

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/132080/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Leno, Virginia Carter, Chandler, Susie, White, Phillipa, Yorke, Isabel, Charman, Tony, Jones, Catherine R. G., Happe, Francesca, Baird, Gillian, Pickels, Andrew and Siminoff, Emily 2021. Associations between theory of mind and conduct problems in autistic and nonautistic youth. *Autism Research* 14 (2) , pp. 276-288. 10.1002/aur.2346

Publishers page: <http://dx.doi.org/10.1002/aur.2346>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Associations between theory of mind and conduct problems in autistic and nonautistic youth

Virginia Carter Leno , Susie Chandler , Phillipa White , Isabel Yorke , Tony Charman , Catherine R. G. Jones , Francesca Happé , Gillian Baird, Andrew Pickles , and Emily Simonoff 

Many autistic young people exhibit co-occurring behavior difficulties, characterized by conduct problems and oppositional behavior. However, the causes of these co-occurring difficulties are not well understood. Impairments in theory of mind (ToM) are often reported in autistic individuals and have been linked to conduct problems in nonautistic individuals. Whether an association between ToM ability and conduct problems exists in autistic populations, whether this association is similar between individuals who are autistic versus nonautistic, and whether these associations are specific to conduct problems (as opposed to other domains of psychopathology) remains unclear. ToM ability was assessed using the Frith–Happé Triangles task in a pooled sample of autistic ($N = 128$; mean age 14.78 years) and nonautistic youth ($N = 50$; mean age 15.48 years), along with parent-rated psychiatric symptoms of conduct problems, hyperactivity/inattention and emotional problems. Analyses tested ToM ability between autistic versus nonautistic participants, and compared associations between ToM performance and conduct problems between the two groups. Where no significant group differences in associations were found, the pooled association between ToM and conduct problems was estimated in the combined sample. Results showed no evidence of moderation in associations by diagnostic status, and an association between poorer ToM ability and higher levels of conduct problems, hyperactivity/inattention and emotional problems across the total sample. However, these associations became nonsignificant when adjusting for verbal IQ. Results provide support for theoretical models of co-occurring psychopathology in autistic populations, and suggest targets for intervention for conduct problems in autistic youth. *Autism Res* 2020, 00: 1–13. © 2020 The Authors. *Autism Research* published by International Society for Autism Research and Wiley Periodicals LLC.

Lay Summary: Many young people with autism spectrum disorder show co-occurring behavior problems, but the causes of these are not well understood. This paper found an association between difficulties recognizing what others think and intend (so-called “theory of mind”) in a simple animated task, and emotional and behavioral problems in autistic and nonautistic young people. However, a substantial part of this association was explained by individual differences in verbal ability. These findings may have implications for intervention efforts to improve young people’s mental health.

Keywords: ADHD; autism; cognition; conduct problems; QUEST; SNAP; theory of mind

BACKGROUND

Autistic youth are at increased risk of exhibiting co-occurring behavior/conduct problems (Gjevik, Eldevik, Fjæran-Granum, & Sponheim, 2011; Simonoff et al., 2008), characterized by aggression, temper outbursts, and severe noncompliance (referred to as conduct problems throughout the current manuscript for consistency, although we acknowledge different terms are used in the literature). One approach to understanding the heightened prevalence of

co-occurring conduct problems in autistic populations is to focus upon the cognitive profile associated with a diagnosis of autism spectrum disorder (ASD). Traditionally these specific cognitive difficulties have been conceptualized as potential mechanisms that may underpin the core symptoms of social communication difficulties and restricted, repetitive behaviors. However, far less research has considered how these difficulties may also be important in understanding co-occurring psychopathology in autistic individuals. Identifying individual characteristics that

From the Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK (V.C.L., A.P.); Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK (S.C., P.W., I.Y., E.S.); Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK (T.C.); South London and Maudsley NHS Foundation Trust (SLaM), London, UK (T.C., E.S.); School of Psychology, Cardiff University, Cardiff, UK (C.R.G.J.); Social, Genetic and Developmental and Psychiatry Research Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK (F.H.); Evelina London Children’s Hospital, Guy’s and St Thomas’ NHS Foundation Trust, London, UK (G.B.); Maudsley Biomedical Research Centre for Mental Health, London, UK (A.P., E.S.)

Received February 14, 2020; accepted for publication May 24, 2020

Correspondence Virginia Carter Leno, Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, 16 De Crespigny Park, London SE5 8AF, UK.

Email: virginia.carter_leno@kcl.ac.uk

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

Published online 00 Month 2020 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.2346

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

predict poor mental health in autistic populations is an important first step in signposting appropriate targets for intervention.

One potentially relevant cognitive risk factor for conduct problems, which is often studied in autistic populations, is difficulties in theory of mind (ToM), characterized by problems understanding the mental states (e.g., beliefs, intentions) of others. While group differences are often present, in that autistic youth show poorer performance on ToM tasks as compared to matched non-autistic youths (see Happé & Conway, 2016 for a review; Salter, Seigal, Claxton, Lawrence, & Skuse, 2008), not all autistic individuals fail tests of ToM (Scheeren, de Rosnay, Koot, & Begeer, 2013). It has been argued that these difficulties in ToM underpin the core symptoms of social impairments (Jones et al., 2018); however, the role of ToM as a risk factor for conduct problems is a relatively understudied area in autistic populations. One study with autistic youth found poorer performance on computerized ToM tasks was associated with higher levels of self-reported aggression (Pouw, Rieffe, Oosterveld, Huskens, & Stockmann, 2013). In contrast to the limited research in autistic populations, more work has been done to examine the link between individual differences in ToM ability and conduct problems in nonautistic (or “typically developing”) populations. Early studies found that although children with clinically high levels of conduct problems, as indicated by a diagnosis of conduct disorder, may pass simple tasks, questionnaire measures suggested subtle difficulties in (particularly prosocial) ToM were present (Happé & Frith, 1996). Others report “difficult to manage” (but otherwise typically developing) 4-year-old children show impairments on ToM false belief tasks, but these differences were largely accounted for by variation in verbal IQ (Hughes, Dunn, & White, 1998). However, Hughes and colleagues have also reported associations between poorer ToM and aggressive behavior problems in typically developing 2-year-old children, even after adjusting for differences in age, sex, verbal ability, social disadvantage, and executive functioning ability (Hughes & Ensor, 2006). Studies of slightly older children (aged 3–7 years) also found those who performed worse on ToM tasks were rated as more aggressive by their teachers (Capage & Watson, 2001). More recently, an association between poorer ToM and higher levels of conduct problems has been found in youth community samples (aged 9–13 years), over and above the effects of age, sex and IQ (Sharp, 2008). Furthermore, poor ToM ability longitudinally predicts increases in teacher-rated physical and relational aggression in children aged 6–11 years, even when adjusting for differences in initial level of verbal ability (Holl, Kirsch, Rohlf, Krahé, & Elsner, 2018).

Some suggest the link between difficulties in ToM and conduct problems in nonautistic populations is due to misinterpretation of social situations leading to

interpersonal aggression (Choe, Lane, Grabell, & Olson, 2013). In support of this proposal, aggressive children more frequently attribute hostile intentions to peers in ambiguous social situations (De Castro, Merk, Koops, Veerman, & Bosch, 2005; De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). It should be noted here that children with depression are also more likely to attribute hostile intentions (Quiggle, Garber, Panak, & Dodge, 1992); however they can be differentiated by their behavioral response to the perceived negative intent. Another interpretation of the link between poor ToM and conduct problems is that the social signals that modulate ongoing interpersonal behaviors do not inhibit certain behaviors in the same manner in those with poor ToM. Others contest this link, and instead argue that *better* social cognition is associated with interpersonal aggression (e.g., the ability to bully successfully may rely on intact ToM) (Sutton, Smith, & Swettenham, 1999), although the empirical support for this stance is mixed (Monks, Smith, & Swettenham, 2005). Differences in the ages of samples, the nature of ToM test (e.g., with or without an affective component, the level of demand on other areas of cognition such as executive function/inhibitory control), and the conceptualization of conduct problems (e.g., proactive vs. reactive aggression) likely contribute to the heterogeneity in findings. This last point is important to note when considering predictors of conduct problems in autistic populations, as proactive and reactive aggression are thought to have separable cognitive correlates (Dodge & Coie, 1987; Marsee & Frick, 2007), and there is evidence to suggest autistic individuals may be more like to display reactive (as opposed to proactive) conduct problems, compared to other clinical groups (Farmer et al., 2015).

CURRENT STUDY

From literature briefly reviewed above, there is strong evidence to suggest poor ToM is associated with increased conduct problems in nonautistic (or “typically developing”) populations, however, few studies have tested whether a comparable association is found in autistic individuals, despite the high levels of conduct problems in autistic populations. A distinction can be made between explanations of increased conduct problems that posit higher rates/increased prevalence of risk factors (such as poor ToM), versus explanations that suggest the same risk factors more strongly predict conduct problems in autistic as compared to nonautistic groups. With regard to the latter hypothesis, theoretically it may be the case that some risk factors have a more negative effect in autistic individuals because of fewer buffering factors (e.g., due to a decreased social support system; Humphrey & Symes, 2010) or compensatory mechanisms

(e.g., due to a diffuse profile of cognitive impairments; Brunson et al., 2015). Alternatively, some established risk factors may not be associated with conduct problems at all in autistic individuals (e.g., as with lower IQ; Simonoff et al., 2008; Kanne & Mazurek, 2011). Although the focus of the current study is on child-intrinsic factors and specifically ToM, there are many other risk factors for mental health problems which are thought to be more prevalent in autistic populations (e.g., bullying; Schroeder, Cappadocia, Bebkco, Pepler, & Weiss, 2014, difficulties with emotion regulation; Mazefsky et al., 2013).

Thus, in the current study we test for differences in ToM ability between autistic and nonautistic individuals using a task-based assessment of ToM, and then compare the strength of associations between ToM ability and parent-reported conduct problems in both groups, to assess if ToM difficulties are associated with in a comparable way in autistic and nonautistic individuals. To the best of our knowledge, only one study has tested the association between ToM ability and conduct problems in autistic youth (Pouw et al., 2013), and no work thus far has compared the strength of association between ToM and conduct problems in autistic versus nonautistic populations. In addition, given the frequent co-occurrence of conduct problems and attention-deficit hyperactivity disorder (ADHD) (Ford, Goodman, & Meltzer, 2003) in the general population, and conduct problems and emotional symptoms in autistic populations (Simonoff et al., 2008), although our primary focus is the association between ToM ability and conduct problems, we also examine associations with these other domains of psychiatric symptoms (i.e., hyperactivity/inattention and emotional problems) known to be prevalent in autistic populations. Including these two domains as additional outcomes allows us to explore specificity of any associations to the conduct problems domain. Furthermore, given that many studies of ToM in autistic populations have used small to modest sized samples, which may contribute to issues with reproducibility in the field, we combine two well-characterized samples of autistic youth who completed the same measures of psychiatric symptoms and ToM ability to provide sufficient statistical power.

METHOD

Sample

The full sample consisted of data pooled from the two cohorts described below. More details on sample ascertainment are given in the Supporting information.

QUEST (target cohort born September 2001–September 2004, South East London area of the UK). A total of 277 children were originally assessed as part of the QUEST study (see Salazar et al., 2015 for details), a

longitudinal community autistic sample recruited at age 4–8 years, part of the wider IAM Health project (<http://iamhealthkcl.net/>). Two hundred seventy-seven children were recruited into the study and split into an “intensively studied” (hereafter intensive; $n = 101$) and “extensively studied” group (hereafter extensive; $n = 176$). This sampling structure was maintained throughout all waves of data collection. All participating girls were invited into the intensive subsample in order to make sex comparisons possible. Although all participants had a clinical diagnosis of ASD upon entry to the study, the intensive group had their diagnosis confirmed at age 11–15 years (Wave 2 of data collection) with the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012), and a subset (66/83 cases) also completed the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003). All participants were above threshold on one or both instruments. All participating families gave written informed consent and the study was approved by Camden and King’s Cross Ethics Subcommittee (14/LO/2098). The present study uses data from intensive participants at Wave 2 who had an $IQ \geq 50$ (to ensure a similar IQ distribution to the SNAP cohort) ($n = 44$, mean age = 13.49 years, 57% male).

SNAP (target cohort born July 1990–December 1991, South Thames area of the UK).

A total of 100 autistic adolescents and 57 nonautistic adolescents who had an $IQ \geq 50$, were assessed as part of the Special Needs and Autism Project (SNAP) cohort (see Baird et al., 2006 for details). Participants were recruited at 12 years and followed up in adolescence and young adulthood (Simonoff et al., 2020). Here we report upon data from adolescence (mean age 16 years) (see Charman et al., 2011 for details). Upon entry to the study all individuals in the autistic group received a consensus clinical ASD diagnosis, made using the ADI-R and ADOS-Generic (ADOS-G; Lord et al., 2000). The participants who made up the nonautistic group consisted of two groups. First, SNAP participants who did not have ASD but had a range of other primary ICD-10 diagnoses ($n = 27$; 15 mild intellectual disability; 4 moderate intellectual disability; 3 specific reading/spelling disorder; 2 ADHD; 1 expressive/receptive language disorder; 2 no diagnosis). To these were added nonautistic participants ($n = 30$) recruited from local mainstream schools. Parent and teacher report confirmed that all children in this subset of the nonautistic group were typically developing; none had a psychiatric diagnosis, a statement of special educational needs nor were receiving medication. The Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) was collected from 27 of the 30 adolescents; no individual was above the cut-off for ASD (≥ 15). Written informed consent was obtained from all parents and, where appropriate, by the participants themselves, if their level of understanding was sufficient. The study was approved by

the South East Multi-Centre Research Ethics Committee (05/MRE01/67). The present study uses data from participants at age 16 years ($n = 84$ in the autistic group, mean age = 15.45 years, 90% male, $n = 50$ in the nonautistic group, mean age = 15.45 years, 98% male).

Measures

Psychiatric symptoms. The Strengths and Difficulties Questionnaire (SDQ; Goodman, Ford, Simmons, Gatward, & Meltzer, 2000) is a 25-item questionnaire measuring psychiatric symptoms over the last 6 months, comprising of subscales of conduct problems, hyperactivity/inattention (ADHD symptoms), emotional problems, peer-relationship problems, and prosocial behavior. Analyses focused upon the parent-rated subscales of conduct problems, hyperactivity/inattention, and emotional problems, each composed of five items that are scored as not true, somewhat true or certainly true. Acceptable internal consistency is found for all three subscales ($\alpha > 0.70$) in typically developing (Goodman, 2001) and autistic samples (Findon et al., 2016). In the current sample of children who completed the ToM task, internal consistency was acceptable for the hyperactivity/inattention and emotional problems subscales (both $\alpha = 0.72$), but only moderate for the conduct problems subscale ($\alpha = 0.53$).

Autistic symptoms. The ADOS (Lord et al., 2000; Lord et al., 2012) is a semi-structured assessment that is considered a gold-standard instrument for assessing current autistic symptoms. A calibrated severity score can be calculated, scored 0–10, which takes into account age and language level (Shumway et al., 2012). Although the ADOS-2 was used in QUEST and the ADOS-G was used in SNAP, ADOS-G raw data used to generate CSS scores is equivalent to that used to create CSS scores from the ADOS-2, meaning that CSS scores from the ADOS-G and ADOS-2 are comparable. The ADOS was not completed on individuals in the nonautistic sample.

Verbal ability. Full scale and verbal IQ were predominantly estimated using the Wechsler Abbreviated Scale of Intelligence (WASI-I in SNAP, WASI-II in QUEST; Wechsler, 1999; Wechsler, 2012) (the Verbal Comprehension Index was used to calculate verbal IQ). The Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV; Wechsler, 2012) was also used with one participant in the QUEST cohort. As the WPPSI was used out of age range, an age-equivalent was calculated and a ratio full scale IQ/verbal IQ derived [ratio IQ = (age-equivalent/chronological age) \times 100] (Terman & Maude, 1960).

ToM ability. Participants completed the Frith–Happé Triangles animation task (Abell, Happé, & Frith, 2000; Castelli, Frith, Happé, & Frith, 2002), which involved viewing six short animations of two cartoon triangles interacting, and

then describing what the two triangles were doing. All responses were recorded for later transcription and scoring. The task consisted of four ToM animations depicting complex social interactions (coaxing, mocking, seducing, surprising), and two goal-directed animations, in which the actions of one object show a simple dependency on those of the other (fighting, chasing). Accurate interpretation of the ToM animations requires understanding of complex mental states to accurately interpret, whereas accurate interpretation of the goal-directed animations requires inference of dependencies but not complex mental states. Two average scores were calculated: intentionality (degree of mental state attribution: scored from 0 to 5) and appropriateness (degree to which participant correctly identified the intended content: scored from 0 to 2). In both cohorts stimuli were presented on a laptop in the presence of a trained experimenter, the same task administration and coding rules were followed, and the responses were double coded by two independent raters (all scripts in QUEST, 56% of scripts in SNAP). Reliability was high in both cohorts (range of intraclass correlations in QUEST = 0.87–0.98, SNAP = 0.82–0.98; Jones et al., 2011).

Statistical analysis

All analyses were completed in Stata 14 (StataCorp, 2015). Variables were assessed for normality; the goal-directed appropriateness variable was highly skewed due to ceiling effects, therefore scores of 0–1.5 were collapsed into a score of 0 and scores of 2 were recoded to 1 to give a binary variable. As there was some evidence of heteroscedasticity of residuals for the conduct problems subscale, all analyses were undertaken using robust standard errors (Hayes & Cai, 2007).

Combining participants from QUEST and SNAP cohorts into one group. The main aim of the study was to test the association between ToM ability and conduct problems in an appropriately powered sample. Hence, before beginning the main analyses we confirmed the validity of combining the QUEST and SNAP autistic participants into one group (and thereby increasing statistical power). We were aware that they would differ in basic demographic characteristics of sex and age due to differences in how the samples were ascertained (i.e., QUEST deliberated over-samples females, QUEST participants were assessed at 10–14 years whereas SNAP were assessed at age 16 years). For the current analyses what was key was whether the SNAP and QUEST autistic participants differed in overall ToM task performance, parent-rated psychiatric symptoms or verbal ability. We also report on cohort differences in parental education and employment to give fuller characterization of potential cohort differences. Both were assessed as binary variables (parental employment scored as 0 = neither parent in employment, 1 = one or more parent in employment; parental education scored as 0 = no qualifications up to GCSEs,

1 = A-Levels or higher). These were both compared using the χ^2 statistic. Cohort differences in all other variables were adjusted for age and sex using a series of one-way ANOVAs. Logistic regression was used for the binary variable of goal-directed appropriateness. We included age and sex as covariates here as we knew they differed between cohorts, and these were planned covariates in the main analysis; sex due to well-recognized sex differences in the prevalence of both ASD and conduct problems (Baird et al., 2006; Collishaw, Maughan, Goodman, & Pickles, 2004), age due to potential effects on task performance. The majority of comparisons between SNAP and QUEST were nonsignificant (see Table S1 for details; apart from differences in hyperactivity/inattention (SNAP had higher scores; $p = 0.03$) and goal-directed appropriateness (QUEST scoring lower; $p < 0.01$). In terms of unadjusted demographic characteristics, SNAP autistic participants also had a significantly higher rate of parental employment (99% vs. 66% in QUEST; $p < 0.01$), but no differences were found in parental education. However, as neither the goal-directed conditions nor hyperactivity/inattention were our primary predictor or outcome of interest, and given the number of comparisons made between the two cohorts (thereby inflating the likelihood of a chance finding), we did not see these two effects as sufficient evidence to suggest combining the samples was inappropriate.

Testing associations between ToM ability and psychiatric symptoms. Before running inferential statistics, we compared demographic factors and SDQ scores between the two groups, using a series of one-way ANOVAs. We also confirmed that the autistic group scored lower on the ToM task using the same approach, aside from for the binary variable of goal-directed appropriateness, where the χ^2 statistic was used. Next, we compared the strength of associations between ToM performance and psychiatric symptoms in the autistic versus non-autistic group. The critical comparison of the associations between the two groups was the significance of the diagnosis-by-task performance interaction term. A significant term indicates that the association between ToM ability and psychiatric symptoms is different in the autistic as compared to nonautistic group. As stated in the aims, our primary outcome of interest was conduct problems, however associations with hyperactivity/inattention and emotional problems were also examined to assess whether there was any specificity in the type of psychiatric symptoms associated with ToM difficulties. We used a seemingly unrelated regression approach as it allowed us to test the association between multiple predictors and multiple outcomes using standard linear regression, while also allowing for correlations between the outcomes (here the SDQ subscales) (Zellner, 1962) (see Figure 1 for a schematic of the analytic model). Each metric of task performance

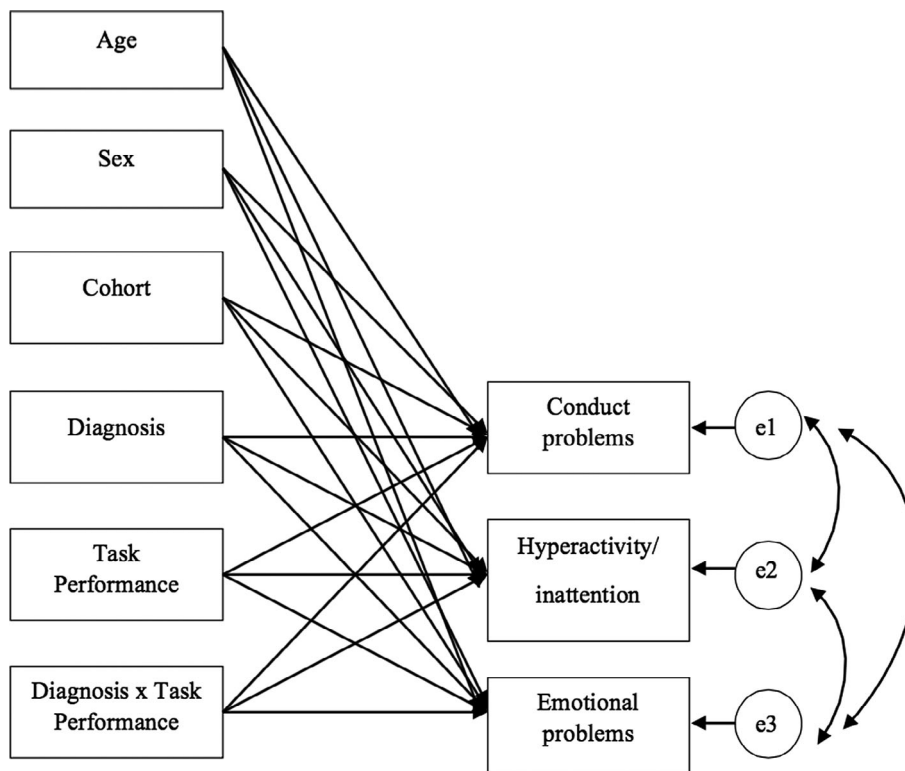


FIGURE 1. Model testing associations between task performance and psychiatric symptoms

TABLE 1. Demographic information, psychiatric symptoms and task performance in autistic and nonautistic groups

Mean (SD) (range)	Autistic (N = 128)	Nonautistic (N = 50)	Group differences
Age	14.78 (1.20) (11.4–16.75)	15.45 (0.50) (14.2–16.9)	$p < 0.01$
% male	79%	98%	$p < 0.01$
Full IQ	86.74 (17.96) (50–129)	88.36 (22.06) (54–133)	$p = 0.61$
Verbal IQ	85.41 (17.30) (54–123)	86.48 (20.39) (55–140)	$p = 0.73$
SDQ conduct problems ^a	1.86 (1.68) (0–8)	1.76 (1.90) (0–10)	$p = 0.74$
SDQ hyperactivity/inattention	5.55 (2.58) (0–10)	3.62 (2.75) (0–10)	$p < 0.01$
SDQ emotional problems	3.98 (2.46) (0–10)	1.84 (2.11) (0–10)	$p < 0.01$
ToM intentionality	2.84 (0.99) (0–4.75)	3.35 (0.85) (1.5–4.75)	$p < 0.01$
ToM appropriateness	0.65 (0.50) (0–2)	0.87 (0.49) (0–2)	$p = 0.01$
Goal-directed intentionality	2.42 (0.71) (0–4.5)	2.73 (0.66) (0.5–4.5)	$p < 0.01$
Goal-directed appropriateness (% scoring 2)	48%	64%	$p = 0.05$

Abbreviations: SDQ, Strengths and Difficulties Questionnaire; ToM, theory of mind.

^aMissing from one participant in the autistic group.

was tested independently. Metrics of task performance were centered to aid interpretation of effects. Our primary hypotheses concerned performance in the ToM condition; the goal-directed metrics were included to provide a check that findings were specific to understanding of mental state terms, rather than inference in general. Results of these analyses are reported in Table S2; in summary, no significant associations were found between goal-directed abilities and psychiatric symptoms. Sex, age and cohort (SNAP = 0, QUEST = 1) were entered as covariates; the latter to account for any additional unmeasured ascertainment differences. Finally, where the diagnosis-by-task performance interaction term was nonsignificant, the model was rerun excluding the interaction term to estimate the pooled association between task performance and psychiatric symptoms across the two groups. Standardized estimates (β) were calculated for the pooled associations to aid

interpretation (where 0.1 = small effect, 0.3 = moderate effect, and 0.5 = large effect; Cohen, 1988).

As the task required participants to verbally state their response, associations from the pooled sample were consequently adjusted for verbal IQ. We also report results of sensitivity analyses excluding participants with verbal IQ < 80 (Table S3). For completeness, bivariate correlations were calculated between all variables (Table S4).

As a final check, we tested for differences in ToM-psychiatric symptoms associations between SNAP versus QUEST in the autistic group only (see Table S5 for full results). To give an estimate of the power to detect associations between ToM performance and conduct problems we conducted post hoc power calculations in the full sample on the two primary paths of interest (ToM intentionality/appropriateness – conduct problems).

TABLE 2. Associations between task performance and psychiatric symptoms

		Unadjusted model			Interaction term removed			Adjust for verbal IQ		
		CPs	HA/IA	EPs	CPs	HA/IA	EPs	CPs	HA/IA	EPs
ToM Intentionality	Task performance	-0.42	-0.81*	0.04	-0.29*	-0.48**	-0.17	-0.12	-0.13	-0.07
	ASD diagnosis	-0.21	1.83**	1.65**	-0.17	1.94**	1.58**	-0.14	2.00**	1.59**
	Diagnosis × task performance	0.17	0.44	-0.27	—	—	—	—	—	—
	Sex	-0.01	-.38	.45	-.01	-.40	.46	-.01	-.40	.46
	Age	0.06	-0.60*	-0.11	0.05	-0.63**	-0.09	0.08	-0.56*	-0.07
	Cohort	0.47	-1.74*	0.85	0.45	-1.79*	0.89	0.63	-1.41	0.10
ToM Appropriateness	Task performance	-0.89*	-1.92**	-0.18	-0.72**	-1.01**	-0.67*	-0.48	-0.44	-0.59
	ASD diagnosis	-0.25	1.80**	1.53**	0.23	1.90**	1.48**	-0.21	1.95**	1.49**
	Diagnosis × task performance	0.24	1.26	-0.68	—	—	—	—	—	—
	Sex	0.06	-0.29	0.52	0.06	-0.29	0.52	0.03	-0.36	0.51
	Age	0.08	-0.60*	-0.06	0.08	-0.59*	-0.07	0.10	-0.55*	-0.06
	Cohort	0.60	-1.63*	1.07	0.61	-1.56*	1.03	0.71	-1.33	1.07

Note: All displayed coefficients are unstandardized.

Abbreviations: CPs, conduct problems; EPs, emotional problems; HA/IA, hyperactivity/inattention; ToM, theory of mind.

* $p < 0.05$.

** $p < 0.01$.

RESULTS

Group differences in demographic variables, psychiatric symptoms and ToM task performance

See Table 1 for group means. The nonautistic group had a higher percentage of male participants (98% vs. 79% in the autistic group) and were older than the autistic group (both $ps < 0.01$), but there were no group differences in full scale or verbal IQ. The autistic and nonautistic group had comparable levels of conduct problems ($p = 0.74$), but the autistic group had higher levels of emotional problems and hyperactivity/inattention (both $ps < 0.01$). Confirming the first

part of our hypothesis, the autistic group scored lower on the ToM task (all $ps < 0.05$), apart from in the goal-directed appropriateness metric, where group differences were at a trend level of significance ($p = 0.05$).

Associations between task performance and psychiatric symptoms

Certain associations remained constant across all analyses, therefore are summarized to avoid repetition (see Table 2 for details). Sex was not a significant predictor of any outcomes (all $ps > 0.10$; Lines 5 and 11), and age was

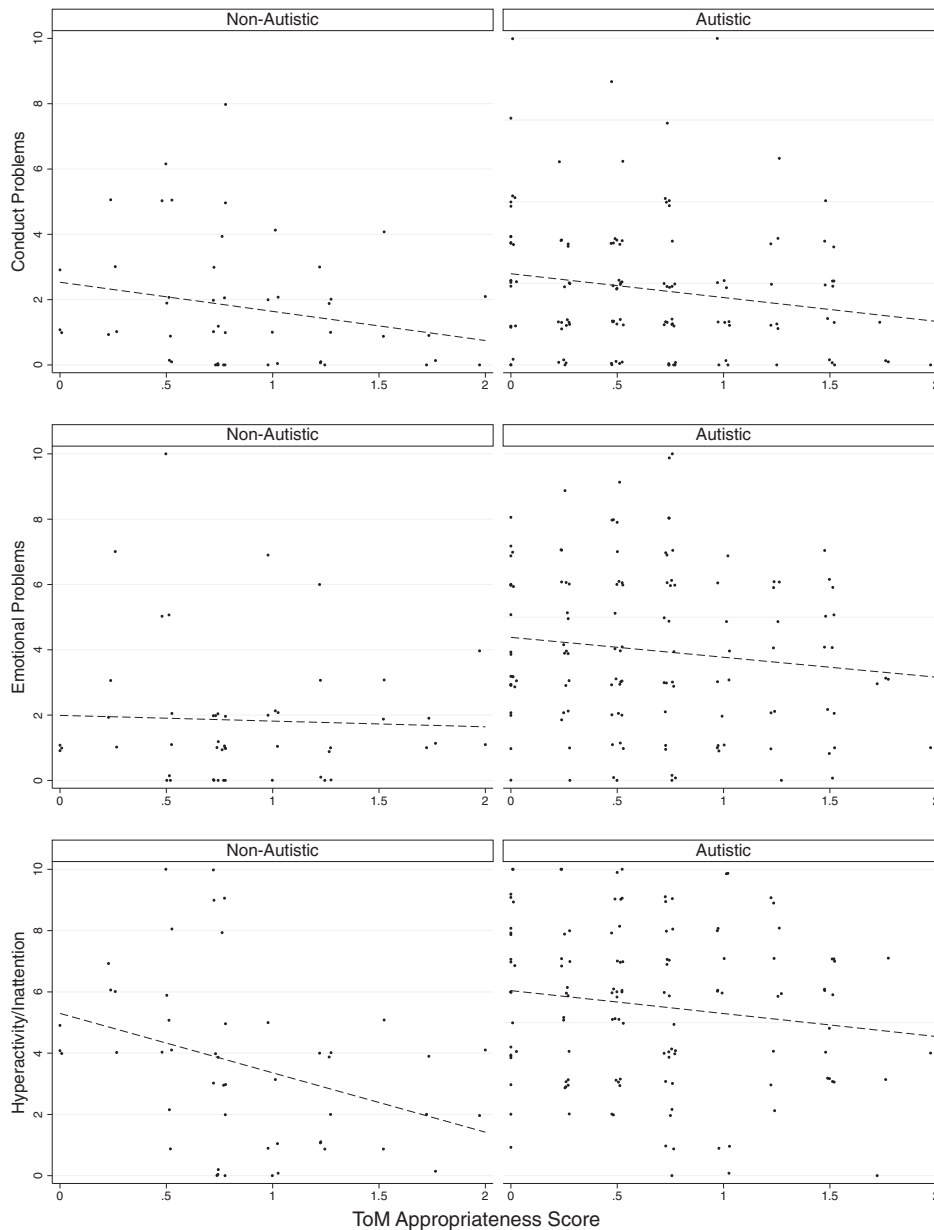


FIGURE 2. Associations between theory of mind (ToM) appropriateness and conduct problems (top), hyperactivity/inattention (middle) and emotional problems (bottom), split by autism spectrum disorder (ASD) diagnostic status. Data have been jittered for illustrative purposes

negatively associated with hyperactivity/inattention (ps ranged from 0.03 to 0.04; Table 2, Lines 6 and 12, Columns 4, 7, and 10). The presence of an ASD diagnosis was positively associated with hyperactivity/inattention (all $ps < 0.01$; Table 2, Lines 3 and 9, Columns 4, 7, and 10) and emotional problems (all $ps < 0.01$; Table 2, Columns 5, 8, and 11).

Intentionality score. The diagnosis-by-task performance interaction term was not significant in the prediction of conduct problems ($b = 0.17, p = 0.65$), hyperactivity/inattention ($b = 0.44, p = 0.31$), or emotional problems ($b = -0.27, p = 0.56$) (Table 2, Line 4, Columns 3–5). The interaction term was removed and the model was rerun in the whole sample (autistic and nonautistic combined). A negative association was found between intentionality and conduct problems ($b = -0.29, \beta = -0.16, p = 0.05$) and hyperactivity/inattention ($b = -0.48, \beta = -0.17, p < 0.01$), but no association was found with emotional problems ($b = -0.17, \beta = -0.06, p = 0.31$) (Table 2, Line 2, Columns 6–8). Both significant associations became nonsignificant with the inclusion of verbal IQ as a covariate ($b = -0.12, \beta = -0.07, p = 0.48$ for conduct problems, $b = -0.13, \beta = -0.04, p = 0.56$ for hyperactivity/inattention) (Table 2, Line 2, Columns 9–11).

Appropriateness score. The diagnosis-by-task performance interaction term was not significant in the prediction of conduct problems ($b = 0.24, p = 0.63$), hyperactivity/inattention ($b = 1.24, p = 0.06$), or emotional problems ($b = -0.68, p = 0.29$) (Table 2, Line 10, Columns 3–5). The interaction term was removed and the model was re-run in the whole sample (autistic and nonautistic combined). Negative associations were found between appropriateness and conduct problems ($b = -0.72, \beta = -0.21, p < 0.01$), hyperactivity/inattention ($b = -1.01, \beta = -0.19, p < 0.01$), and emotional problems ($b = -0.67, \beta = -0.13; p = 0.03$) (see Figure 2) (Table 2, Line 8, Columns 6–8), however all became nonsignificant adjusting for verbal IQ ($b = -0.48, \beta = -0.14, p = 0.08$ for conduct problems, $b = -0.44, \beta = -0.08, p = 0.23$ for hyperactivity/inattention; $b = -0.59, \beta = -0.12, p = 0.09$ for emotional problems) (Table 2, Line 8, Columns 9–11).

Cohort differences in task performance-psychiatric symptoms associations

We found two instances of sample-specific effects; the association between ToM appropriateness and emotional problems, and the association between goal-directed appropriateness and emotional problems, where both cohort-by-task performance interaction terms were significant (both $p < 0.05$) (see Table S5 for full output). There was a significant association between ToM appropriateness and emotional problems in QUEST ($b = -1.67, p < 0.01$) but only a marginal association in SNAP

($b = -0.63, p = 0.07$), and a significant association between goal-directed appropriateness and emotional problems in SNAP ($b = -1.17, p = 0.02$) but no association in QUEST ($b = 1.14, p = 0.11$).

Post hoc power calculations

The calculations for the two paths of primary interest (ToM intentionality/appropriateness scores – SDQ conduct problems) suggested that the power to detect effects for the ToM intentionality–conduct problems association was questionable (51%), but satisfactory for the ToM appropriateness–conduct problems association (86%) (both coefficients estimated at two-tailed 95% significance).

DISCUSSION

There is strong evidence to suggest difficulties in ToM are associated with increased conduct problems in nonautistic populations (Capage & Watson, 2001; Holl et al., 2018; Hughes & Ensor, 2006; Sharp, 2008). However, few studies have tested whether a comparable association is found in autistic individuals, despite the high levels of conduct problems in autistic populations. The current study sought to not only test the association between experimentally-assessed ToM difficulties and parent-reported conduct problems in autistic youth (thereby minimizing the impact of shared rater variance), but also to compare the strength of associations between ToM ability and conduct problems in autistic versus nonautistic youth, the latter to better understand how established risk factors function in populations of youth with and without a diagnosis of ASD. Analyses found negative associations between ToM performance and conduct problems, hyperactivity/inattention and emotional problems in the full sample, and no evidence of moderation of ToM-behavior associations by diagnosis, as evidenced by the lack of significance of all diagnosis-by-task performance interaction terms. Pooled associations became nonsignificant when verbal IQ was included as a covariate.

Notwithstanding the fact we found associations not only with conduct problems, but also with hyperactivity/inattention and emotional problems, our primary hypothesis that the association between ToM difficulties and *conduct problems* reported in nonautistic populations would also be found in autistic youth was confirmed. The lack of moderation by diagnosis and the significant unadjusted association with conduct problems suggests that in the current sample, both autistic and nonautistic youths with ToM difficulties had higher levels of conduct problems. This replicates and extends findings reported from nonautistic child and adolescent samples (Capage & Watson, 2001; Holl et al., 2018; Hughes & Ensor, 2006;

Sharp, 2008). With regard to the lack of specificity of associations with poorer ToM, we note that there is some existing evidence for links between ToM impairment and emotional symptoms in nonautistic populations (Colonnesi, Nikolić, de Vente, & Bögels, 2017; Hezel & McNally, 2014; Lee, Harkness, Sabbagh, & Jacobson, 2005). The specificity of the association with hyperactivity/inattention remains unclear; this may simply reflect difficulties in paying attention to cognitive tasks or executive functioning impairments in youth with ADHD symptoms rather than evidence for a specific difficulty in ToM (e.g., Mary et al., 2016). We highlight that statistical model we used accounted for cross-domain correlations, therefore associations with emotional problems and hyperactivity cannot simply be due to overlap with conduct problems.

Results showed significant associations with performance in the ToM, but not the goal-directed, condition, suggesting the link with psychiatric symptoms is due to difficulties in higher-level understanding of mental state terms, rather than general inference. In nonautistic populations, the link between ToM and conduct problems is proposed to be due to poorer ToM leading to misinterpretation of social cues, which in turn prompts interpersonal aggression (Choe et al., 2013; De Castro et al., 2002; De Castro et al., 2005). Our results could be seen to support this suggestion, as stronger effects were found for the appropriateness score, reflecting the accuracy of the interpretation of the animated cartoons, rather than the intentionality score, which indexes the quality of mental state terms used (regardless of their appropriateness). Tasks specifically designed to measure variation in valence of mental state attributions in ambiguous social situations (e.g., De Castro et al., 2002; De Castro et al., 2005) could test this hypothesis.

In previously work (Hughes & Ensor, 2006; Sharp, 2008), adjusting for verbal ability or IQ did not alter the significance of associations between ToM and conduct problems, whereas in our sample inclusion of verbal IQ as a covariate caused associations to fall to nonsignificant trends or become fully nonsignificant. A similar drop in significance of ToM-conduct problems association when accounting for verbal ability was also reported from one study of typically developing four-year old children (Hughes et al., 1998). Inspection of the change in standardized estimates between task performance and the three domains of psychiatric symptoms when verbal IQ was included in the model suggests around one third of the association between ToM appropriateness and conduct problems ($\beta = -0.21$ fell to $\beta = -0.14$) and around 40% of the association between ToM appropriateness and hyperactivity/inattention ($\beta = -0.19$ fell to $\beta = -0.08$) was due to variation in verbal IQ. The change in estimates for emotional problems was negligible ($\beta = -0.13$ fell to $\beta = -0.12$). It should be noted that some

have argued that controlling for verbal ability likely removes some “true” ToM effect, given how strongly inter-related ToM and language/verbal ability are in development (Happé, 2015), and that autistic groups appear to rely more strongly on verbal ability to pass ToM tests (Happé, 1995). Disagreement regarding the role of verbal ability between current results and previous literature may be in part because studies solely drawn from nonautistic samples generally have a narrower range of cognitive ability and do not include individuals with $IQ < 70$ (e.g., in the study by Sharp, 2008, the lower bound on their IQ range is similar to the mean IQ in the current study), but may also be due to differences in the ages of samples (e.g., toddlerhood vs. adolescence). Research in this area would benefit from comprehensive measurement of cognitive functioning across a range of domains combined with multivariate statistical approaches, to test which relevant factors are driving different associations, and if they are domain-specific (e.g., associations are only found with tasks tapping ToM) or general (e.g., Carter Leno et al., 2018).

The lack of moderation by diagnostic status suggests that the strength of association between ToM ability and the three domains of psychiatric symptoms included at present was comparable in autistic and nonautistic participants (although we note there will be less power in the test of the interaction term as compared to the tests of associations in the pooled sample). Results suggests that certain risk factors (which are associated with psychopathology in nonautistic populations; here ToM difficulties) may function in a similar manner in the whole population (e.g., in both autistic and nonautistic individuals), however the fact they are more common in certain subgroups (e.g., autistic individuals) could partially explain the heightened prevalence of psychiatric symptoms in that subgroup. Thus we propose that the increased rates of ToM difficulties in autistic individuals (we found overall ToM performance was poorer in the autistic group, similar to that reported elsewhere; Salter et al., 2008), can not only be conceptualized as a driver of core autism symptoms, but also a risk factor for additional psychiatric difficulties.

The current paper only tests one established predictor of conduct problems; there are many more risk factors for conduct problems that could also contribute to the increased rates of behavioral difficulties in autistic individuals (e.g., inconsistent parenting strategies; Dretzke et al., 2009; Gardner, 1989; bullying; Singham et al., 2017; Wolke, Woods, Bloomfield, & Karstadt, 2000). Additionally, it remains possible that these and other risk factors for conduct problems function differently in autistic populations, in that they may have a more negative impact or less influence (as appears to be the case for low IQ; Kanne & Mazurek, 2011; Simonoff et al., 2008) compared with other groups. Understanding how established risk factors

function in autistic as compared to nonautistic populations is key to understanding whether existing interventions developed for nonautistic individuals should also be offered to autistic people.

The current study has a number of strengths: first, the use of objective metrics specifically designed to measure ToM, which were not confounded by other cognitive domains (e.g., face processing). Second, both cohorts included well-characterized autistic individuals with a wide range of verbal IQ, who had their ASD diagnoses confirmed with “gold-standard” instruments. Most studies of cognitive functioning (in both autistic and nonautistic populations) are with samples with a narrow range of verbal/full scale IQ, which is not representative of autistic populations as a whole. Third, utilizing a well-powered sample allowed us to test for a moderating effect of diagnosis upon the association between ToM and psychiatric symptoms, although it should be noted that the nonautistic group was considerably smaller in size and a nearly half had intellectual disability. However, as our hypothesis under question was whether autism moderated the association between ToM and conduct problems, what was crucial was that the nonautistic group did not meet diagnostic criteria for ASD. Furthermore, given that autistic individuals are substantially more likely to have intellectual or learning disabilities (Charman, Jones, et al., 2011; Charman, Pickles, et al., 2011), there is still debate over the validity of using a control group with a higher IQ than the autistic group when asking questions about cognitive functioning. We also note that full scale IQ and verbal ability were not significantly different between the autistic and nonautistic participants. Finally, we note that although we had a relatively modest sample size in the nonautistic group, links between poorer ToM and higher levels of conduct problems in nonautistic individuals are well-established by previous literature (e.g., Capage & Watson, 2001; Holl et al., 2018; Hughes & Ensor, 2006), including in studies that have used the same ToM task as used currently (Sharp, 2008).

In terms of limitations, the ToM task employed in the current study, although widely used, requires participants to verbally communicate their responses, so may have underestimated ToM abilities. Although not they key task metric of ToM ability (but rather a “control” condition to check findings were specific to mental state understanding), ceiling effects in the goal-directed appropriateness condition meant a binary variable was used in these analyses. This may have limited our power to detect effects, and thus results here should be interpreted with caution. Additionally, more controlled tasks such as the one included currently could be argued to be lower in ecological validity, as in real life contexts decoding social situations often involves an affective component and multiple competing inputs. Future studies should employ a battery of ToM measures, including those that do not rely on

spoken communication, such as eye-tracking paradigms (Senju, Southgate, White, & Frith, 2009), and those with an affective component (Sebastian et al., 2012; Shamay-Tsoory, Tomer, Berger, Goldsher, & Aharon-Peretz, 2005) to test the impact of task content and presentation. Finally, the current analyses present cross-sectional association in a sample of adolescents; whether similar associations are found at different developmental stages, and the directionality of associations between cognition and behavior cannot be inferred from the current results.

Theoretically, as none of the diagnosis-by-task performance interaction terms were significant, current results suggest that ToM difficulties are associated with conduct problems in a similar manner in autistic and nonautistic populations, and therefore interventions focused on improving the accuracy of mental state attribution may be beneficial in improving conduct problems in general. Furthermore, as similar associations were found with hyperactivity/inattention and emotional problems, interventions may be beneficial for symptoms beyond those in the domain of conduct/oppositional defiant disorder. However, although meta-analyses report an overall improvement in ToM following targeted intervention in autistic and nonautistic samples (Hofmann et al., 2016), others contest that this is based on studies of low quality, and suggest there is little robust evidence to suggest that improvement of the knowledge acquired through training translates to changes in everyday functioning (Fletcher-Watson, McConnell, Manola, & McConachie, 2014). Instead of a sole focus on ToM, appreciation of the heterogeneity of cognitive profiles in autistic individuals, and consideration of the impact of particular individual characteristics in the context of evidence-based interventions may be of merit, although the added impact of adding a ToM component to existing interventions should be formally tested. Furthermore, the finding that verbal ability may account for a substantial proportion (but not all) of the association between ToM and psychiatric symptoms suggests that interventions should also consider language skills as a target of interest (in addition to specific domains of cognitive functioning). This is in line with previous work highlighting the role of poor communication in conduct problems in both autistic and nonautistic populations (McClintock, Hall, & Oliver, 2003; Moffitt & Silva, 1988). Conducting similar analyses in longitudinal cohorts, followed from early in childhood, along with high-quality randomized control trials, will be key to testing developmental hypotheses regarding the impact of impaired or delayed ToM and verbal abilities. Furthermore, as noted in the Introduction, conduct-disordered behavior such as aggression can be partitioned into both reactive and proactive behaviors, which are thought to have independent cognitive correlates (Dodge & Coie, 1987; Marsee & Frick, 2007). Whether a similar differentiation is found in autistic samples has not yet been well explored

(although see Pouw et al., 2013). Finally, a comprehensive etiological model of conduct problems in autistic populations should not only consider child-level characteristics such as ToM abilities, but also how these may interact with other established environment-level risk factors (e.g., inconsistent parenting strategies, bullying).

CONCLUSION

The current study found that lower ToM ability was associated with higher levels of conduct problems, but also hyperactivity/inattention and emotional problems, in autistic and nonautistic youth. However, results also suggest that a substantial part of the pooled association between ToM and psychiatric symptoms was driven by individual differences in verbal IQ ability. Understanding the drivers of conduct problems in autistic individuals, and whether they are comparable to those found in nonautistic individuals, is important for accurate etiological models and will inform selection of appropriate targets for intervention.

ACKNOWLEDGMENTS

We would like to thank all the participants and their caregivers involved in this research. We also thank Drs Jacqueline Bold and Mark O'Leary for assistance with the QUEST study. The original QUEST sample was funded by Clothworkers' Foundation, brokered by Research Autism (R011217 Autism M10 2011/12). The IAM Health research programme was funded by the National Institute for Health Research (NIHR) (RP-PG-1211-20016). The SNAP study was supported by the Medical Research Council (G0400065). ES and AP receive NIHR Senior Investigator Awards (NF-SI-0514-10073, NF-SI-0617-10120). The study was partially supported through the NIHR Maudsley Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust in partnership with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the MRC, the NIHR or the Department of Health.

CONFLICT OF INTEREST

AP receives royalties from the Social Communication Questionnaire. There are no other conflicts of interest to report.

REFERENCES

- Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15* (1), 1–16.
- Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *The Lancet, 368*(9531), 210–215.
- Brunsdon, V. E. A., Colvert, E., Ames, C., Garnett, T., Gillan, N., Hallett, V., ... Happé, F. (2015). Exploring the cognitive features in children with autism spectrum disorder, their co-twins, and typically developing children within a population-based sample. *Journal of Child Psychology and Psychiatry, 56* (8), 893–902.
- Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behaviour, and social skills in young children. *Early Education and Development, 12*(4), 613–628.
- Carter Leno, V., Vitoratou, S., Kent, R., Charman, T., Chandler, S., Jones, C., ... Simonoff, E. (2018). Exploring the neurocognitive correlates of challenging behaviours in young people with autism spectrum disorder. *Autism, 23*(5), 1152–1164.
- Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839–1849.
- Charman, T., Jones, C. R. G., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011). Defining the cognitive phenotype of autism. *Brain Research, 1380*, 10–21. <http://dx.doi.org/10.1016/j.brainres.2010.10.075>.
- Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). *Psychological Medicine, 41*(3), 619–627.
- Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology, 49*(12), 2245.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.), Hillsdale, NJ: Erlbaum.
- Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004). Time trends in adolescent mental health. *Journal of Child Psychology and Psychiatry, 45*(8), 1350–1362.
- Colonnesi, C., Nikolić, M., de Vente, W., & Bögels, S. M. (2017). Social anxiety symptoms in young children: Investigating the interplay of theory of mind and expressions of shyness. *Journal of Abnormal Child Psychology, 45*(5), 997–1011.
- De Castro, B. O., Merk, W., Koops, W., Veerman, J. W., & Bosch, J. D. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 105–116.
- De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behaviour: A meta-analysis. *Child Development, 73*(3), 916–934.
- Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*(6), 1146.
- Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., ... Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct

- problems: A systematic review of randomised controlled trials. *Child and Adolescent Psychiatry and Mental Health*, 3(1), 7.
- Farmer, C., Butter, E., Mazurek, M. O., Cowan, C., Lainhart, J., Cook, E. H., ... Aman, M. (2015). Aggression in children with autism spectrum disorders and a clinic-referred comparison group. *Autism*, 19(3), 281–291.
- Findon, J., Cadman, T., Stewart, C. S., Woodhouse, E., Eklund, H., Hayward, H., ... McEwen, F. S. (2016). Screening for co-occurring conditions in adults with autism spectrum disorder using the strengths and difficulties questionnaire: A pilot study. *Autism Research*, 9(12), 1353–1363.
- Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). Interventions based on the theory of mind cognitive model for autism spectrum disorder (ASD). *Cochrane Database of Systematic Reviews*, (3). <http://dx.doi.org/10.1002/14651858.cd008785.pub2>.
- Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(10), 1203–1211.
- Gardner, F. E. M. (1989). Inconsistent parenting: Is there evidence for a link with children's conduct problems? *Journal of Abnormal Child Psychology*, 17(2), 223–233.
- Gjevik, E., Eldevik, S., Fjæran-Granum, T., & Sponheim, E. (2011). Kiddie-SADS reveals high rates of DSM-IV disorders in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 41(6), 761–769.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177(6), 534–539.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–855.
- Happé, F. (2015). Autism as a neurodevelopmental disorder of mind-reading. *Journal of the British Academy*, 3, 197–209.
- Happé, F., & Conway, J. R. (2016). Recent progress in understanding skills and impairments in social cognition. *Current Opinion in Pediatrics*, 28(6), 736–742.
- Happé, F., & Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *British Journal of Developmental Psychology*, 14(4), 385–398.
- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39, 709–722.
- Hezel, D. M., & McNally, R. J. (2014). Theory of mind impairments in social anxiety disorder. *Behavior Therapy*, 45(4), 530–540.
- Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., ... Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, 150, 200–212.
- Holl, A. K., Kirsch, F., Rohlf, H., Krahé, B., & Elsner, B. (2018). Longitudinal reciprocity between theory of mind and aggression in middle childhood. *International Journal of Behavioural Development*, 42(2), 257–266.
- Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(7), 981–994.
- Hughes, C., & Ensor, R. (2006). Behavioural problems in 2-year-olds: Links with individual differences in theory of mind, executive function and harsh parenting. *Journal of Child Psychology and Psychiatry*, 47(5), 488–497.
- Humphrey, N., & Symes, W. (2010). Perceptions of social support and experience of bullying among pupils with autistic spectrum disorders in mainstream secondary schools. *European Journal of Special Needs Education*, 25(1), 77–91.
- Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., ... Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, 11(1), 95–109.
- Jones, C. R. G., Swettenham, J., Charman, T., Marsden, A. J., Tregay, J., Baird, G., ... Happe, F. (2011). No evidence for a fundamental visual motion processing deficit in adolescents with autism spectrum disorders. *Autism Research*, 4(5), 347–357.
- Kanne, S. M., & Mazurek, M. O. (2011). Aggression in children and adolescents with ASD: Prevalence and risk factors. *Journal of Autism and Developmental Disorders*, 41(7), 926–937.
- Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders*, 86(2), 247–258.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The autism diagnostic observation schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S. L., ... Guthrie, W. (2012). *The Autism Diagnostic Observation Schedule, Second Edition (ADOS-2)*. San Antonio, TX: Pearson Assessments.
- Marsee, M. A., & Frick, P. J. (2007). Exploring the cognitive and emotional correlates to proactive and reactive aggression in a sample of detained girls. *Journal of Abnormal Child Psychology*, 35(6), 969–981.
- Mary, A., Slama, H., Mousty, P., Massat, I., Capiou, T., Drabs, V., & Peigneux, P. (2016). Executive and attentional contributions to theory of mind deficit in attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology*, 22(3), 345–365.
- Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L., White, S. W. (2013). The Role of Emotion Regulation in Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(7), 679–688. <http://dx.doi.org/10.1016/j.jaac.2013.05.006>.
- McClintock, K., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: A meta-analytic study. *Journal of Intellectual Disability Research*, 47(6), 405–416.

- Moffitt, T. E., & Silva, P. A. (1988). IQ and delinquency: A direct test of the differential detection hypothesis. *Journal of Abnormal Child Psychology*, 17, 330–333.
- Monks, C. P., Smith, P. K., & Swettenham, J. (2005). Psychological correlates of peer victimisation in preschool: Social cognitive skills, executive function and attachment profiles. *Aggressive Behaviour*, 31(6), 571–588.
- Pouw, L. B. C., Rieffe, C., Oosterveld, P., Huskens, B., & Stockmann, L. (2013). Reactive/proactive aggression and affective/cognitive empathy in children with ASD. *Research in Developmental Disabilities*, 34(4), 1256–1266.
- Quiggle, N. L., Garber, J., Panak, W. F., & Dodge, K. A. (1992). Social information processing in aggressive and depressed children. *Child Development*, 63(6), 1305–1320.
- Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.
- Salazar, F., Baird, G., Chandler, S., Tseng, E., O’Sullivan, T., Howlin, P., ... Simonoff, E. (2015). Co-occurring emotional and behavioural disorders in preschool and elementary school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(8), 2283–2294.
- Salter, G., Seigal, A., Claxton, M., Lawrence, K., & Skuse, D. (2008). Can autistic children read the mind of an animated triangle? *Autism*, 12(4), 349–371.
- Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 54, 628–635.
- Schroeder, J. H., Cappadocia, M. C., Bebko, J. M., Pepler, D. J., & Weiss, J. A. (2014). Shedding light on a pervasive problem: A review of research on bullying experiences among children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44(7), 1520–1534.
- Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S.-J., De Brito, S., McCrory, E., & Viding, E. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, 7(1), 53–63.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, 325(5942), 883–885.
- Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired ‘affective theory of mind’ is associated with right ventromedial prefrontal damage. *Cognition Behaviour and Neurology*, 18(1), 55–67.
- Sharp, C. (2008). Theory of mind and conduct problems in children: Deficits in reading the “emotions of the eyes”. *Cognition and Emotion*, 22(6), 1149–1158.
- Shumway, S., Farmer, C., Thurm, A., Joseph, L., Black, D., & Golden, C. (2012). The ADOS Calibrated Severity Score: Relationship to phenotypic variables and stability over time. *Autism Research*, 5(4), 267–276.
- Simonoff, E., Kent, R., Stringer, D., Lord, C., Briskman, J., Lukito, S., ... Baird, G. (2020). Trajectories in symptoms of autism and cognitive ability in autism from childhood to adult life: Findings from a longitudinal epidemiological cohort. *Journal of the American Academy of Child & Adolescent Psychiatry*. <https://doi.org/10.1016/j.jaac.2019.11.020>.
- Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(8), 921–929.
- Singham, T., Viding, E., Schoeler, T., Arseneault, L., Ronald, A., Cecil, C. M., ... Pingault, J.-B. (2017). Concurrent and longitudinal contribution of exposure to bullying in childhood to mental health: The role of vulnerability and resilience. *JAMA Psychiatry*, 74(11), 1112–1119.
- StataCorp. (2015). *Stata statistical software: Release 14*.
- Sutton, J., Smith, P. K., & Swettenham, J. (1999). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology*, 17(3), 435–450.
- Terman, L., & Maude, M. (1960). *Stanford–Binet Intelligence Scale: Manual for the Third Revision Form*. Boston, MA: Houghton Mifflin.
- Wechsler, D. (1999). *The Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Wolke, D., Woods, S., Bloomfield, L., & Karstadt, L. (2000). The association between direct and relational bullying and behaviour problems among primary school children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(8), 989–1002.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298), 348–368.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Supporting information