Do cognitive deficits persist into adolescence in autism?

Cantio, Cathriona; White, Sarah; Madsen, Gitte Falcher; Bilenberg, Niels; Jepsen, Jens Richardt Møllegaard

Published in:
Autism Research

DOI:
10.1002/aur.1976

Publication date:
2018

Document version
Accepted manuscript

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
Do Cognitive Deficits Persist Into Adolescence In Autism?

Cathriona Cantio, MSc., PhD (1, 2), Dr. Sarah White (4), Gitte Falcher Madsen, MD, PhD (1), Niels Bilenberg, Professor, MD, PhD (1, 3), and Jens Richardt Møllegaard Jepsen, MSc., PhD (5, 6)

(1) Child and Adolescent Psychiatric Department, Odense, Mental Health Services, Southern Denmark

(2) Department of Psychology, University of Southern Denmark

(3) Institute of Clinical Research, University of Southern Denmark

(4) Institute of Cognitive Neuroscience, University College London

(5) Centre for Neuropsychiatric Schizophrenia Research & Centre for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Copenhagen University Hospital, Psychiatric Hospital Centre Glostrup, Denmark

(6) Child and Adolescent Mental Health Centre, Mental Health Services Capital Region of Denmark, University of Copenhagen, Copenhagen, Denmark

Number of text pages: 15

Number of Tables: 2

Number of Figures: 4

Corresponding authors contact information:

Cathriona Cantio, Child and Adolescent Psychiatric Department, Odense, J.B. Winsløws Vej 16, 5000 Odense C, Denmark, Tel.: +45 24658731, Email: cathriona.cantio@rsyd.dk
Grant information

This study was supported by Grant sponsor: The Psychiatric Research Foundation in the Region of Southern Denmark; Grant sponsor: The PhD grant in Region of Southern Denmark; Grant sponsor: Sofiefonden; Grant sponsor: Royal Society (Dorothy Hodgkin Research Fellowship DH150167), and; Grant sponsor: Tryg foundation, Denmark

Acknowledgements

The authors wish to thank the participating children and families.

We also wish to thank Uta Frith for her insight and helpful feedback.

We are grateful to the foundations that have supported this project: The Psychiatric Research Foundation in the Region of Southern Denmark; The PhD grant in Region of Southern Denmark; Sofiefonden, and; Tryg foundation.
Abstract

Several theories have attempted to characterise autism spectrum disorders (ASD) at the cognitive level, most notably: Theory of Mind (ToM), executive function (EF) and a local processing bias (LB). The aim of this study was to investigate how these cognitive functions develop over time.

The three cognitive domains (ToM, EF and LB) were examined in a group of high-functioning children (age: 8-12, mean 10.85; IQ: 78-139, mean 105.48) with ASD and a matched group of children with neurotypical development (IQ: 75-145, mean: 109.47), and several tasks were used within each domain to ensure the validity of the cognitive measures. Approximately three years later (mean age: 14.34), all children and their families were invited to participate in the follow-up (ASD, N=21; NTD, N=30).

While the understanding of other’s minds does improve from childhood to adolescence, Theory of Mind (ToM) impairment persists in adolescents with ASD relative to their peers. Likewise, a development in Executive Function (EF) were observed in the ASD group, while no significant improvement was seen in the NTD group, leading the ASD group to catch up in this domain. We did not detect any group differences at any time point regarding Local Bias processing (LB).

Individual patterns of development were seen, but remarkably, ToM deficits were present in every child with ASD in whom we could detect any cognitive impairment at baseline, and a similar pattern was found at follow-up.

These findings indicate that ToM is a persistent cognitive deficit in ASD.

Lay summary: This was the first study to investigate the development of three well-known cognitive functions into adolescence: While the understanding of other’s minds improves from childhood to adolescence, adolescents with ASD are still impaired relative to their peers. The executive functions, however, seem to improve to a neurotypical level in ASD as children enter adolescence, while Local Processing bias seems to differentiate the groups only in early childhood.
Do Cognitive Deficits Persist Into Adolescence in Autism?

An ASD-diagnosis is defined as a set of behavioural criteria (social communication symptoms and repetitive behaviours and interests; DSM-5, American Psychiatric Association, APA, 2013). However, it is acknowledged both clinically and scientifically that different individuals vary in their personal presentation of this shared diagnostic label (Geschwind, 2009; Munson, Faja, Meltzoff, Abbott, & Dawson, 2008; Ronald, Happé, Price, Baron-Cohen, & Plomin, 2006), and likewise the same heterogeneity is clearly evident at the genetic level (Abrahams & Geschwind, 2008; O'Roak et al., 2012).

This has led the field to pursue endophenotypic markers that can homogeneously draw subgroups of individuals together (Charman et al., 2007), and a number of cognitive theories have been proposed in this matter, most notably the theories of impairments in Theory of Mind (ToM, Frith, Morton, & Leslie, 1991), impairments in executive functioning (EF, Hill, 2004), and a local processing style (LB, also called weak central coherence, Happé & Frith, 2006). The ToM account of ASD, a difficulty representing the mental states of others, was proposed by Baron-Cohen, Leslie, and Frith (1985) as a cognitive explanation for the socio-communicative impairment. The EF account on the other hand originally aimed to explain the repetitive and stereotyped behaviours through a lack of higher order control processes such as planning, flexibility and inhibition (Ozonoff, Pennington, & Rogers, 1991), and the third account, LB, aims to explain why individuals with ASD often have trouble integrating information but also show “islets of ability”, through a tendency to process local details at the expense of global meaning (Frith, 1989; Happé & Frith, 2006).
The development of these cognitive functions has primarily been investigated in early childhood, where ToM has been found to improve in ASD across time, although not to a NTD level of performance (Pellicano, 2010a; Serra, Loth, van Geert, Hurkens, & Minderaa, 2002; Steele, Joseph, & Tager-Flusberg, 2003). Some improvement has been found in EF, where children with ASD in general seems to improve at the same rate as NTDs (Griffith, Pennington, Wehner, & Rogers, 1999; Pellicano, 2010a), and the development of LB has not been well investigated, but the results from Pellicano (2010a) suggest a loss of the advantage in the ASD-group across a three-year period.

In adolescence, very little is known about the development of any of these three domains, however; stable performance in ASD has been suggested in ToM (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994) and EF (Ozonoff & McEvoy, 1994), while LB has to the best of our knowledge not been investigated at this stage of development. No studies have yet examined all of the three cognitive domains simultaneously in transition to adolescence.

ASD remains to be understood as a life-long condition (American Psychiatric Association, 2013), and adolescence is a time characterized by hormonal, physical, mental and social changes (Blakemore, 2012a). In individuals with neurotypical development (NTD), adolescence is characterized by psychological changes, including changes in identity and changes in social relationships and interactions. It is also known to be a period of significant structural and functional change in the brain, especially in areas relating to social function (for reviews, see Blakemore, 2012a; Blakemore, 2012b; Blakemore & Choudhury, 2006; Pfeifer & Blakemore, 2012). Yet, very little is known about these psychological and physiological changes during adolescence in ASD; one study did find accelerated cortical thinning in adolescents with ASD in brain areas commonly associated with social and executive functions, and that the rate of cortical thinning was related to executive impairment and
social symptoms (Wallace et al., 2015), while other studies have found an improvement in the ASD symptomatology in adolescence (Gray et al., 2012; Seltzer, Shattuck, Abbeduto, & Greenberg, 2004).

Only a few studies have examined all three cognitive domains in ASD simultaneously at any stage in development (Brunsdon et al., 2014; Cantio, Bilenberg, Madsen, Jepsen, & White, 2016; Lai et al., 2012; Lam, 2013; Pellicano, Maybery, Durkin, & Maley, 2006), and only one of these has studied all three cognitive domains over time (Pellicano, 2010a).

The children in Pellicano’s (2010a) longitudinal study were 4-7 years of age at baseline and follow-up data was collected approximately three years later. At baseline, LB was found to differentiate the ASD and NTD group the best, but deficits in ToM and EF also differentiated the groups. At follow-up the individuals with ASD still showed ToM and EF deficits and enhanced LB. However, while a significant improvement was seen in ToM and EF over time in both groups, LB was stable in the ASD group, while the NTD group improved, decreasing the gap between the groups.

It is therefore a crucial question as to how cognition manifests later in development in individuals with ASD and how the cognitive domains are linked over time, hence why the present study aims to look at the next stage of development. Thus the objective is to investigate how the cognitive profile in ASD develops into adolescence.

Methods

We examined all three cognitive domains in a group of high-functioning children with ASD as well as a group of neurotypical developed (NTD) children, comparable in age, gender, and IQ at two time points (see table 1).

The study was approved by the Ethical Committee, Region of Southern Denmark (20090071) and the Danish Data Protection Agency (2010-41-4350).
**Participants.** In the initial study (Time 1, T1) a total of 82 children (age: 8-12 years) with an ICD-10 diagnosis of Pervasive Developmental Disorder (F84.0-84.9) and an IQ>70 were identified and invited from two Child and Adolescent Mental Health Service Clinics in the Region of Southern Denmark by searching the Patient Administrative System (PAS). Children were excluded if they were hearing impaired, had an organic cerebral disorder, or had first relatives suffering from schizophrenia or manic-depressive disorder. Of these children, 37 families agreed to participate and 3 additional children were recruited from schools of special education in the area. All families gave informed consent, and the children were re-assessed by trained clinicians at the time of testing on the basis of ADOS, ADI-R and the previous diagnostic information and re-diagnosed according to DSM-IV criteria, and at that time 5 children no longer fulfilled diagnostic criteria for an ASD. Additionally 4 children had an estimated IQ<70 (using 3 verbal and 3 non-verbal tasks from WISC-III), leaving a total of 31 participants (Cantio et al., 2016).

Approximately three years later all of the participating families were contacted once again and invited to participate in the follow-up study (Time 2, T2). Of the 31 children with ASD, 22 were willing to participate (71%). Six families turned the invitation down without any given reason, while three families chose to withdraw because of the children’s current (poor) mental health. One further child refused to participate in all tasks at T2, so his performance was at floor levels and he was therefore considered an outlier and removed from the sample. In the control group a total of seven families out of 37 chose to withdraw from the follow-up (leaving 81% of the original participants); three families because the testing was to be conducted in school time, and four families without giving a reason. All NTD participants were screened for ASD traits at T1 and T2 and none of them had elevated scores on the Social Responsiveness Scale (SRS, Constantino et al., 2003) or the Social Communication Questionnaire (SCQ, formerly the Autism Screening Questionnaire, Berument, Rutter,
Lord, Pickles, & Bailey, 1999). We found no significant differences between the participating and non-participating children in terms of gender (ASD: $X^2(1)=3.044, p=0.081$; NTD: $X^2(1)=0.712, p=0.399$), age (ASD: $t(25.68)=1.743, p=0.093$; NTD: $t(35)=0.763, p=0.549$), or IQ (ASD: $t(29)=0.193, p=0.849$; NTD: $t(35)=1.402, p=0.170$), while participants with parents with higher education (>12 years of school) were more likely to drop out at T2 in the ASD group (ASD: $X^2(1)=4.178, p=0.041$; NTD: $X^2(1)=0.059, p=0.809$). No differences were found between the ASD and the NTD group in age, IQ or parental education at T2 (see table 1).

Table 1 about here

Instruments.

Theory of Mind tasks (ToM). At both time points, we used White et al.’s (2009) modified Strange Stories for children and the Frith-Happé Animations (Abell, Happé, & Frith, 2000). For the Strange Stories, we included all 5 sets of 8 stories at Time 1 but only 2 sets at Time 2: the mental stories (as a measure of ToM performance) and nature stories (as a control measure). At the end of each story, a short question was given to the child and the child’s answer was scored according to White et al. (2009). Only the mental stories were used as a measure of ToM. The Frith-Happé Animations were administered identically at the two time points, starting with two practice trials from the random animations, followed by the Goal Directed (GD) and ToM animations presented alternately. The children were asked about the content of each of the movies, and their responses were recorded and later scored for intentionality and appropriateness according to Castelli, Frith, Happé, and Frith (2002).

Executive Function tasks (EF). At both time points, we included tests of generativity (Verbal Fluency and Pattern Meanings), as well as the Cambridge Neuropsychological Test Automated Battery (CANTAB, 1996) which taps into planning (Stockings of Cambridge, SOC), working memory (Spatial
Span, SSP, and Spatial Working Memory, SWM) and flexibility (Intra-Extra Dimensional Set Shift, IED).

**Generativity.** Two Verbal Fluency tasks were used: letter and category fluency (Turner, 1999). Participants were required to generate as many different words as possible in 60 seconds for each condition: letter (F) and category (animal). Answers could not be proper nouns. We recorded the number of correct answers after removing doublets and incorrect answers. The Pattern Meanings task involved six meaningless line drawings (Benton, 1968) and the children were shown one meaningless drawing at a time and asked what the drawing looked like. The investigator gave alternative suggestions for the first picture (practice item) and the children were allowed to return to previously viewed items and give as many responses as possible within the time limit of 90 seconds (scored according to Bishop & Norbury, 2005). The total number of answers was recorded.

**CANTAB.** The CANTAB was administered according to the user manual: instructions were read aloud and the tasks were presented on a tablet. As a measure of planning the Stocking of Cambridge (SOC) task was included. The outcome measures were number of problems solved in the minimum number of moves, initial thinking time, and subsequent thinking time. As a measure of working memory the Spatial Span task (SSP) was included, where the span length was recorded, as well as the Spatial Working Memory task, where strategy scores and double errors at 4-, 6- and 8-box problems were recorded. As a measure of flexibility, the Intra-extra-dimensional shift task (IED) was included, where the errors at stage 8 was the outcome measure, as it is at this stage the intra- to extra-dimensional shift occur.

**Local bias.** At time 1, two tasks within Local Bias was included: Hooper Visual Organisation Test (Hooper, 1983) and Embedded Figures Test (Spreen & Benton, 1969). None of these could detect any significant differences between the groups as both showed ceiling effects. Therefore an alternative,
and more widely used, task was included at Time 2 instead of the two previous tasks: Children’s Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971). The test was presented in two laminated picture books and each child was instructed to search for the target shape (triangle or house shape) in each picture and to draw over the target with a whiteboard pen. The test was administered according to the manual allowing up to 180s for each picture and the child had to tell whenever he/she had given his/her answer. Outcome measures were number of correct answers and response time. However, all children were relatively fast responders (the majority of answers were faster than 5 seconds and most children gave up after about 10 seconds if they were unable to find the task figure right away). The response times were therefore too imprecise to use for further analysis. The participants were not allowed to have more trials at the same target.

Results

As the primary aim was to investigate how cognition manifests over time in individuals with ASD, we converted all raw scores to z-scores in relation to the NTD group’s mean and standard deviation at Time 1, and used these values for all further analyses. Table 2 shows the z-scores for all the cognitive measures for both groups at both time points.

Table 2 about here

Theory of Mind (ToM).

Strange stories. A 2 (story type) by 2 (time) by 2 (group) ANOVA revealed a main effect of story type ($F(1,49)=58.214, p<0.001$) with performance being generally worse on the mental state stories, a main effect of time ($F(1,49)=4.332, p=0.043$) with performance generally improving from T1 to T2, and a main effect of group ($F(1,49)=30.021, p<0.001$) with poorer performance in the ASD group. A significant interaction between story type and group was identified ($F(1,49)=37.371, p<0.001$); the ASD group performance was worse than that of the NTD group for the mental stories ($t(49)=5.814,$
but not for the nature stories ($t(49)=1.257, p=0.215$). Time did not interact with diagnostic group, indicating a relative stability of mentalizing deficits for the ASD group over time, which was supported by a correlation on the mental stories between the two time points within this task ($r=0.453, p=0.001$), although this did not hold in the groups individually (ASD: $r=-0.028, p=0.903$; NTD: $r=0.344, p=0.063$).

*Frith-Happé Animations:* A composite was computed by averaging the intentionality and appropriateness scores, and a 2 (animation type) by 2 (time) by 2 (group) ANOVA revealed a main effect of time ($F(1,49)=22.969, p<0.001$) with performance improving over time. Additionally, a main effect of group was found ($F(1,49)=33.363, p<0.001$), where the ASD group performed worse than the NTD group. We did not find a main effect of animation and no significant interactions were found; as time did not interact with group again, this indicates a stability of these difficulties for the ASD group over time, supported by a significant correlation across the time points (ToM condition: $r=0.474, p<0.001$; within groups, ASD: $r=0.352, p=0.117$, NTD: $r=0.399, p=0.029$).

*ToM composite:* We found a strong correlation between the tasks at T1 ($r=0.563, p<0.001$), with a trend towards a correlation in each group separately (ASD: $r=0.371, p=0.089$; NTD: $r=0.206, p=0.054$), and at T2 ($r=0.401, p=0.004$), which held in the ASD group ($r=0.574, p=0.005$) and a trend was seen in the NTD group ($r=0.356, 0.054$). We combined each individual’s z-scores on these two ToM measures to create a ToM composite at each time point. We found that ToM performance was correlated across these two time points in the whole sample ($r=0.494, p<0.001$) and within the NTD group ($r=0.378, p=0.039$), but not in the ASD group ($r=0.086, p=0.711$). An ANOVA on this total ToM composite revealed a main effect of group ($F(1,49)=47.457, p<0.001$) with the ASD group consistently performing worse than the controls, and a main effect of time ($F(1,49)=9.612, p=0.003$)
indicative of improvement across the whole sample, but no interaction between the two
\(F(1,50)=0.742, p=0.393\), see Figure 1), indicating stability of the ToM impairment.

Executive function (EF).

Generativity. As the tasks within generativity were generally correlated (e.g. Pattern Meanings at
T2 with Verbal Fluency Total at T1: \(r=0.300, p=0.032\); and at T2: \(r=0.394, p=0.004\)) a generativity
composite score was made by averaging the two Verbal Fluency conditions, and thereafter averaging
these with the number of correct answers on Pattern Meanings. On this generativity composite score,
an ANOVA revealed that there was no main effect of group \(F(1,49)=3.850, p=0.055\) or time
\(F(1,49)=0.008, p=0.929\), but an interaction between the two \(F(1,49)=8.144, p=0.006\), indicating that
the ASD group improved across time \((t(21)=-2.206, p=0.039)\) while the NTD group was stable \((t(29)=-0.970, p=0.391)\). A strong correlation was found across time on the generativity composite \((r=0.463, p=0.001)\) and in the groups separately (ASD: \(r=0.611, p=0.003\); NTD: \(r=0.468 p=0.009\)).

CANTAB: Although we only found a correlation between two tasks on the CANTAB battery
(SSP T1 and SWM T1: \(r=0.491, p=0.024\); T2: \(r=0.453, p=0.039\)), we chose to combine the four
CANTAB tasks into a composite on a theoretical basis by averaging the individual test scores. While
different tasks showed improvement, degradation or a lack of change over time, none of the individual
tasks showed any main effects of group or interactions between group and time. An ANOVA on the
combined CANTAB score revealed that there was a main effect of time \(F(1,49)=28.833, p<0.001\) but
not group \(F(1,49)=0.019, p=0.901\), and no interaction between the two \(F(1,49)=0.164, p=0.687\). Additionally, a strong correlation was found across time \((r=0.435, p=0.001)\), which held in the NTD
group \((r=0.489, p=0.006)\) but not in the ASD group \((r=0.320, p=0.158)\).
EF tasks combined. Although we did not find a correlation between generativity and CANTAB at T1 ($r=0.080$, $p=0.731$) or at T2 ($r=0.333$, $p=0.140$), the CANTAB and generativity composite scores were averaged to provide an EF total score based on the theoretical assumption that EF is a unitary concept (Hill, 2004). EF at T1 was strongly correlated to EF at T2 ($r=0.563$, $p<0.001$). An ANOVA revealed that there was a significant main effect of time ($F(1,49)=14.232$, $p<0.001$), with increasing performance over time, but there was no main effect of group ($F(1,49)=1.830$, $p=0.182$, see Figure 2). There was a significant interaction between time and group ($F(1,49)=5.872$, $p=0.019$), indicating a significantly improved performance in the ASD group ($t(20)= 3.769$, $p=0.001$) that was not seen in the NTD group ($t(29)= 1.110$, $p=0.276$, see Figure 2).

Figure 2 about here

Local Bias (LB). As the tasks changed between the two time points, the T2 results could not be $z$-scored on the basis of the NTD group’s performance at T1. Instead, performance at T2 was $z$-scored on the basis of NTD performance at T2; while this would not allow us to detect any absolute change in LB ability over time, we were primarily interested in the relative change between the groups. We found a non-significant trend for the two embedded figures tasks to be correlated ($r=0.271$, $p=0.057$) that was highly significant in the ASD group ($r=0.756$, $p<0.001$), but not in the NTD group ($r=-0.077$, $p=0.690$). No such relationship was found between HVOT and either of the embedded figures tasks and, given the HVOT taps global rather than local processing, we chose only to include the EFT and the CEFT in further analyses. An ANOVA revealed that there was no main effect of time as expected ($F(1,49)=0.010$, $p=0.922$), but also no main effect of group ($F(1,49)=1.009$, $p=0.320$) and no interaction between time and group ($F(1,49)=0.035$, $p=0.852$, see Figure 3).

Figure 3 about here
Relationship between cognitive domains. At T1 we found a strong correlation between the ToM composite and the EF composite across the whole sample ($r=0.383$, $p=0.006$), although this did not hold in the groups separately (ASD: $r=0.218$, $p=0.342$; NTD: $r=0.287$, $p=0.125$).

At T2 we found a similar trend towards a correlation between ToM and EF ($r=0.266$, $p=0.060$), which was significant in the NTD group ($r=0.428$, $p=0.018$), but not in the ASD group ($r=0.222$, $p=0.333$). No correlations were found with the LB domain.

Across time, we found correlations between T1 EF performance and T2 ToM performance ($r=0.331$, $p=0.018$), although this did not hold in either group separately. The ToM composite at T1 was not predictive of any cognitive domains at T2 but we did find a relationship between the Frith-Happé Animations at T1 and generativity at T2 ($r=0.284$, $p=0.044$), although this would not withstand Bonferroni correction; this correlation was seen in the NTD group ($r=0.452$, $p=0.012$) but not in the ASD group ($r=0.212$, $p=0.357$).

Likewise, the LB composite at T1 was not found to be predictive of any cognitive domain at T2, nor to any of the individual tasks in the two other domains (all $r$’s<0.264).

A step-wise regression analysis using cognitive composite scores as dependent variables was performed to determine the early predictors of later cognitive skills. T1 Full-Scale IQ (FSIQ) in combination with group significantly predicted ToM scores at Time 2 ($R^2=.44$, $F(2,50)=19.05$, $p=0.007$), with group accounting for 34.9% of the variance. Only EF performance at T1 was found to significantly predict later EF performance ($R^2=.32$, $F(1,50)=6.04$, $p=0.018$). FSIQ predicted LB performance at T2 ($R^2=.18$, $F(1,50)=10.69$, $p=0.002$).

Given that the Frith-Happé animations and the generativity tasks seemed to be driving the relationship between ToM and EF, a further step-wise regression was performed with generativity/Frith-Happé animations at T2 as dependent variables, and FSIQ, age, generativity and Frith-Happé
animations at T1 as predictors. This showed that T1 generativity predicted T2 generativity ($R^2=.21$, $F(1,50)=13.34, p=0.001$). When generativity was removed as predictor variable, only FSIQ significantly predicted the variance of generativity ($R^2=.11$, $F(1,50)=6.011, p=0.018$), and when FSIQ was removed, the Frith-Happé animations accounted for 8% of the variance ($R^2=.08$, $F(1,50)=4.33$, $p=0.043$). Likewise, the T1 Frith-Happé animations predicted performance on the Frith-Happé animations at T2 ($R^2=.22$, $F(1,50)=14.17, p<0.001$), and when removed, only FSIQ significantly predicted the variance ($R^2=.09$, $F(1,50)=5.84, p=0.019$). Generativity did not significantly contribute to the variance even when FSIQ was removed from the analysis.

When step-wise regression was performed on the ASD group data alone, similar results appeared: FSIQ predicted ToM at T2 ($R^2=0.19$, $F(1,20)=4.43, p=0.049$); T1 EF performance predicted T2 EF performance ($R^2=0.40$, $F(1,20)=12.51, p=0.002$); and T1 LB and FSIQ together predicted T2 LB performance ($R^2=0.67$, $F(1,20)=18.60, p<0.001$) with LB accounting for 57.4% of the variance. When the relationship between generativity and the Frith-Happé animations was examined further in the ASD group, nothing predicted the variance in the T2 Frith-Happé animations, but T1 generativity and FSIQ predicted variance in T2 generativity ($R^2=.45$, $F(1,20)=9.22, p=0.002$). When generativity was removed from the latter analysis, no predictive variables remained.

**Individual patterns of impairment**

Patterns of impairment in each individual with ASD were studied. To detect individuals in the ASD group with deviant performance on each measure, any NTD group outliers performing more than 1.65 standard deviations (SDs) below the NTD group mean were removed in order to obtain a better estimate of normal performance, regardless of NTD children who might have performed abnormally on any one task. Deviant performance was defined as below the 5th centile (1.65 SD) of NTD group
performance (White et al., 2006).

Figure 4 shows that the distribution of significant impairments changed very little over time, with the majority of children having a ToM impairment at both time points. Only two children had impairments in EF at each time point and, while both of these children also had a ToM impairment at T1, one of these children was only categorised as impaired in EF at T2. This child still performed well below the NTD group mean on the ToM domain at T2, however ($z=-1.32$). Two further children with ToM impairments at T1 were categorised as showing no impairments at T2.

Discussion

This three-year follow-up cohort study investigated cognitive changes in ASD from late childhood into adolescence. It was found that while understanding of other’s minds does improve between the ages of 10 and 14, Theory of Mind (ToM) impairment persists in adolescents with ASD relative to their peers. We similarly observed development in Executive Function (EF) in both ASD and NTD groups, but the ASD group actually progressed to the extent that an EF impairment was no longer detectable relative to the NTD group. Remarkably, ToM difficulties were present in the majority of children with ASD at both time points.

Compared to previous literature, our results support Pellicano (2010a); Serra et al. (2002); Steele et al. (2003) who found that ToM does improve over time. However, it seems from the literature (Pellicano, 2010a; Serra et al., 2002) and the present study that ToM improves at about the same rate in ASD and NTD, while the ASD group do not catch up with the NTD group across time. Other studies regarding the development of ToM (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994) have not been able to detect any absolute improvement in the ASD group. However, these studies are more than 20 years old when ToM tasks may have been less sensitive in the adolescent populations tested. In
conclusion, these improvements indicate that development of ToM is present into adolescence in individuals with ASD as well as individuals with neurotypical development.

We found improvements in EF in both groups over time, which differs from the majority of previous findings (Griffith et al., 1999; Ozonoff & McEvoy, 1994; Pellicano, 2010a). However, evidence of age related improvements in both ASD and NTD children has been reported previously by Happé, Booth, Charlton, and Hughes (2006), who discovered that children with ASD performed better at EF tasks at older ages and performed at a level equivalent to the NTD children in several EF measures.

No change in the presence of a LB relative to the NTD group was found. We chose to focus on local processing, using only the EFT for analysis at T1, and to change the task version at T2 to avoid ceiling effects, but there was little evidence of a local processing bias in the ASD group compared to the NTD group at either time point. This differs from Pellicano (2013), who found that the ASD group showed a distinct LB at T1 but lost their advantage on this domain at T2. Interestingly, Pellicano’s participants at T2 were of a similar age to our individuals at T1. This may indicate that a LB is primarily seen in early childhood but reduces in later childhood and adolescence in ASD. It is equally possible that our LB tasks were not sensitive enough to detect this processing bias.

Although many children with ASD improved over time, there were varying individual patterns of development between T1 and T2. A few individuals showed immense improvements in one domain (e.g. two individuals with ASD in the ToM-domain), while others had a slight decreased performance.

The percentages of poor performers in ToM and EF did not change much, but two additional individuals showed no impairment in any of the domains at T2, indicating that individuals with ASD can improve their cognitive functions over time. Moreover, one individual had improved his ToM skills at T2, so he had EF impairment only. However, the ToM performance by this individual was still not at
a NTD level (borderline). Nevertheless, at T2 the majority of the individuals with ASD still had ToM impairments.

Hence, the results indicate an inter-relationship between ToM and EF can be interpreted in two ways: 1) ToM impairments are much more common than EF impairments, possibly indicating a greater role in the cognitive impairments in ASD. 2) ToM is dependent on EF since intact ToM is not seen without executive control and impaired EF is solely seen with impaired ToM. This last hypothesis was proposed by Perner and Lang (1999) and later used as an argument of EF as the primary deficit by Pellicano (2007).

While the first hypothesis presupposes that ToM and EF are separable and independent, or their relationship exist because of interfering elements in the tasks that are used (EF argument: ToM tasks contain EF components, J. Russell, Saltmarsh, & Hill, 1999; ToM argument: EF tasks require mentalizing abilities in order to understand the experimenter's expectations for the task, White, 2013), the second hypothesis presupposes that EF consists of a number of basic functions that must be acquired in order to develop ToM.

Although developing, ToM impairments are still observed in older children and young adolescents with ASD, while the ASD group seem to catch up to a NTD level in terms of EF. Pellicano (2010a) hypothesized, on the basis of previous literature, that ToM improvements could be more apparent in the early development in young children. Although development was seen in the ASD-group, it seems that ToM impairments continue to be apparent into adolescence in high-functioning individuals with ASD and presumably beyond. This next step of development from adolescence into adulthood is still to be investigated. These findings indicate that there is reason to prioritize the ToM domain in clinical practice in older children and adolescents with ASD, as well as in young children.
References


Figure Legends

Figure 1: Individual development in the ToM domain (z-scores based on individuals with neurotypical development, NTD, performance at Time 1) for both groups.

Figure 2: Individual development in the EF domain (z-scores based on individuals with neurotypical development, NTD, performance at Time 1) for both groups.

Figure 3: Individual development in the LB domain (z-scores based on individuals with neurotypical development, NTD, performance at Time 1) for both groups.

Figure 4: Venn diagram to show the number of children with ASD, who participated at both time points and who displayed significant ToM, LB or EF impairments (below the 5th centile of the NTD group); the children outside the diagram represent those in which none of these impairments were detectable.
Table Legends

Table 1: Characteristics of participants

Table 2: Mean z-scores for cognitive measures at baseline (T1) and follow-up (T2) for the ASD group (N=21) and the NTD group (N=30).
Table 1: Characteristics of participants

<table>
<thead>
<tr>
<th></th>
<th>ASD N=21</th>
<th></th>
<th>NTD N=30</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>10.70 (1.50)</td>
<td>14.36 (1.45)</td>
<td>10.96 (1.26)</td>
<td>14.32 (1.15)</td>
</tr>
<tr>
<td>Performance IQ (SD)</td>
<td>104.38 (15.61)</td>
<td>-</td>
<td>111.30 (18.56)</td>
<td>-</td>
</tr>
<tr>
<td>Verbal IQ (SD)</td>
<td>104.05 (17.42)</td>
<td>-</td>
<td>103.03 (16.06)</td>
<td>-</td>
</tr>
<tr>
<td>Full-scale IQ (SD)</td>
<td>105.48 (15.95)</td>
<td>-</td>
<td>109.47 (18.58)</td>
<td>-</td>
</tr>
<tr>
<td>Parent’s educational level (higher/lower)</td>
<td>15/6</td>
<td>15/6</td>
<td>20/10</td>
<td>20/10</td>
</tr>
<tr>
<td>ADOS total</td>
<td>10.91 (3.77)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADI total</td>
<td>28.86 (11.73)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2: Mean z-scores for cognitive measures at baseline (T1) and follow-up (T2) for the ASD group (N=21) and the NTD group (N=30).

<table>
<thead>
<tr>
<th>Measure</th>
<th>ASD T1, Mean (SD)</th>
<th>ASD T2, Mean (SD)</th>
<th>NTD T1, Mean (SD)</th>
<th>NTD T2, Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strange Stories, Mental</td>
<td>-1.82 (1.66)</td>
<td>-1.55 (0.98)</td>
<td>0.00 (1.00)</td>
<td>0.05 (1.00)</td>
</tr>
<tr>
<td>Strange Stories, Nature</td>
<td>-0.35 (0.76)</td>
<td>0.16 (0.75)</td>
<td>0.00 (1.00)</td>
<td>0.40 (0.61)</td>
</tr>
<tr>
<td>Animations, ToM</td>
<td>-1.54 (1.54)</td>
<td>-0.26 (1.35)</td>
<td>0.00 (1.00)</td>
<td>0.52 (0.86)</td>
</tr>
<tr>
<td>Animations, GD</td>
<td>-1.54 (1.66)</td>
<td>-0.40 (1.56)</td>
<td>0.00 (1.00)</td>
<td>0.45 (0.60)</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>-0.73 (0.90)</td>
<td>0.64 (1.81)</td>
<td>0.00 (1.00)</td>
<td>0.48 (0.89)</td>
</tr>
<tr>
<td>Pattern Meanings</td>
<td>-0.64 (0.75)</td>
<td>-1.40 (0.59)</td>
<td>0.00 (1.00)</td>
<td>-1.14 (0.64)</td>
</tr>
<tr>
<td>IED*</td>
<td>0.00 (1.19)</td>
<td>0.49 (0.93)</td>
<td>0.00 (1.00)</td>
<td>-0.13 (1.23)</td>
</tr>
<tr>
<td>SWM*</td>
<td>-0.34 (0.92)</td>
<td>0.49 (0.91)</td>
<td>0.00 (1.00)</td>
<td>0.53 (0.79)</td>
</tr>
<tr>
<td>SOC*</td>
<td>0.19 (0.68)</td>
<td>0.35 (0.75)</td>
<td>0.00 (1.00)</td>
<td>0.49 (1.14)</td>
</tr>
<tr>
<td>SSP</td>
<td>-0.06 (0.94)</td>
<td>0.51 (0.91)</td>
<td>0.00 (1.00)</td>
<td>0.76 (0.88)</td>
</tr>
<tr>
<td>EFT* / CEFT*&lt;sup&gt;A&lt;/sup&gt;</td>
<td>-0.25 (0.86)</td>
<td>-0.20 (1.14)</td>
<td>0.00 (1.00)</td>
<td>0.00 (1.00)</td>
</tr>
</tbody>
</table>

* Reversed because of expected direction for group effect

<sup>A</sup> Based on NTD T2 mean/SD