic experience[23].

Many studies have now explored these 3 constructs both separately, and to a lesser extent, in some combination within a single autism sample. This is notable given the growing evidence that these processes typically work together to form a unified percept[23]. Thus it would be difficult, for example, to speculate on the underpinnings of emotional empathy differences and how these may relate to autism symptomatology, without assessing these separate but related constructs simultaneously. To this end, empathy differences between autistic and neurotypical (NT) individuals have been primarily assessed using multidimensional self-report tools like the Interpersonal Reactivity Index[24] (IRI), Empathy Quotient[14] (EQ), and Questionnaire of Cognitive and Affective Empathy[25] (QCAE) (for a review, see Song et al., 2019). While these well-validated self-report measures have informative potential, they are not without limitations. Primarily, they are subject to social desirability biases[26], and subjective variability in interpretation that can ultimately result in under- or over-reporting.

To address the limitiations of widely-used self-report measures, Dziobek and colleagues (2008) developed the Multifaceted Empathy Test (MET) as a performance-based measure of empathy. The MET is a computerized task that assesses both emotional and cognitive empathy in response to a series of emotionally charged facial expressions. According to the creators of the MET, this task is more ecologically valid than self-report measures of empathy, as it does not rely as heavily on the level of insight an individual has into their own emotions. The MET was also designed to mitigate potential social desirability biases in empathy ratings by asking participants to rate their level of arousal in response to each stimulus, which ostensibly serves as an implicit measure of emotional empathy. This measure has contributed meaningfully to the study of empathy in autism, as large effects of diagnostic group on the cognitive empathy subcomponent of the MET[26–28] indicate that group differences in empathy demonstrated using questionnaire measures are not simply due to differences in emotional self-awareness or social desirability biases.

The ability to perceive and distinguish positive from negative emotional situations is an important feature of empathy and drives our tendency to engage in prosocial behaviors. This process is also mediated by the limbic system and allows for intentional and motivated reactions to our ever-changing environment[29]. Indeed, face perception studies suggest that both limbic neural activity[4,30,31], and gaze patterns[32] are differentially responsive to faces of negative versus positive emotion. For example, faces displaying negative emotion typically elicit stronger activation in the left amygdala[4], a larger early negative component (N300)[30], and increased scanning of the eyes[32]. Though there is precedent for a link between the amygdala and valence-specific emotion processing patterns in autism, the role of emotion in autism is not yet clear[33]. Further, the capacity to distinguish positive from negative emotions in others is important for the ability to produce appropriate resonant behaviors that ultimately help build social rapport[34].

Although some prior work has indicated that emotional valence may moderate diagnostic group differences in empathy[27,35], studies on this topic are scarce and inconsistent, and it remains unclear whether these differences relate to autism symptomatology. Studies directly comparing cognitive and affective empathy in response to positive and negative faces in autism reported impaired emotion recogniton of positive but not negative emotions[33,35], while another reported emotional empathisizing differences in autism for negative but not positive emotions[27]. Notably, participant samples from these studies were predominantly male, providing little to add to reports of sex differences in empathy from theoretical[36] and performance-based studies alike, and thus warrant further exploration and confirmation in single cohorts that include females.

The present study

In light of this emerging but incomplete picture of empathy in autism, the present study explores empathy as a multifaceted construct that is potentially modulated by valence in individuals with autism. To extend previous findings, an age-appropriate adapted version of the MET, the MET-Juvenile[28], consisting of 16 positive and 16 negative emotional images (decreasing testing time to accommodate younger participants), was administered to a large sample with a wide age range. Notably, though component-specific age effects on empathy have been reported[37,38], there is no clear agreement on how these capacities may be related to maturation[11]. For example, while some studies suggest that empathy increases with age[37], others suggest that older adults have lower empathic abilities[39], see Sun et al., 2017 for a review[40]. Thus, participants from a broad age range were included in the present study thereby ensuring grounds for exploration of any age or age group related effects in empathy responses. Group differences in cognitive and emotional empathy were assessed globally in the context of the hypothesis that empathy is a complex, developmental process involving emotional and cognitive components that are moderated by stimulus valence.

Results

Demographic variables. In total, 184 participants were included in the analysis sample: 45 children and adolescents (35 male; mean age = 11.53) with autism, and 43 neurotypical children and adolescents (34 male; mean = 11.86), 38 adults with autism (21 male; mean age = 27.65), and 58 neurotypical adults (36 male; mean age = 32.69). The two diagnostic groups were approximately equivalent in terms of sex ratio (OR = 1.091, 95% CrI [0.583, 2.003], BF_{10} = 0.177), although they significantly differed with respect to age (d = 0.457 [0.146, 0.773], BF_{10} = 22.2), full-scale IQ (d = 0.532 [0.224, 0.846], BF_{10} = 94.4), VIQ (d = 0.657 [0.326, 1.010]), and SRS-2T-scores (d = − 2.896 [− 3.478, − 2.349]). Additionally, the NT group had numerically higher PIQ scores on average than the autistic group, although the Bayes factor indicated only "anecdotal" evidence in favor of a group difference (d = 0.284 [− 0.021, 0.588], BF_{10} = 1.59).

Best fit predictor	OR [95% CrI]	BF_{10}	BF_{inc}	BF_{ROPE}	P_{ROPE}
Cognitive empathy					
Diagnosis (AUTISM)	**0.726 [0.587, 0.906]**	**6.854**		**3.677**	**0.00**
Sex (F)	1.113 [0.917, 1.352]	0.179		0.110	0.00
Age group (Child/Adolescent)	**0.604 [0.462, 0.795]**	**58.271**		**37.80**	**0.00**
Valence (Negative)	0.545 [0.267, 1.071]	1.633		1.609	0.00
Verbal IQ Z-score	1.126 [1.037, 1.218]	2.949	2.760	0.213	0.00
Best fit predictor	**β [95% CrI]**	**BF_{10}**	**BF_{inc}**	**BF_{ROPE}**	**P_{ROPE}**
Emotional empathy					
Diagnosis (AUTISM)	− 0.059 [− 0.255, 0.142]	0.121		0.057	0.594
Sex (F)	− 0.073 [− 0.275, 0.138]	0.135		0.072	0.544
Age group (Child/Adolescent)	0.025 [− 0.175, 0.214]	0.103		0.042	0.671
Valence (Negative)	**− 0.647 [− 0.80, − 0.49]**	**1.61×10^{-6}**		**4.89×10^{5}**	**0.000**
Diagnosis × Valence	0.188 [0.066, 0.307]	3.521	5.59	1.075	0.076
Arousal empathy					
Diagnosis (AUTISM)	− 0.045 [− 0.232, 0.145]	0.104		0.046	0.653
Sex (F)	− 0.024 [− 0.234, 0.179]	0.110		0.048	0.640
Age group (Child/Adolescent)	0.084 [− 0.118, 0.289]	0.143		0.079	0.517
Valence (Negative)	− 0.137 [− 0.342, 0.071]	0.253		0.160	0.345
Age group x Valence	**− 0.206 [− 0.390, − 0.149]**	**89.300**	**147**	**16.3**	**0.005**

Table 1. Bayesian model analysis results. Practically significant predictors for each empathy component are shown in bold.

Hierarchical Bayesian models. Model selection procedures indicated that cognitive empathy scores were best predicted by a model including all baseline predictors and VIQ score ($P(\mathcal{M}_{BL+VIQ}|Data) = 0.405$; $R^2_{Bayes} = 0.257$ [0.242, 0.272]). The posterior inclusion probability of VIQ was relatively high ($P_{VIQ} = 0.734$), although the inclusion Bayes factor for VIQ did not meet the a priori threshold of 3 ($BF_{inc} = 2.76$). Inclusion Bayes factors also provided substantial evidence *against* the inclusion of all interaction terms, PIQ, and SRS-2 T-score as predictors of cognitive empathy (Table 1).

In the best-fitting model, autism diagnosis was associated with a practically significant reduction in performance on cognitive empathy (CE) trials (OR = 0.726, 95% CrI [0.587, 0.906], $BF_{ROPE} = 3.70$). This effect of group on CE is depicted in Fig. 1A. Moreover, age group was an even larger predictor of performance, with children across both diagnostic groups displaying significantly lower emotion recognition accuracy than adults (OR = 0.604 [0.462, 0.795], $BF_{ROPE} = 37.80$). Neither female sex (OR = 1.113 [0.917, 1.352], $BF_{ROPE} = 0.110$) or negative valence (OR = 0.545 [0.267, 1.071], $BF_{ROPE} = 1.609$) significantly predicted performance on the cognitive empathy trials, although there was insufficient evidence to conclude that valence was unrelated to the chance of a correct response. Lastly, although higher VIQ significantly predicted higher performance on cognitive empathy trials (OR = 1.126 [1.037, 1.218] per standard deviation increase in VIQ), there was substantial evidence that this effect was too small to be practically significant ($BF_{ROPE} = 0.213$).

When predicting emotional empathy, model selection procedures indicated that the most likely model included all baseline predictors as well as the interaction between diagnosis and valence ($P(\mathcal{M}_{BL+Dx*Valence}|Data) = 0.452$; $R^2_{Bayes} = 0.505$ [0.491, 0.518]). Inclusion Bayes factors supported only the inclusion of diagnosis × valence interaction term in the final model ($BF_{inc} = 5.59$), along with the exclusion of most other predictors (Table 1). The best-fitting model displayed a large effect of valence, with lower reports of shared feelings for negative emotions (β = − 0.647 [− 0.800, − 0.489], $BF_{ROPE} = 4.89 \times 10^5$), as well as small and practically insignificant main effects of diagnostic group (β = − 0.059 [− 0.255, 0.142], $BF_{ROPE} = 0.057$), sex (β = − 0.073 [− 0.275, 0.138], $BF_{ROPE} = 0.072$), and age group (β = 0.025 [− 0.175, 0.214], $BF_{ROPE} = 0.042$). The interaction between diagnosis and valence was statistically significant, although the ROPE Bayes factor was equivocal with regard to its practical significance (β = 0.188 [0.066, 0.307], $BF_{ROPE} = 1.075$). Although both diagnostic groups reported empathizing more with positive than negative emotions, this difference was larger in the NT group (d = 0.647 [0.489, 0.800]) than the autism group (d = 0.459 [0.315, 0.614]). This interaction is depicted in Fig. 1B.

When predicting ratings of arousal, the best-fitting model was found to include all baseline predictors as well as the interaction between age group and valence ($P(\mathcal{M}_{BL+Age*Valence}|Data) = 0.677$; $R^2_{Bayes} = 0.466$ [0.451, 0.480]). Inclusion Bayes factors demonstrated strong support for the inclusion of the age group × valence interaction ($BF_{inc} = 147$), as well as the exclusion of all other potential predictors (Table 1). Coefficients in the best-fitting model indicated practically insignificant effects of autism diagnosis (β = − 0.045 [− 0.232, 0.145], $BF_{ROPE} = 0.046$), age (β = 0.084 [− 0.118, 0.289], $BF_{ROPE} = 0.079$), sex (β = − 0.024 [− 0.234, 0.179], $BF_{ROPE} = 0.048$), and emotional valence (β = − 0.137 [− 0.342, 0.071], $BF_{ROPE} = 0.160$). However, these effects were qualified by a statistically and practically significant interaction between age and valence (β = − 0.269 [− 0.390, − 0.149], $BF_{ROPE} = 16.3$). Across both diagnostic groups, children reported significantly higher arousal ratings for positive stimuli compared to negative stimuli (d = 0.403 [0.195, 0.609]), whereas no significant effect of valence was seen in adult participants (d = 0.137 [− 0.071, 0.342]). This interaction effect is illustrated in Fig. 2.
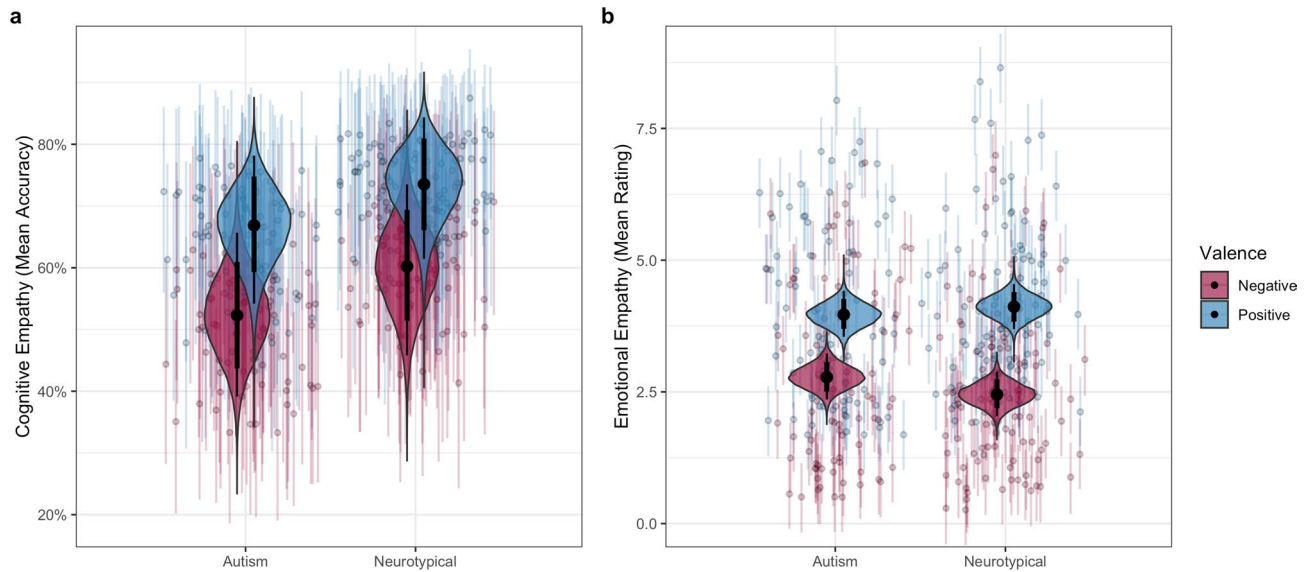
Figure 1. Group comparisons for (**a**) mean accuracy in emotion recognition for cognitive empathy surveys, and (**b**) mean resonance rating for emotional empathy surveys. Figure generated in R[41].
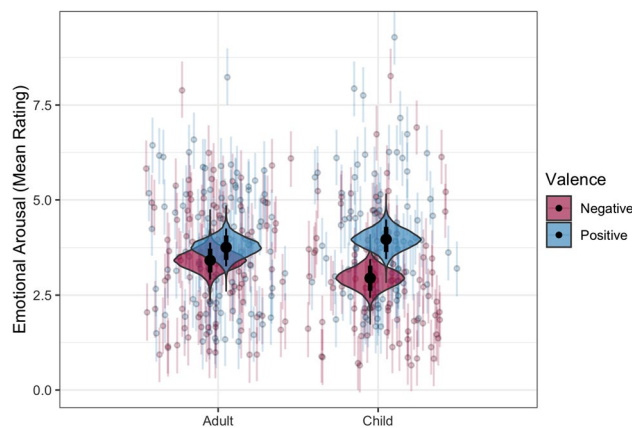

Figure 2. Age group comparisons for mean emotional arousal ratings to positive and negative emotionally charged facial expressions. Figure generated in R[41].

Discussion

Using a multidimensional approach, this study replicates previous findings of practically equivalent levels of emotional empathy between autistic and NT groups despite a significant effect of diagnosis on cognitive empathy. These findings also complement Dziobek's et al[26], by extending the comparison to test for valence effects. To our knowledge, only one other group has assessed emotional valence effects in autism using the MET[27], which reported that autistic adolescents displayed reduced emotional empathy for facial expressions of negative emotions compared to controls, a difference that was not present for positive emotions. Notably, the findings of Mazza and colleagues differ from those of the current study, which found that ratings of emotional empathy were approximately equal between groups for negatively-valenced stimuli, although neurotypical individuals provided higher average ratings for positively-valenced stimuli. The present study further extends Mazza's et al[27], results by assessing age group effects and utilizing a substantially larger child/adolescent autism sample. With regards to age, our Bayesian analysis indicates that children and adolescents across both diagnostic groups had greater difficulty with emotion recognition than adults, although main effects of age group on self-ratings of arousal and emotional empathy were small and practically insignificant.

Valence effects in our analyses were only significant in models predicting emotional empathy and arousal. For emotional empathy, the significant interaction between valence and diagnostic group indicated that positive facial expressions elicited greater emotional resonance than negative facial expressions across both groups, but that this difference was greater in neurotypicals compared to autistic individuals. In the arousal empathy Bayesian Model Averaging (BMA) analysis, the small and practically insignificant main effects for emotional valence and age group were qualified by a strong and practically significant interaction between these predictors. That

is, the increase in arousal elicited by positive facial expressions relative to negative ones was more pronounced in our child and adolescent group compared to the adult group.

In the current study, we observed higher ratings for shared feelings to positive versus negative facial expressions that interacted with diagnostic group but not with age. This may reflect a neurotypical advantage for better self-other distinction for negative valence compared to autistic individuals that persists across age. In other words, contrary to previous reports of unequivocally intact emotional empathy in autism[26,42], emotional empathy may actually differ somewhat in autism when valence is considered. The contrast between greater resonance to positive emotions compared to negative emotions across neurotypical development may also be partially attributable to the effect of positivity-biases for ambiguous emotions. In some neurotypical adults, ambiguous facial expressions like 'surprise' tend to get rated as positive, an effect that seems to be moderated by the regulatory influence of prefrontal cortex[43]. Thus, the lower degree of separation between emotional resonance ratings to positive and negative emotions in autism may also be reflecting neural differences in top-down emotion processing.

Supporting the cognitive vs. affective empathy dichotomy, Cox et al[18] reported distinct intrinsic functional connectivity (FC) dynamics in healthy adult brains for self-reported cognitive and emotional empathy. Using a difference score (cognitive – emotional), this group found emotional empathy dominance (negative scores) to be associated with stronger functional connectivity between social-emotional brain regions like the amygdala and ventral anterior insula[18]. By contrast, cognitive empathy dominance (positive scores) correlated with greater FC in areas like the ventral anterior insula and superior temporal sulcus, both of which have been implicated in social-cognitive processes[18,44,45]. This distinction also becomes relevant when considering the implications of valence-specific evidence for empathy differences in autism like an impairment in the ability to understand when social experiences warrant resonating with negative emotions, a feature that has been linked to aggressive behaviors[46–48].

The present study had various strengths and limitations to consider. Amongst the strengths, our study included a wide age range in which we were able to replicate previous findings from research conducted primarily in adolescent and adult samples, and extend findings to a larger group of younger individuals. Additionally, use of robust statistical approaches allowed us to report on the practical significance of our findings as well as to test for equivalence between diagnostic groups on these empathy measures. Ours is among the first to combine several important approaches in a single study: examining empathy multidimensionally, testing both children and adults, and examining positive and negative emotional valence separately. These advances facilitate a significant step forward in our understanding of empathy in autism.

This study was limited by exclusion of individuals with low IQ (FSIQ < 70) as a way of ensuring proper understanding of the task instructions. Future efforts to address this limitation should include objective psychophysiological empathy measures that do not require explicit behavioral response or abstract thinking. An additional limitation is the fact that the current study did not include measurements of trait alexithymia[49,50], which has been proposed to mediate the relationship between diagnoses of autism and performance on tasks tapping multiple facets of empathy[17,51,52]. Another limitation is that we did not test for empathic concern and prosocial behavior propensity. In retrospect, we would have also liked to complement our findings by including at least one additional validated empathy questionnaire to compare and enrich our understanding on multimodal empathy assessment.

Within the context of social neuroscience, our measure of emotional empathy aligns more with indices of emotional accuracy which may be more relevant for generating social rapport compared to emotional concern and thus its relationship with prosocial behavior remains unexamined in this multidimensional way. Changes in the wording for indexing emotional empathy were modelled after the MET-J and pursued under the premise that the language used in the original MET may be too abstract for younger participants[28]. The implications of this change became apparent only after careful review of recent literature pointing to the differences between empathic accuracy and empathic concern[11]. Thus, future research designs ought to take this distinction into consideration and address as many of empathy components as possible within a single cohort in order to obtain a more holistic insight and understanding of empathy.

It should also be noted that a growing literature further improving ecological validity by utilizing dyadic conversational interactions (typically involving richer but less emotionally charged stimulation than the static but highly emotional faces used in the MET) has described a "double empathy" problem with assumptions made from empirical research in empathy and autism[53–55]. These reports critique the use of neurotypical people as a reference point and conclude that differences are diminished when autistic individuals are partnered with one another, suggesting that an empathy deficit (at least in the relatively emotionally neutral context of an initial conversation with a new person) should be reframed as a feature of the interaction rather than the individual[55]. Finally, The MET utilizes static emotional faces, which limits the ecological validity of the task. Following the example of recent studies that use dyadic interactions to characterize social differences in autism[55–57], future studies should continue to balance ecological validity with experimental control under a framework that does not assume a neurotypical reference point.

The self-report nature in the design of the MET, albeit task-based, still confers some susceptibility to social desirability biases. This is one potential explanation for the increased emotional relatedness feelings reported by our neurotypical adult sample, who may be more impacted by social desirability bias[58]. The specificity of this difference to facial expressions depicting positive emotions and not negative emotions, however, warrants further investigation. Future efforts should aim to collect more implicit measures of emotional empathy such as skin conductance, spontaneous facial expressions, and neural measures of empathic response. Within the context of conflated definitions and methodologies employed in previous research, the resulting confusion from conflicting academic reports on the autistic 'empathy deficit' has not been without effect on the autistic population. Overgeneralizations on this matter have been described as 'unwarranted' and 'dehumanizing' by autistic self-advocates[59]. Because of this negative impact potential, future work should also take great care to develop

methodologies based on clearly defined empathy concepts as well as reporting and interpreting results using a more humane framework.

In conclusion, the current study finds component-specific differences in empathy. We report impaired cognitive empathy in autism, a valence by diagnostic group effect in emotional empathy, and valence by age group effect for arousal empathy. In both diagnostic groups, emotion recognition as measured by our cognitive empathy survey was significantly greater in adults than in children and adolescents. Improved emotion recognition by adulthood across groups may potentially reflect training effects of social interventions in autism and a larger repertoire of references for positive interactions in neurotypicals. Further investigations would benefit from an analysis that accounts for potential confounds like co-occuring mental health symptoms or training effects on the ability of individuals with autism to recognize emotions. These findings suggest that emotional stimuli, specifically hedonically negative stimuli, may be actively recruiting separate perceptual pathways that are differentially altered in autism. Better elucidating component-specific impairments is crucial for the generation of target-specific social interventions to improve empathic capacities and social outcomes.

Methods

Participants.   Participants in this study were recruited from the community through posters and social media postings as well as from a pool of participants in previous larger and longitudinal lab studies who consented to be recontacted. Over the course of the recruitment period, our stance on study design for case–control comparisons has evolved; early in this period, the goal was to achieve a "clean" ASD sample and thus excluded for most co-occurring psychiatric and developmental conditions within the ASD group. As time has gone by, we have seen that these samples are not representative of the population, and have opted to be more inclusive and attempt to control for co-occurring conditions in both groups analytically instead, though we have been slower to adopt this for control groups. Thus, this sample represents a blend as our approach has evolved.

Co-occurring mental health symptoms in our autism group were screened for using the ASEBA School Age (6–18) & ASEBA Adult (18–59) forms. Overall, we collected ASEBA data on 90/184 participants (24 ASD, 66 NT). Of these, 11 ASD participants endorsed clinically significant symptoms of depression, 8 endorsed clinically significant symptoms of anxiety, and 9 endorsed clinically significant symptoms of ADHD. Of the 66 NT participants with ASEBA data, 3 endorsed clinically significant symptoms of depression, and 1 participant endorsed clinically significant symptoms of ADHD.

Individuals with autism and co-occurring ADHD, anxiety, or depression were included, while those with other psychiatric diagnoses within the past five years or co-occurring neurogenetic syndromes were excluded. Stimulant medication use was screened for before study participation but there was no exclusionary criteria based on this. Of our 184 participants, 36 endorsed 'Yes' to taking medications and of these, 11 (all ASD) reported using stimulant medication. In our adult subgroup, ten autistic adults (26%) reported taking antidepressant medications, and in our children/adolescents subgroup, ten autistic participants' parents (22%) reported their child taking antidepressant medications.

Adults.   Thirty-eight adult participants with autism (21 male; mean age = 27.65) and 58 neurotypical (NT) adults (36 male; mean age = 32.69) were included in the study. All adult participants were between the ages of 18 and 59 years and achieved full scale IQ scores of ≥ 70 as measured by the Wechsler Abbreviated Scale of Intelligence-Second Edition[60] (WASI-II). Autism diagnoses were confirmed by the clinical judgment of a licensed psychologist specializing in the assessment of autism, supported by research-reliable administration of the Autism Diagnostic Observation Schedule-2[61] (ADOS-2).

Exclusion criteria for both groups included the presence of other neurological and genetic disorders, non-autism related sensory impairments (e.g., uncorrected visual or hearing impairments), and substance/alcohol abuse or dependence during the past two years. Further, individuals in the NT group were excluded if they had reported a previous psychiatric history, cognitive or sensory impairment, use of psychotropic medications, or clinically elevated scores on the Social Communication Questionnaire[62] (SCQ Total score > 15).

Children/adolescents.   Forty-five autistic children/adolescents (35 male; mean age = 11.53) and 43 neurotypical (NT) children/adolescents (34 male; mean age = 11.86) were included in the study. All child/adolescent participants were between the ages of 8 and 17 years and achieved full scale IQ (FSIQ) scores of ≥ 70 as measured by the WASI-II. Autism diagnoses were confirmed by the clinical judgment of a licensed psychologist specializing in the assessment of autism, supported by research-reliable administration of the ADOS-2 and, when available, parent interviews (n = 30) that included algorithm items from the Autism Diagnostic Interview, Revised[63] (ADI-R).

Exclusion criteria for children/adolescents were similar to those for adults with some additional considerations. Mainly, for children and adolescents, behavior and co-occurring psychiatric conditions were screened for using parent and guardian reports.

Ethical considerations.   The study was conducted in accordance with the Declaration of Helsinki and all participants were compensated $20 per hour of their time following each session. Written informed consent or assent forms were signed by all participants, while informed consent was obtained from parents or guardians of minors. All methods and procedures were approved by the Institutional Review Board for human subjects at Vanderbilt University Medical Center and carried out in accordance with relevant guidelines and regulations on ethical human research.
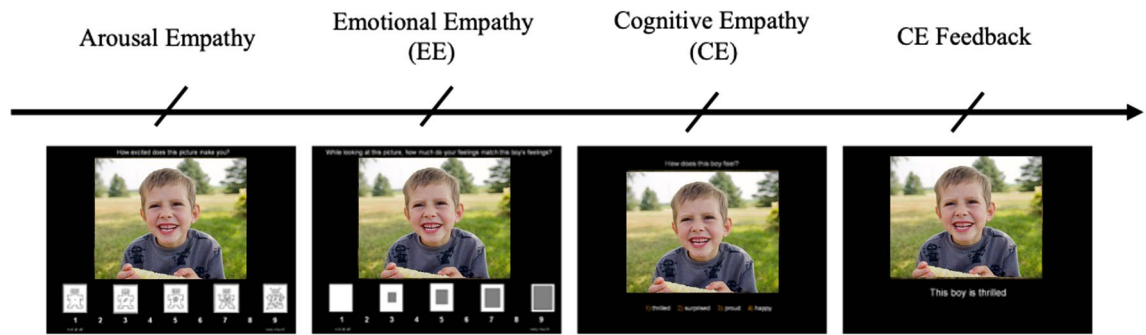
Figure 3. Example trial for 'thrilled' emotion depicted on MET-J task. The surveys read 1) "How excited does this picture make you" (implicit emotional empathy/arousal empathy), 2) "While looking at the picture, how much do your feelings match the boy's feelings (emotional empathy; EE, and 3) "How does this boy feel?" (cognitive empathy; CE). Cognitive empathy emotion label options are 1) thrilled, 2) surprised, 3) proud, 4) happy. The slides for this trial example were designed on PowerPoint using a free-use stock image from the Canva.com image database.

Measures. The Social Responsiveness Scale–Second Edition[64] (SRS-2) was used to measure autistic traits dimensionally across the full sample. Adult participants in both diagnostic groups completed the SRS-2 adult self-report form, whereas parents or guardians of children/adolescents in both groups completed the analogous caregiver-report questionnaire, the SRS-2 School Age form. To facilitate comparison across the different groups, the SRS-2 total scores were converted to T- scores ($M = 50$, $SD = 10$).

Empathy. Empathy was assessed multi-dimensionally using an adapted version of the Multifaceted Empathy Test, the MET-J[28], a validated performance-based test that separates cognitive and emotional empathy based on responses to emotional faces presented with context in the background. The original MET includes 50 still images depicting emotionally charged facial expressions of 25 positive (e.g., joy, happiness) and 25 negative (e.g., sadness, anger) emotions. The adapted MET-J version used in the present study included only 16 images each for positive and negative valence. The photographs are taken from the International Affective Picture System[65] (IAPS), a well-validated database of photographs designed for standardized emotion and attention testing. On each trial, participants viewed an emotional image and were first asked to rate their level of arousal, followed by explicit emotional empathy ratings, and a cognitive empathy (i.e., emotion recognition) multiple choice question. Figure 3 depicts an example trial on the task, recreated using a free-use stock image from the Canva.com image database.

As described by Dziobek et al. (2008), to minimize demands of self-reflection and thereby also mitigate social desirability bias, we included an implicit assessment of emotional empathy by asking participants to rate how calm/aroused the emotional stimuli made them feel using the Self-Assessment Manikin (SAM). The SAM is a visual-analogue scale providing scores ranging from 1 (very calm) to 9 (very aroused). Thus for each picture, participants were asked (1) "How excited does this picture make you" (implicit emotional empathy; subsequently described as arousal empathy); (2) "While looking at the picture, how much do your feelings match the X's feelings" (emotional empathy; EE) measured on a visual Likert scale (1–9); and finally (3) "How does this X feel?" (cognitive empathy; CE). Here an "X" represents the noun used to describe the individuals (boy/girl/man/woman) in the image, who varied across trials. Each trial ended with a final presentation of the emotional stimulus that provided feedback for the cognitive empathy question by displaying the correct emotion label from among the four choices. Note that this order and wording for EE surveys are slightly different from the original MET and MET-J, which provided feedback on CE surveys prior to presenting explicit emotional empathy surveys. We adapted this order to ensure that EE and arousal responses were made as reflexively as possible to the perceived emotion upon initial presentation, rather than being adjusted based on CE feedback. All stimuli were presented as slides of variable duration (ad libitum) in random order on a black screen.

Statistical analyses. Differences in demographics (e.g., age, sex, VIQ, PIQ) and SRS-2 scores were compared between the autism and NT groups within a Bayesian framework. When the outcome of interest was categorical (e.g., correct or incorrect emotion recognition), group differences were examined using a Bayesian analogue of the Pearson chi-squared test[66,67]. When the outcome of interest was a continuous variable (e.g., age), we examined mean differences using a Bayesian analogue of the Welch (unequal-variances) t-test[68]. Effect sizes from each of these tests (i.e., Cohen's d and the odds ratio [OR]) were summarized as the posterior median and 95% highest-density credible interval (CrI). Additionally, for all group comparisons, evidence for or against the point null hypothesis ($\mathcal{H}_0$; i.e., no differences between groups) was quantified with a Bayes factor[67,69], defined as the ratio of how likely the data are under the alternative hypothesis ($\mathcal{H}_1$; i.e., the difference between group is nonzero) divided by how likely the data are under $\mathcal{H}_0$. In concordance with widely-used guidelines on Bayes factor interpretation[70,71], we considered BF_{10} values > 3 as indicating substantial evidence for $\mathcal{H}_1$, BF_{10} values < 0.333 as indicating substantial evidence for $\mathcal{H}_0$, and BF_{10} values between 0.333 and 3 as providing inconclusive and only "anecdotal" evidence for $\mathcal{H}_0$ or $\mathcal{H}_1$. All group comparisons were performed in the R statistical computing

platform using open-source R code written by author ZJW[72]. Additional details on the specifics of the models underlying Bayesian t-test and Chi-squared test analogues are presented in "Supplemental Methods".

To determine the effects of various predictor variables on arousal, emotional, and cognitive empathy while controlling for possible covariates, we used R[41] to analyze the data at the single-trial level using hierarchical Bayesian modeling. Trial-level MET data for arousal, emotional, and cognitive empathy were analyzed using (generalized) linear mixed effects models ([G]LMEMs), which allowed us to model the correlations between responses derived from the same participants as well as the same stimuli[73]. LMEMs were used to model arousal and emotional empathy, as the 9-item scale used to derive these outcomes had enough points to be approximated as a continuous variable[74]. However, we used a logistic GLMEM to model cognitive empathy, as individual trial data from this part of the task consisted of binary "correct/incorrect" responses. The baseline [G]LMEM for each MET-derived outcome included fixed effects of age group (child vs. adult), sex, autism diagnosis, and emotional valence (positive vs. negative), as well as random intercepts for participant and stimulus (see example below for CE, Eq. 1). Random slopes were also included in this baseline model for all subject-level predictors, allowing the effects of age group, sex, and autism status to vary by stimulus. The decision to treat age as categorical in the BMA was driven by the finding that performance on the CE task increased with age throughout childhood, reaching an asymptote at approximately age 18–20, thereby indicating a difference between children and adults rather than a true linear age trend.

$$\text{Cognitive Empathy Dx Group + Age Group + Sex + Valence + (1|Participant) +(1 + Stimulus |Dx Group + Sex + Age Group)}$$

(1)

For each of the three outcomes, we additionally determined if several other predictors beyond the baseline model contributed to task performance, including the two-way and three-way interactions between age, diagnosis, and valence; verbal IQ (VIQ); performance IQ (PIQ); and overall level of autistic traits (SRS-2 T-score). In order to determine whether any given predictor should be added to the baseline model, we fit candidate models that included all combinations of potential predictors ($n = 40$ potential models including the baseline). Then, using bridge sampling[75], we calculated marginal likelihood of each candidate model, deriving posterior model probabilities in a manner equivalent to the process of Bayesian model averaging[76]. The model with the highest posterior probability was considered the final model for each outcome. Using these model weights, we also computed inclusion Bayes factors[76] (BF_{inc}), allowing us to determine the degree of evidence for or against the inclusion of each predictor in the model. Inclusion Bayes factors are interpretable in the same manner as BF_{10}, with $\mathcal{H}_0$ being the exclusion of the variable from the model and $\mathcal{H}_1$ being the inclusion of the variable in the model.

Once the final model for each outcome was selected, we additionally tested all regression slopes in a Bayesian framework, using the 95% CrI to determine whether each slope was likely to be nonzero in magnitude. If the full 95% CrI excluded zero, we rejected the point null hypothesis that the effect was exactly zero. However, as this point null hypothesis is always false at the population level[77], we also tested these effects for *practical* significance[78]. The Bayesian framework allows for a probabilistic view of the parameter estimates so that we can infer whether an effect is practically meaningful at the population level. This is done by defining a region of practical equivalence[79] (ROPE), an interval of parameter values considered small enough to be equivalent to zero in practice (in this case $\beta_{Std} = [-0.1, 0.1]$ for linear models and $e^{\beta_{Std}} = [0.909, 1.10]$ for logistic models). Evidence both for and against the true parameter value falling within the ROPE can be quantified by calculating a ROPE Bayes factor (BF_{ROPE}), defined as the odds of the prior parameter distribution falling within the ROPE divided by the odds of the posterior effect size distribution falling within the ROPE[80,81]. These Bayes factors can be interpreted on the same scale as previously discussed for BF_{10} *and* BF_{inc}[70,71]. In the case that a parameter was nonzero or a given variable was included within the final model but the BF_{ROPE} value was smaller than 0.333, we considered this variable as not predicting the MET outcome of interest to a practically meaningful extent. Lastly, in order to assess the predictive power of the final model, we calculated the Bayesian R^2 coefficient proposed by Gelman et al.[82].

All Bayesian [G]LMEMs were fit in Stan using the *brms* R package[83,84] with weakly informative priors, including Normal(0, 1) priors on all (standardized) regression slopes and intercept terms, as well as default half-Student $t_3(0, 2.5)$ priors on the standard deviation of each random slope or intercept term. Model parameters were estimated via Markov chain Monte Carlo (MCMC) using the No U-turn Sampler implemented in Stan[85], with posterior distributions of each parameter estimated using 21,000 post-warmup MCMC draws from seven Markov chains (14,000 in cases where missing data were present). Parameter summaries from these posterior distributions were operationalized as the posterior median and the 95% CrI. Convergence for each model was confirmed by examination of Markov chain trace plots, as well as values of the Gelman–Rubin (Rubin & Gelman, 1992) convergence diagnostic < 1.01. Missing data were handled using five-fold multiple imputation based on the random forest imputation algorithm implemented in the *missForest* R package[86,87].

Data availability
The datasets generated for this study can be found in the National Database for Autism Research (NDAR) repository (https://nda.nih.gov).

References

1. Kanner, L. Autistic disturbances of affective contact.pdf. *Nerv. Child* **2**, 217–250 (1943).
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. vol. 5 (American Psychiatric Association, 2013).

3. Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc. Natl. Acad. Sci. USA* **104**, 12831–12836 (2007).
4. Wright, P. *et al.* Dissociated responses in the amygdala and orbitofrontal cortex to bottom-up and top-down components of emotional evaluation. *Neuroimage* **39**, 894–902 (2008).
5. Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a 'theory of mind'?. *Cognition* **21**, 37–46 (1985).
6. Astle, D. E. & Fletcher-Watson, S. Beyond the Core-deficit hypothesis in developmental disorders. *Curr. Dir. Psychol. Sci.* **29**, 431–437 (2020).
7. Gernsbacher, M. A. & Yergeau, M. Empirical failures of the claim that autistic people lack a theory of mind 102–118 (2019).
8. Happé, F., Ronald, A. & Plomin, R. Time to give up on a single explanation for autism. *Nat. Neurosci.* **9**, 1218–1220 (2006).
9. Baron-Cohen, S. & Wheelwright, S. EQ-an investigation of adults with AS or HFautism and normal sex differences. *J. Autism Dev. Disord.* **34**, (2004).
10. Harmsen, I. E. Empathy in autism spectrum disorder. *J. Autism Dev. Disord.* **1**, 3939–3955 (2019).
11. Song, Y., Nie, T., Shi, W., Zhao, X. & Yang, Y. Empathy impairment in individuals with autism spectrum conditions from a multidimensional perspective: A meta-analysis. *Front. Psychol.* **10**, (2019).
12. Blair, R. J. R. Psychophysiological responsiveness to the distress of others in children with autism. *Personal. Individ. Differ.* https://doi.org/10.1016/S0191-8869(98)00154-8 (1999).
13. Trimmer, E., McDonald, S. & Rushby, J. A. Not knowing what i feel: Emotional empathy in autism spectrum disorders. *Autism* **21**, 450–457 (2017).
14. Baron-Cohen, S. & Wheelwright, S. The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* **34**, 163–175 (2004).
15. Frankel, R. M. The evolution of empathy research: Models, muddles, and mechanisms. *Patient Educ. Couns.* **100**, 2128–2130 (2017).
16. Preston, S. D. & de Waal, F. B. M. Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.* **25**, 1–20 (2002).
17. Fletcher-Watson, S. & Bird, G. Autism and empathy: What are the real links?. *Autism* **24**, 3–6 (2020).
18. Cox, C. L. *et al.* The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. *Soc. Cogn. Affect. Neurosci.* **7**, 727–737 (2012).
19. Shamay-Tsoory, S. G., Aharon-Peretz, J. & Perry, D. Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* **132**, 617–627 (2009).
20. Jordan, M. R., Amir, D. & Bloom, P. Are empathy and concern psychologically distinct?. *Emotion* **16**, 1107–1116 (2016).
21. Zhao, X., Li, X., Song, Y. & Shi, W. Autistic traits and prosocial behaviour in the general population: Test of the mediating effects of trait empathy and state empathic concern. *J. Autism Dev. Disord.* **49**, 3925–3938 (2019).
22. Zickfeld, J. H., Schubert, T. W., Seibt, B. & Fiske, A. P. Empathic concern is part of a more general communal emotion. *Front. Psychol.* **8**, (2017).
23. Håkansson Eklund, J. & Summer Meranius, M. Toward a consensus on the nature of empathy: A review of reviews. *Patient Educ. Couns.* **104**, 300–307 (2020).
24. Davis, M. H. A multidimensional approach to individual differences in empathy. *JSAS Cat. Sel. Doc. Psychol.* **10**, (1980).
25. Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M. & Völlm, B. A. The QCAE: A questionnaire of cognitive and affective empathy. *J. Pers. Assess.* **93**, 84–95 (2011).
26. Dziobek, I. *et al.* Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *J. Autism Dev. Disord.* **38**, 464–473 (2008).
27. Mazza, M. *et al.* Affective and cognitive empathy in adolescents with autism spectrum disorder. *Front. Hum. Neurosci.* **8**, 1–6 (2014).
28. Poustka, L. *et al.* Dissociation of cognitive and emotional empathy in adolescents with autism spectrum disorders. *Kindh. Entwickl.* **19**, 177–183 (2010).
29. Berridge, K. C. Affective valence in the brain: Modules or modes?. *Nat. Rev. Neurosci.* **20**, 225–234 (2019).
30. Dawson, G., Webb, S. J., Carver, L., Panagiotides, H. & McPartland, J. Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Dev. Sci.* **7**, 340–359 (2004).
31. Moradi, A., Mehrinejad, S. A., Ghadiri, M. & Rezaei, F. Event-related potentials of bottom-up and top-down processing of emotional faces. *Basic Clin. Neurosci.* **8**, 27–36 (2017).
32. de Wit, T. C. J., Falck-Ytter, T. & von Hofsten, C. Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces. *Res. Autism Spectr. Disord.* **2**, 651–659 (2008).
33. Ashwin, C., Chapman, E., Colle, L. & Baron-Cohen, S. Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. https://doi.org/10.1080/17470910601040772 (2007).
34. Luberto, C. M. *et al.* A systematic review and meta-analysis of the effects of meditation on empathy, compassion, and prosocial behaviors. *Mindfulness* **9**, 708–724 (2018).
35. Rueda, P., Fernández-Berrocal, P. & Baron-Cohen, S. Dissociation between cognitive and affective empathy in youth with Asperger Syndrome. *Eur. J. Dev. Psychol.* **12**, 85–98 (2015).
36. Singer, T. & Lamm, C. The social neuroscience of empathy. *Ann. N. Y. Acad. Sci.* **1156**, 81–96 (2009).
37. Peterson, C. Theory of mind understanding and empathic behavior in children with autism spectrum disorders. *Int. J. Dev. Neurosci.* **39**, 16–21 (2014).
38. Schulte-Rüther, M. *et al.* Age-dependent changes in the neural substrates of empathy in autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* **9**, 1118–1126 (2014).
39. Phillips, L. H., MacLean, R. D. J. & Allen, R. Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **57**, 526–530 (2002).
40. Sun, B., Luo, Z., Zhang, W., Li, W. & Li, X. Age-related differences in affective and cognitive empathy: self-report and performance-based evidence. *Aging Neuropsychol. Cogn.* **25**, 655–672 (2018).
41. R Core Team. *R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing.* (2021).
42. Baron-Cohen, S. *Zero Degrees of Empathy: A New Theory of Human Cruelty and Kindness.* (Penguin/Allen Lane, 2011).
43. Tottenham, N. *et al.* The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Res.* **168**, 242–249 (2009).
44. Fan, Y. T., Chen, C., Chen, S. C., Decety, J. & Cheng, Y. Empathic arousal and social understanding in individuals with autism: Evidence from fMRI and ERP measurements. *Soc. Cogn. Affect. Neurosci.* **9**, 1203–1213 (2014).
45. Gu, X. *et al.* Autonomic and brain responses associated with empathy deficits in autism spectrum disorder. *Hum. Brain Mapp.* **36**, 3323–3338 (2015).
46. Eisenberg, N. & Miller, P. A. The relation of empathy to prosocial and related behaviors. *Psychol. Bull.* **101**, 91–119 (1987).
47. Miller, P. A. & Eisenberg, N. The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychol. Bull.* **103**, 324–344 (1988).
48. Pouw, L. B. C., Rieffe, C., Oosterveld, P., Huskens, B. & Stockmann, L. Reactive/proactive aggression and affective/cognitive empathy in children with ASD. *Res. Dev. Disabil.* **34**, 1256–1266 (2013).
49. Bagby, R. M., Parker, J. D. A. & Taylor, G. J. Twenty-five years with the 20-item Toronto Alexithymia Scale. *J. Psychosom. Res.* **131**, 109940–109940 (2020).
50. Williams, Z. J. & Gotham, K. O. Improving the measurement of alexithymia in autistic adults: A psychometric investigation and refinement of the twenty-item Toronto Alexithymia Scale. *Mol. Autism* **12**, 20 (2021).

51. Bird, G. & Cook, R. Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Transl. Psychiatry* **3**, e285–e285 (2013).
52. Bird, G. & Viding, E. The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neurosci. Biobehav. Rev.* **47**, 520–532 (2014).
53. Milton, D. E. M. On the ontological status of autism: The 'double empathy problem'. *Disabil. Soc.* **27**, 883–887 (2012).
54. Edey, R. *et al.* Interaction takes two: Typical adults exhibit mind-blindness towards those with autism spectrum disorder. *J. Abnorm. Psychol.* **125**, 879–885 (2016).
55. Morrison, K. E. *et al.* Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism* **24**, 1067–1080 (2020).
56. Qualls, L. R. & Corbett, B. A. Examining the relationship between social communication on the ADOS and real-world reciprocal social communication in children with ASD. *Res. Autism Spectr. Disord.* **33**, 1–9 (2017).
57. Rolison, M. J., Naples, A. J. & McPartland, J. C. Interactive social neuroscience to study autism spectrum disorder. *Yale J. Biol. Med.* **88**, 17–24 (2015).
58. Kirchner, J. C., Schmitz, F. & Dziobek, I. Brief report: Stereotypes in autism revisited. *J. Autism Dev. Disord.* **42**, 2246–2251 (2012).
59. Cohen-Rottenberg, R. Unwarranted Conclusions and the Potential for Harm: My Reply to Simon Baron-Cohen. *Autism and Empathy* https://autismandempathyblog.wordpress.com/unwarranted-conclusions-and-the-potential-for-harm-my-reply-to-simon-baron-cohen/ (2013).
60. Wechsler, D. *WASI-II: Wechsler Abbreviated Scale of Intelligence.* (Pearson, 2011).
61. Lord, C. *et al. Autism Diagnostic Observation Schedule (ADOS-2): Manual.* (Western Psychological Services, 2012).
62. Rutter, M., Bailey, A. & Lord, C. *Social Communication Questionnaire.* (Western Psychological Services, 2003).
63. Lord, C., Michael, R. & Couteur, A. L. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24**, (1994).
64. Constantino, J. N. Social responsiveness scale, Second Edition (SRS-2) [Assessment Instrument]. 2015–2015 (2011).
65. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Cent. Study Emot. Atten.* 39–58 (1997).
66. Gûnel, E. & Dickey, J. Bayes factors for independence in contingency tables. *Biometrika* **61**, 545–557 (1974).
67. Jamil, T., Ly, A., Morey, R. D., Love, J. & Wagenmakers, M. M. E. Default "Gunel and Dickey" Bayes factors for contingency tables. 638–652. https://doi.org/10.3758/s13428-016-0739-8 (2017).
68. Kruschke, J. K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).
69. Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H. & Grasman, R. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognit. Psychol.* **60**, 158–189 (2010).
70. Harold, J. *Theory of Probability.* (Clarendon Press; Oxford University Press, 1961).
71. Wagenmakers, E. J., Wetzels, R., Borsboom, D. & van der Maas, H. L. J. Why psychologists must change the way they analyze their data: The case of Psi: Comment on Bem (2011). *J. Pers. Soc. Psychol.* **100**, 426–432 (2011).
72. Williams, Z. J. No Title. *BayesianTools R Funct. Perform Gen.-Purp. Bayesian Estim. Hypothesis Test. Using Brms R.* https://doi.org/10.13140/RG.2.2.26089.31845 (2020).
73. Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**, 390–412 (2008).
74. Rhemtulla, M., Brosseau-Liard, P. É. & Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **17**, 354–373 (2012).
75. Gronau, Q. F. *et al.* A tutorial on bridge sampling. *J. Math. Psychol.* **81**, 80–97 (2017).
76. Hinne, M., Gronau, Q. F., van den Bergh, D. & Wagenmakers, E.-J. A conceptual introduction to Bayesian model averaging. *Adv. Methods Pract. Psychol. Sci.* **3**, 200–215 (2020).
77. Cohen, J. The earth is round (p < .05). *Am. Psychol.* **49**, 997–1003 (1994).
78. Kirk, R. E. Practical significance: A concept whose time has come. *Educ. Psychol. Meas.* **56**, 746–759 (1996).
79. Kruschke, J. K. & Liddell, T. M. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon. Bull. Rev.* **25**, 178–206 (2018).
80. Makowski, D., Ben-Shachar, M. & Lüdecke, D. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **4**, 1541–1541 (2019).
81. Makowski, D., Ben-Shachar, M. S., Chen, S. H. A. & Lüdecke, D. Indices of effect existence and significance in the Bayesian framework. *Front. Psychol.* **10**, 1–14 (2019).
82. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. R-squared for Bayesian Regression models. *Am. Stat.* **73**, 307–309 (2019).
83. Bürkner, P. C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, (2017).
84. Bürkner, P. C. Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411 (2018).
85. Hoffman, M. D. & Gelman, A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
86. Stekhoven, D. J. & Bühlmann, P. Missforest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
87. Stekhoven, D. J. missForest: Nonparametric missing value imputation using random forest. (2013).

Acknowledgements

Author contributions

J.M.Q.-Z. and C.C. conceptualized the study, and wrote the main manuscript text. M.G., L.E.M., and B.H.H. collected and organized data. J.M.Q.-Z. and Z.J.W. performed background reading, data cleaning and management, and statistical analysis. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-98516-5.

Correspondence and requests for materials should be addressed to J.M.Q.-Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.