

University of Groningen

The development of theory-of-mind and the theory-of-mind storybooks

Blijd-Hoogewys, Els Maria Arsene; Blijd-Hoogeweys, E.M.A.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2008

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Blijd-Hoogewys, E. M. A., & Blijd-Hoogeweys, E. M. A. (2008). *The development of theory-of-mind and the theory-of-mind storybooks*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 5



Temporal patterns in the development of Theory of Mind. A longitudinal study in children with PDD-NOS.

Abstract: Children with PDD-NOS are often not included in Theory-of-Mind research, although they experience specific Theory-of-Mind problems. This study is aimed at establishing whether their problems are only of quantitative nature, whether they display a spontaneous yet delayed Theory-of-Mind development, whether their Theory-of-Mind problems are subject to age changes and whether Theory-of-Mind growth is reflected in more adequate social behavior over time. Thirty children with PDD-NOS (3-8 years old) were tested six times over a period of 20 months with the Theory-of-Mind Storybooks, which is a comprehensive test measuring basic Theory-of-Mind knowledge. Their Theory-of-Mind knowledge was compared to two control groups. At the first measurement, children with PDD-NOS displayed quantitative problems with belief, desire and mental-physical tasks. There were no qualitatively different Theory-of-Mind score profiles indicative of a deviant development. The Theory-of-Mind problems and developmental progress were subject to age trends. At the last measurement, children with PDD-NOS have caught up their delay, on the one side as a result of aging and on the other side as a learning effect due to repeated measurements. The correlation between Theory-of-Mind and everyday social behavior was positive but relatively weak. Children with PDD-NOS display Theory-of-Mind problems. Their development is delayed, not deviant. One of the most interesting findings of the present study is the observation of a temporary anomaly and regression in the Theory-of-Mind development of children with PDD-NOS. This anomaly is qualitatively similar to the one found in typically developing children. Consistent with the observed delay in Theory-of-Mind development, it occurs at a later age than seen in typically developing children.

This chapter is submitted as: Blijd-Hoogewys, E.M.A., Van Geert, P.L.C., Serra, M., & Minderaa, R.B. (submitted). Temporal patterns in the development of ToM. A longitudinal study in children with PDD-NOS.

INTRODUCTION

Theory-of-Mind problems in autism

Theory-of-Mind (abbreviated as ToM) is the ability to attribute mental states to oneself and others and to use these attributions in understanding, predicting and explaining behavior of oneself and others (Premack & Woodruff, 1978; Mitchell, 1997). ToM is considered an important condition for understanding the social environment and for showing socially adequate behavior (Astington & Jenkins, 1995). Children with autism have serious problems with social interaction. Linked with this, it has been suggested that they lack a ToM (postulated by Baron-Cohen et al., 1985). They are known to have problems with first-order beliefs (inferring about what another person thinks) and second-order beliefs (inferring about what another person thinks about yet another person's thoughts) (for reviews see Baron-Cohen, 2000; Perner et al., 1989; Yirmiya et al., 1998). This difficulty in mental state attribution has been found in both children and adults with autism (Beaumont & Newcombe, 2006).

In the beginning, ToM research in autism concentrated on the absolute presence or absence of ToM. Later, this point of view was nuanced; it was postulated that ToM impairments may occur in various degrees (Baron-Cohen et al., 1993) and that children with autism may not be ToM blind but rather ToM weak-sighted. This degree of weak-sightedness is subject to inter- and intra-individual variability. ToM may differ over individuals but also within an individual the level of ToM comprehension is not fixed, but may develop over the years. It has been reported that children with autism can eventually develop early ToM aspects (such as the understanding of desire and pretending) and basic ToM aspects (such as the understanding of perception/knowledge, false beliefs and hiding) (Peterson et al., 2005). However, more advanced ToM aspects appear more problematic (such as the understanding of second-order false beliefs, lies and jokes) (Steele et al., 2003; Brent et al., 2004). Still, there are individuals with autism who not only pass first-order belief tasks but also second-order belief tasks, although they are either older than typically developing children or have a higher mental age (Baron-Cohen, 1989b; Happé, 1995; Sparrevohn & Howie, 1995; Yirmiya et al., 1996; Ziatas et al., 1998). It often concerns individuals with AS (Asperger's syndrome), an ASD (autism spectrum disorder, which refers to the whole spectrum of autism, also referred to as autism) with no cognitive or language delays. It appears that they have less severe ToM problems than individuals with HFA (high

functioning autism, an autistic disorder of normal intelligence) (Baron-Cohen et al., 1997a, b, 2001; Dissanayake & Macintosh, 2003; Jolliffe & Baron-Cohen, 1999; Klin, 2000) or even do not differ from TD (typically developing) (Bowler, 1992; Dahlen & Trillingsgaard, 1993; Ozonoff et al., 1991; Ziatas et al., 1998). However, other studies found that individuals with AS do evidence qualitative difficulties in using mental state terms context-appropriately (Kaland et al., 2005).

Passing ToM tasks seems strongly related to a higher verbal mental age (Dissanayake & Macintosh, 2003; Happé, 1995; Sparrevohn & Howie, 1995; Tager-Flusberg & Sullivan, 1994; Ziatas et al., 1998). Children with lesser variants of ASD often have better developed language skills, hence resulting in better results on ToM tasks. However, equating the social cognitive ability of these individuals with a typical ToM seems not legitimate, since their apprehension is probably not achieved through social cognitive mechanisms as seen in TD but rather through learning experiences and acquired compensation strategies (Hermelin & O'Connor, 1985; Tager-Flusberg & Sullivan, 2000), also known as the 'hacking' hypothesis (Bowler, 1992; Happé, 1994; for a critical note see Brent et al., 2004). This compensation strategy has been illustrated both in AS (Frith & Happé, 1999) and AD (autistic disorder) (Dissanayake & Macintosh, 2003). So, possibly through compensation, individuals with ASD can to a certain degree mask their social cognitive deficits. Nonetheless, several studies have shown that they perform much slower (Bowler, 1992; Kaland et al., 2007) and that more advanced ToM tasks or more naturalistic tasks are often failed (Abell et al., 2000; Baron-Cohen et al., 2001; Heavey et al., 2000; Klin, 2000; Klin et al., 2004; Roeyers et al., 2001). Though, it has been reported that adults with ASD can succeed on even the latter sort of tasks (more complex tasks), if they are provided with sufficient support through structure and context in these tasks (Ponnet et al., 2005). However, in daily live they often remain awkward and socially impaired, since daily live is far more complicated than a test situation, less structured and often requires rapid and intuitive understanding of mental states of others. A more cognitively mediated route seems not always sufficient.

A continuum of Theory-of-Mind problems

In recent years, the focus of ToM research has shifted from a categorical approach to observing individual differences and nuances in ToM functioning (for a substantial overview, see Repacholi & Slaughter, 2003).

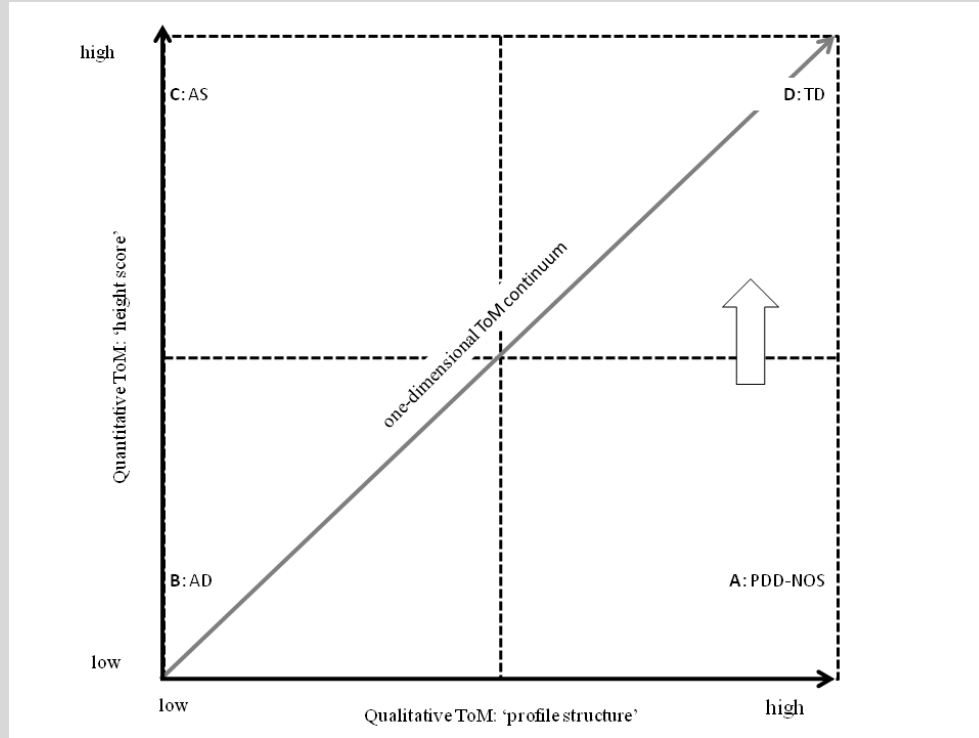
Looking at the apparent degree in ToM impairments, the idea of a continuum of ToM functioning comes to mind, with at the extreme low end a complete lack of ToM and at the extreme high end a complete mastery of ToM as seen in TD. However, the disadvantage of such a one-dimensional continuum is that it impedes the distinction among subgroups with different ToM problems, for instance among children with AD, AS or PDD-NOS (Pervasive Developmental Disorder Not Otherwise Specified, a lesser variant of autism), whereas there exists support for the assumption that the degree of ToM impairment is linked to the severity of the autism diagnosis independent of the level of intelligence and language ability (Tager-Flusberg, 2003).

To improve the making of distinctions among subgroups, we suggest a continuum which distinguishes two dimensions, one referring to the quantitative aspects and one to the qualitative aspects of ToM. Quantitative aspects relate to the degree of successful use of the skill in question, for instance as measured by the score on a ToM test. Qualitative aspects concern the composition of a ToM-score in terms of underlying components (e.g. first-order and second-order beliefs), it concerns the profile structure of the ToM skill in question. In this way, four quadrants can be distinguished on the basis of these two theoretical dimensions (see Figure 1).

Children in the 'B' quadrant show both quantitative and qualitative deficits in ToM, which can be considered to be characteristic of children with severe AD and a deviant ToM development. Children in quadrant 'C' also show a deviant qualitative ToM profile, but the few skills they are equipped with are not differently used in comparison with the TD group. Children with AS might be seen as representatives of this type of ToM profile, being able to successfully answer first-order belief tasks, but having more problems with second-order belief tasks and more naturalistic tasks. Children in the 'A' quadrant are characterized by a normal qualitative profile, but they are delayed in terms of successful use of these skills. Hypothetically, children with PDD-NOS belong to this quadrant. It is imaginable that a significant subgroup of these children will be able to eventually complete the developmental trajectory but with some delay (therefore the upward arrow from A to D quadrant). The 'D' quadrant defines the group of children with a typically developing ToM. The B, C and D quadrants have been well established through research. However, the

A quadrant, where children with PDD-NOS are supposedly situated, is at present less clearly understood¹.

Figure 1: Two-dimensional ToM continuum depicting four quadrants of ToM functioning



Legend. AS= Asperger syndrome, AD= autistic disorder, TD= typically developing.

Development of Theory-of-Mind in autism

It is clear that children with autism evidence ToM problems, but also that they undergo some ToM development: they do come to master some ToM abilities. How this development takes place, is as yet unclear and it is also

¹ Alternatively, the A quadrant could be indicated as the group with 'Acquired ToM', the B quadrant as 'Bad ToM', the C quadrant as 'Compensated ToM' and the D quadrant as 'Developed ToM'.

unclear whether their ToM development is either delayed or deviant (a question that has been brought up for the first time in Baron-Cohen, 1991b).

A delay is when the same developmental sequence is followed but at a slower pace. A deviance is when a child wanders off from the TD trajectory: in such cases the level of ToM functioning is lower as can be expected by the mental age; the ToM ability is out of line with other skills from related domains (Burack, 1992), and the pattern of abilities regarding ToM components is qualitatively different from the TD group.

Research has not been conclusive on whether children with autism evidence either a delayed or deviant ToM development, also because, according to Burack (1992), the definitions of delay and deviance have not been used correctly. Burrack (1992) argued that the evidence Baron-Cohen (1989b) brought up for a delay was actually pointing out a deviant ToM development. Subsequent studies continued referring to a delay in ToM (for reviews see Happé, 1995; Yirmiya et al., 1998). Illustrative is a recent study, which found that the developmental order of ToM components in children with AD differs from that in TD (Peterson et al., 2005). These differences were highly regular and scalable. Children with autism had a distinctive, autism-specific difficulty with false belief understanding (Peterson et al., 2005). This evidences a deviant ToM development. Yet again the difference was referred to as a delay. Oppositely, in children with PDD-NOS a delayed ToM development has been demonstrated, though a deviant ToM development could not be ruled out (Serra et al., 2002).

Next to the question whether there is a delayed or deviant ToM development in children with autism, one can also question if and how this development is reflected in their social behavior. Does it lead to less social inadequate behavior? It has been reported that ToM functioning correlates with everyday social behavior in children with autism, as measured by the parent-rated VABS (Vineland Adaptive Behaviour Scales, sub-domain Socialization; Tager-Flusberg, 2003) and the VABS additional items (Frith et al., 1994). However, there is also research that did not confirm this correlation (Dissanayake & Macintosh, 2003).

Relevance of longitudinal Theory-of-Mind studies

Despite the fact that the development of ToM ability in children with autism is plausible, significant ToM changes over time are rarely recorded (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994; considering a time period of 7 and 3 years respectively). In our view, this might be

attributed to the fact that mainly older children were tested (too old to find meaningful ToM changes, though it has been shown that there is no evidence for a plateau in ToM development in children with autism in their early adolescence, see Steele et al., 2003), also perhaps too many children were situated in the B quadrant (where no development can be expected). Next to that, often only one or a few ToM tasks were used (these tasks have higher standard errors and lower reliability than more comprehensive tasks using compound scores; if a greater number of tasks is used, ToM changes are indeed found, again see Steele et al., 2003). Finally, longitudinal research has often confined itself to a very limited amount of repeated measurements (often just two measurements, one at the beginning and one at the end), not giving much evidence of the developmental process, and most conclusions on ToM development are even based on extrapolation of findings from different single cross-sectional measurements.

Ideally, research looking into ToM development in children with autism should involve subjects in an age range in which ToM can be expected to develop, using a comprehensive ToM task and applying more frequent, repeated measurements over an extended period of time. Despite the recommendations of prominent researchers to undertake longitudinal research (e.g. Baron-Cohen, 1989b; Sparrevohn & Howie, 1995) and using comprehensive ToM tests (Astington, 2001; Hughes et al., 2000; Serra et al., 2002; Tager-Flusberg, 2003; Wellman & Liu, 2004), little longitudinal research has been carried out in children with autism up till now (for exceptions see Serra et al., 2002; Steele et al., 2003). Consequently, our knowledge of the developmental trajectory of ToM is mainly based on single, cross-sectional measurements (for meta-analyses, see Wellman et al., 2001; Wellman & Liu, 2004). However, in order to understand real individual trajectories, subjects should most preferably be followed from the stage of ToM precursors to the level of mature ToM.

Following a substantial number of children during their entire trajectory of ToM development is virtually undoable. A compromise between following real individual trajectories and confining oneself to single cross-sectional measurements is to measure temporal trajectories using a dense time-serial design. In such research, a group of children with various starting ages are followed for a limited period of time, which does not cover the entire developmental trajectory, but which lasts long enough to cover a substantial part of the total trajectory, and thus to give information about the temporal properties of change. Characteristics of developmental routes may give more insight in ToM abilities of children

with ASD, and may also give insight in whether they have a deviant or delayed ToM development.

Relevance of studying children with PDD-NOS

Research into ToM problems involves mostly individuals with AD. Individuals with AS are included far less frequently and few studies include children with PDD-NOS. It is not surprising that the latter group is left out in research, since the diagnosis PDD-NOS is ambiguous (it is considered a heterogeneous rest group) and can often not easily be measured by means of standardized clinical instruments. However, it has been speculated that PDD-NOS is more prevalent in the population than AD (Chakrabarti & Fombonne, 2001; Fombonne, 2003; Lingam et al., 2003). Furthermore, in the scope of a ToM continuum, they may represent an important link. If ToM problems are related to severity in pervasive developmental diagnosis (see Tager-Flusberg, 2003), it can be expected that children with PDD-NOS, a lesser variant of autism, have lesser ToM problems. We hypothesized that they are situated more at the higher functioning end of the qualitative dimension, but do evidence some quantitative ToM problems (the supposed A quadrant).

It has been postulated that children with PDD-NOS have a lesser variant of autism because they evidence fewer autistic symptoms, especially in the domain of repetitive, stereotyped activities (Walker et al., 2004). They also have milder problems with expressive communication; they seem to be better in their use of syntax and pragmatics (Paul et al., 2004). Since language is considered an important factor in ToM functioning (Astington, 2001; Astington & Baird, 2004), especially syntactic ability, both in TD (de Villiers, 2000) and autism (Tager-Flusberg, 2000; compare with Colle et al., 2007), one could reason that children with PDD-NOS might also perform better on ToM tasks. Indeed, the scarce research undertaken in children with PDD-NOS has shown that, compared to children with AD, they perform better (Buitelaar et al., 1999). Compared to TD (Serra et al., 2002) or compared to children with language impairment (Sicotte & Stemberger, 1999), they perform worse. Compared to socially immature children, with acting-out behavior and social anxiety, no differences were found (Muris et al., 1997). However, it should be noted that the latter research was based on a single first-order-belief task (Smarties task), which is likely to have lower discriminative power for this clinical group.

For the current study, we have chosen to test solely children with PDD-NOS. Taking into account the ToM two-dimensional continuum, we expect children with PDD-NOS to have specific ToM problems, which are primarily of quantitative nature (they perform worse than TD children, but do not show a different ToM profile). In addition, we expect to find a delay and not a deviance in their ToM development.

The present study

In summary, three points are pertinent. First, studies have shown that there are individual differences in ToM functioning in children with ASD, characterized by varying degrees of ToM impairments. We assume there is a two-dimensional ToM continuum, consisting of a qualitative and a quantitative dimension. Based on these dimensions, four quadrants were formulated, with at the extreme lower left severe ToM problems and at the extreme upper right typically developing ToM. Second, a comparable continuum is presumed also within ASD ranging from AD over AS to PDD-NOS (Mayes et al., 1993; Sicotte & Stemberger, 1999). Children with PDD-NOS are an overlooked group in research and could also be considered an overlooked quadrant within the two-dimensional ToM continuum. Concerning this continuum, children with PDD-NOS are supposed to be situated in the A-quadrant: on the qualitative dimension of ToM they are comparable to TD children, but they are low on their respective quantitative level of ToM. Third, ToM develops over time, also in children with ASD. We assume that children with PDD-NOS evidence a delayed ToM development, not a deviant one.

In this article, a PDD-NOS group is compared with a TD group. In doing so, cross-sectional and longitudinal researches are combined, with the aim of obtaining more insight into the nature of their ToM problems and ToM development.

As regards the cross-sectional part, ToM problems can express themselves through quantitative and qualitative differences in comparison with the TD group. Applied to test scores, quantitative differences imply that the ToM total score is lower than in the TD group. Qualitative differences mean that there are differences in the ToM score profile (differences in ToM sub-scores) in comparison with the TD group. The latter can be considered an indication for a deviance in the underlying developmental patterns. The quantitative and qualitative differences between the TD and PDD-NOS group are likely to show age differences,

with ToM difficulties being more obvious in some age periods than in others.

Concerning the longitudinal part, the PDD-NOS group is repeatedly tested over a period of 20 months, with intervals of four months. As already mentioned, we expect a delayed ToM development in the PDD-NOS group. Next to that, we expect the ToM development to depend on age, with learning gradually declining with age, as expected from TD data. Considering this, we would like to remark that ToM development need not be continuous. Temporary regressions can occur, as has been described in motor and verbal development (Gershkoff-Stowe and Thelen, 2004). After mastering an ability children can have a temporary relapse before the ability consolidates. This is a well-known characteristic in developmental psychology, which has also been demonstrated in ToM development in TD children. In a previous study on ToM development in TD, we reported two regressions in ToM functioning, namely one at the age of 56 months and one at the age of 72-78 months (Blijd-Hoogewys et al., submitted b). Such anomalies can find their counterpart in research of children with PDD-NOS. However, we expect this anomaly to occur at a somewhat later age, because we expect children with PDD-NOS to have a delayed ToM development. Serra and colleagues (2002; Figure 2, page 9) already reported such a regression in ToM functioning in children with PDD-NOS, namely approximately at the age of 60 months, which was indeed later as the one seen in TD.

Additionally, the relationship between ToM problems and social behavior is examined. It has been reported that ToM functioning correlates with everyday social behavior, as measured by the parent-rated VABS (Tager-Flusberg, 2003). Thus, it can be expected that ToM problems to correspond with these behavioral scores and that gains in ToM score over time also correspond with gains in social functioning as measured by these behavioral scores. However, if ToM scores of children with PDD-NOS are mainly an expression of the children's cognitive and reflective understanding (referring to the 'hacking' hypothesis), it is likely that progress in ToM test scores does not or only weakly relate to everyday social functioning.

In this study, five questions are addressed: 1) Do children with PDD-NOS evidence ToM problems? 2) Are their ToM problems more distinct at certain ages? (For instance, ToM problems are more pronounced in younger children than in older children.) 3) Do children with PDD-NOS evidence a developmental progress in ToM functioning? 4) Is the nature of this

developmental progress subject to age changes? 5) Is there a meaningful relationship between their ToM scores and behavioral scores? Or differently put, if there is a ToM growth, is this also visible in their daily social behavior?

An indirect goal of this article is to present evidence for a two-dimensional ToM continuum. The presence of ToM weaknesses but also of ToM strengths and the susceptibility to learning effects may locate children with PDD-NOS more towards the right end of the supposed ToM continuum, closer to ToM functioning as seen in TD.

METHOD

Subjects

Children with PDD-NOS

These children were referred to an outpatient clinic for child and adolescent psychiatry. After an extensive psychiatric examination (which included parent interviews and play observations with the child), the children were diagnosed as having PDD-NOS according to DSM-IV criteria (APA, 1994). All children demonstrated, in varying degrees, the triad of impairments observed in pervasive developmental disorders with a central deficit in social relatedness.

Originally the clinical group consisted of 40 children. Their ages ranged from three up to and including eight years ($M=78.88$ months old, $sd=22.13$; gender ratio=4:1). All children participated in an extensive psychological examination which included the assessment of intelligence and language comprehension, and tests that could validate the diagnosis. Depending on the age of a child, different tests were applied.

Intelligence of children up to 6 years was tested with the SON-R 2½-7 years (Snijders-Oomen Non verbal intelligence scale: Tellegen et al., 2003); for children older than 6 years, the WISC-R was applied (Wechsler Intelligence Scale for Children-Revised: Wechsler, 1974; Dutch version, 1986).

Concerning language comprehension, two Dutch language tests were used. For 3-6 year olds, the Reynell was administered (test for receptive language comprehension; Van Eldik et al., 1997); and for 6-9 year olds, the TvK (Taaltest voor Kinderen, Language Test for Children; Van Bon, 1982) was used (subtests 'vocabulary' and 'sentence construction').

To support the clinical diagnosis, two additional tests were administered: the Vineland Adaptive Behavior Scales (VABS) (Sparrow et al., 1984; Dutch version: Research group Developmental Disorders, State University Leiden, 1995) and the Children's Social Behavior Questionnaire (CSBQ; Luteijn et al., 2000; Hartman et al., 2006; Dutch version: VISK; Luteijn et al., 2002). The VABS is an interview in which parents are questioned about the daily social skills of their child. The following parts of the VABS were used: Communication, Daily Living Skills and Socialization. For each child the discrepancy between the Vineland age equivalent (in months) and the chronological age (in months) was computed (VA-CA). The parents also filled in the CSBQ, which focuses on autism-related behavior. This questionnaire can be used to facilitate selection of ASD samples for research purposes (Hartman et al., 2006).

Based on the results of the aforementioned psychological examination, we excluded 10 children. Six children were excluded due to gross intellectual delay ($IQ < 70$)²; they also had extremely low language scores. Four other children were excluded due to extremely low scores on the CSBQ (score < 16), indicative of a questionable validity of their clinical diagnoses. In the end, 30 children remained in the sample ($M = 78.95$ months old, $sd = 15.58$)³. There were 24 boys and 6 girls, resulting in a gender ratio of 4 to 1, which is the average gender ratio found in children with autism (compare Yeargin-Allsopp et al., 2003). Concerning the VABS, they had large and negative discrepancy scores in Receptive Language, Playing Skills, Interpersonal Relationships and Coping Skills as can be expected in children with ASD (Serra et al., 2002). Their problems with Expressive Language and Daily Living Skills were less profound (compare Paul et al., 2004) (see Table 1). The CSBQ scores of our group are comparable to those known for children with HFA (High Functioning Autism) and PDD-NOS (Hartman et al., 2006, Table V) (see Table 2).

² We excluded children with an IQ-score lower than 70 because we expected ToM-data to become unreliable in these children, due to for example not understanding task instructions or to difficulties processing the verbal information.

³ This clinical research group has already been reported on in a previous paper which was aimed at the psychometric properties of the ToM Storybooks (Blijd-Hoogewys et al., 2008).

Table 1: Test results of the children with PDD-NOS at the first measurement

	Age (in years)							Total n=30
	3 n=2	4 n=3	5 n=4	6 n=11	7 n=5	8 n=5		
ToM-TB	42.50 (2.12)	42.67 (7.51)	53.50 (7.55)	72.55 (16.42)	68.00 (6.96)	76.40 (7.16)	64.90 (16.34)	
VIQ	97.50 (10.61)	95.00 (24.43)	94.25 (14.45)	80.73 (14.53)	94.60 (11.13)	94.40 (11.57)	92.97 (13.48)	
PIQ	116.00 (46.67)	104.00 (18.52)	109.50 (21.92)	94.32 (19.30)	108.80 (17.04)	107.20 (13.42)	103.32 (19.92)	
VABS								
Communication								
Receptive language	35.50 (13.44)	33.00 (13.00)	43.33 (10.69)	45.00 (8.74)	48.00 (1.00)	44.20 (8.17)	43.25 (9.21)	
Expressive language	-16.21 (19.91)	-27.42 (14.65)	-30.70 (12.22)	-36.88 (9.87)	-43.42 (3.17)	-55.89 (9.43)	-38.11 (14.18)	
Expressive language	47.50 (9.19)	43.33 (14.19)	73.33 (27.79)	66.00 (19.21)	67.60 (23.83)	68.40 (18.06)	63.75 (20.38)	
Community	-4.21 (15.67)	-17.08 (11.57)	-0.70 (24.49)	-15.88 (17.87)	-23.82 (21.15)	-30.69 (17.30)	-17.61 (19.12)	
Daily living skills	44.00 (7.07)	41.00 (10.82)	63.33 (47.35)	67.60 (20.61)	70.20 (10.03)	74.60 (20.50)	64.32 (19.57)	
Socialization	-7.71 (13.55)	-19.42 (8.47)	-10.70 (19.71)	-14.28 (20.77)	-21.22 (6.84)	-24.49 (20.94)	-17.04 (16.87)	
Interpersonal relationships	32.50 (13.44)	35.33 (15.37)	61.33 (47.35)	50.00 (23.09)	64.80 (22.90)	53.40 (33.63)	51.64 (26.72)	
Play and leisure time	-20.71 (4.83)	-26.08 (10.36)	-2.37 (38.51)	-32.58 (25.87)	-26.62 (21.33)	-45.69 (35.40)	-29.08 (27.15)	
Coping skills	19.50 (0.71)	37.00 (14.11)	51.33 (27.06)	47.50 (15.30)	54.60 (11.33)	63.80 (26.53)	48.96 (19.96)	
Total	-21.21 (9.78)	-23.08 (11.47)	-36.03 (17.47)	-33.98 (15.35)	-36.82 (11.31)	-36.29 (16.87)	-32.86 (16.33)	
Coping skills	38.50 (17.68)	39.33 (10.41)	42.00 (17.35)	53.90 (21.40)	62.20 (11.68)	60.80 (16.72)	52.68 (18.27)	
Total	-16.21 (28.40)	-14.08 (7.39)	-24.70 (28.63)	-29.38 (21.90)	-29.22 (10.10)	-38.29 (17.27)	-27.86 (19.17)	

Note. VABS: means and standard deviations; every first row depicts VABS interview age equivalent; every second row (in italic) depicts VABS interview discrepancy score (for each child the discrepancy between the Vineland age equivalent in months and the chronological age in months was computed for the different subscales).

Table 2: *CSBQ results of the children with PDD-NOS: means (standard deviations)*

Subscales	This study PDD-NOS (N=30)	Hartman et al., 2006		
		HFA (N=102)	PDD-NOS (N=544)	NC (N=232)
Total score	48.30 (17.72)	47.22 (15.37)	37.84 (15.94)	10.28 (9.05)
Tuned	10.03 (4.52)	12.14 (5.23)	12.36 (5.60)	4.24 (3.71)
Social	13.67 (4.98)	10.22 (4.66)	7.75 (4.91)	1.38 (1.99)
Orientation	8.07 (3.10)	7.71 (3.80)	6.42 (3.80)	1.45 (2.10)
Understanding	5.27 (3.03)	8.89 (3.29)	6.12 (3.91)	1.94 (2.13)
Stereotyped	6.40 (3.37)	5.19 (4.22)	2.90 (3.18)	0.74 (1.24)
Change	2.24 (1.43)	3.09 (1.27)	2.30 (2.16)	0.55 (1.13)

Legend. NC = normal control

Two control groups of typically developing children

We had a cross-sectional TD pool of 324 children (3-11 years old) at our disposal, who were tested in a previous study (Blijd-Hoogewys et al., submitted a). These children had no language acquisition problems that could have hampered their ToM performance. From this pool of children two control groups were selected. One control group was matched for age with the clinical group at the first measurement (referred to as group A) and a second control group was matched for age with the clinical group at the fifth measurement (referred to as group B) (see later for the design of this study, Table 3). By doing so, we were able to compare scores obtained through longitudinal testing with scores for which we are sure that no learning effects due to repeated measurements have taken place.

The 30 children in the PDD-NOS group were matched on the basis of age alone, with children from the TD group⁴. Matching consisted of looking for the closest age mate in the typically developing group, and four additional age mates with ages around that of the PDD-group child in question. The difference with the closest age mate was on average 6 days. The average difference in age between the youngest and oldest member of each group of five was about 40 days. The choice of five age matches is based on the wish to obtain the most reliable estimation of the scores, while

⁴ Gender differences in ToM functioning have been established before (Blijd-Hoogewys et al., submitted a), however they do not apply when the entire age range is concerned. Therefore, matching based on age alone was considered to be sufficient.

keeping the ages of the matches as close as possible to the age of the PDD-NOS child in question. The estimated age-specific score of the TD children is calculated by taking the average of the scores of the 5 age matches. By taking the average of five, the estimated scores come close to the ideal of a norm score on the variable of choice.

Material

The ToM Storybooks

We used the ToM Storybooks (see Blijd-Hoogewys et al., 2008; Serra et al., 2002). This instrument consists of six storybooks in which a main protagonist, named Sam, experiences all kinds of feelings, desires and thoughts. The child is asked a variety of questions about Sam's experiences.

The tasks focus on ToM and associated aspects that children develop between the ages of three to six years old. They cover five components: 1) Recognition of emotion, 2) Distinction between physical and mental entities, 3) Understanding that seeing leads to knowing, 4) Prediction of behaviors and emotions from desires, and 5) Prediction of behaviors and emotions from beliefs (see Appendix A).

There are 34 tasks in total. The tasks are spread over the six storybooks and are naturally interwoven in the stories (for the order of tasks and questions, see Appendix B). Each task is illustrated with one or more full color pictures. Transitions between tasks are also accompanied by drawings, to keep the story going and furthermore to avoid too much switching between situations. The test takes 40 to 50 minutes, including a short break.

Scoring procedure

Each task incorporates one to five questions, including both test questions and justification questions (see Appendix B). There are in total 74 test questions and 18 justification questions. The answers to the test questions (for instance: Where will Sam look for his rollerblades? In the toy trunk or in the box?) result in 1 or 0 points (correct or incorrect; maximum score=74). Because justifications are considered to better reflect the ToM knowledge of a child, most tasks also include 'justification questions' (for instance: Why will Sam look in the box?). The justification questions result in 2, 1 or 0 points, depending on the sort and correctness of the mental state terms spontaneously used by a child (maximum score=36). In order to enable a standardized evaluation of the justifications, a category system is

developed, based on the category system used by Rieffe (1998), on different categories from Wellman (1990), and on an exploration of the empirical data. For each justification question, correct answer categories were determined (see the right four columns of Appendix B).

On the basis of these scores, a ToM total score can be calculated (maximum=110), which can also be transformed into a ToM quotient (abbreviated as ToM-Q). This is a standardized norm score, with an average of 100 and a standard deviation of 15. There are norm scores for children from three up to twelve years old. Norms were obtained by applying a non-linear smoothing method over the raw data (for more details on the norming procedure of the ToM Storybooks, we refer to Blijd-Hoogewys et al., submitted a). In order to chart ToM-strengths and ToM-weaknesses of a child, sub-scores can be calculated: 1) Emotion recognition (maximum score= 14), 2) Distinction between physical and mental entities (real-mental, real-imaginary and close impostors; maximum score=24+8+12), 3) Predicting behaviors and emotions on the basis of desires (maximum score=5+12), and 4) Predicting behaviors and emotions on the basis of beliefs (maximum score=26+6). Understanding that seeing leads to knowing does not form a sub-score, for this component contains too few questions.

Psychometric qualities

The ToM Storybooks have good psychometric qualities: the internal consistency is good (Cronbach's $\alpha=.90$, with age correction), as is the test-retest reliability ($r=.86$) and the inter-rater reliability (Cohen's Kappa =.97-.99) (Blijd-Hoogewys et al., 2008). A regression model showed that chronological age, mean verbal intelligence and mean non-verbal intelligence explain a significant proportion of variance in ToM-score of the children with PDD-NOS from the current study ($R=.77$, $R^2=.59$, $F=19.32$, $p<.001$). Considering the contribution of the individual predictors, the results indicated that both chronological age and verbal intelligence are clearly linked to the ToM total score (chronological age: $\beta=.70$, $p<.001$; verbal IQ: $\beta=.34$, $p=.05$). This is in concordance with other studies that have found that ToM-knowledge can be predicted by chronological age and verbal intelligence (e.g. Hughes et al., 1999; Serra et al., 2002) and studies that illustrate the strong connection between linguistic skills and ToM ability (e.g. Astington & Baird, 2004). Some studies also emphasize non-verbal intelligence as an important factor (Muris et al., 1997; Carlson et al., 2002). In the case of the ToM Storybooks however, non-verbal intelligence lost predictive power when verbal intelligence was also included ($\beta=-.06$,

$p=.65$), probably due to the correlation between verbal and non-verbal intelligence ($r=.37$, $p=.05$).

Parallel versions of the ToM Storybooks

There are three additional versions of the ToM Storybooks. Each of them contains identical tasks and is based upon the same underlying test structure, but consist of different protagonists (version Lotje, version Pieter and version Hanna) and different stories (e.g. going to the zoo instead of going to the park). In comparison with version Sam, each of the additional versions consists of six storybooks, holding 34 tasks in total. The additional versions can be considered parallel versions since they mutually correlate highly ($r=.87-.95$ for the ToM total score, $p<.001$). The parallel versions are designed to be used in longitudinal and repeated testing research, avoiding the effect of habituation, recognition and trivial learning effects that might result from mere repetition with the same instrument, say version Sam.

The VABS

As mentioned earlier, the VABS (Vineland Adaptive Behavior Scales; Sparrow et al., 1984) was used in the screening phase, as a support for the clinical diagnosis. At the end of the research period, a second VABS interview was administered from the parents. Since this instrument provides an objective and standardized assessment of socialization skills (Fombonne et al., 1994; Frith et al., 1994), it was used as a validity check for the presumed ToM growth at the last measurement.

To compare pre- and post-experiment VABS results, a VABS developmental quotient was computed by dividing the VABS developmental age by the chronological age (according to the method as used by Gevers et al., 2006). Such a quotient accounts for maturation during the period in which the measurements took place.

Procedure and Design

The children with PDD-NOS were tested individually at the out-patient clinic. Their ToM-knowledge was assessed at six different times with an interval of approximately four months (resulting in a total research period of 20 months) (see Table 3). During the research period, children did not attend a ToM training or a social skills training.

The comparisons of the PDD-NOS group with the TD group were done at measurement one and five, which both involved the Sam version of

the test. The reason for not taking measurement six as a point of comparison is that this involved another ToM Storybook version, whereas the cross-sectional TD data were entirely based on the Sam version of the ToM Storybooks. For comparisons within the PDD-NOS group, all six measurements were taken into account.

Table 3: *Overview of groups tested*

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Clinical (n=30) ¹	+ Sam	+ Lotje	+ Pieter	+ Hanna	+ Sam	+ Lotje
Control						
A (n=30)	+ Sam	-	-	-	-	-
B (n=30)	-	-	-	-	+ Sam	-

Note. ¹ n=27 with 6 measurements, n=3 with 5 measurements (two measurements missing at time 2, one at time 6), mean number of months between times 1 and 2: 3.96 (SD=.33), between times 2 and 3: 4.15. (SD=.39), between times 3 and 4: 3.99 (SD=.32), between times 4 and 5: 3.88 (SD=.32), between times 5 and 6: 4.07 (SD=.54), and overall between times 1 and 5: 20.07 (SD=.41).

Statistical method

In view of the complex nature of some of the statistical tests required by our research questions and the small and irregularly distributed samples, we used random permutation techniques, and more generally, Monte-Carlo analyses (abbreviated as MC). This method entails a simulation of the test statistic at issue as based on the null hypothesis (Good, 2001; Manly, 1997; Todman & Dugard, 2001). It is highly flexible and robust, and is free of parametric assumptions.

RESULTS

ToM problems in children with PDD-NOS

Concerning the first question – whether children with PDD-NOS evidence ToM problems – the ToM scores of the PDD-NOS group at the first

measurement were compared with predicted scores from control group A (obtained through stratified sampling, see before). Quantitative differences in ToM were charted by comparing ToM total scores and ToM Q-scores; qualitative differences were charted by comparing the profiles of ToM sub-scores.

For each variable, the 30 scores of the PDD-NOS children were compared with the 30 scores based on averages of five age mates. The hypothesis is that TD children score higher on the variable than their age mates with PDD-NOS. The hypothesis was tested by means of a procedure in which the scores of the PDD-NOS and TD group were randomly permuted over two groups of 30 subjects, and differences in average scores were compared with the observed differences. Random permutations were done pairwise, i.e. over pairs of similar-age children. By doing so, the age composition of the randomly permuted groups is always similar to that of the real PDD-NOS.

Children with PDD-NOS had significantly lower ToM total scores (MC, $M=67.60$, $sd=18.23$ versus $M=77.70$, $sd=12.14$; $p<.001$) and ToM-Q scores (MC, $M=85.10$, $sd=21.28$ versus $M=100.00$, $sd=6.76$; $p<.001$) at the first measurement than control group A (see Table 4, column 'Measurement 1'). More concretely, in terms of norm scores, children with PDD-NOS scored one standard deviation lower than their TD peers (theoretically ToM-Q=100, with $sd=15$). In terms of total scores, children with PDD-NOS scored about 5/6th pooled standard deviation lower than their TD peers.

Table 4 also gives an overview of ToM-strengths and -weaknesses in the PDD-NOS group based on their ToM sub-scores (note that these are relative scores, with a maximum score of 10). At the first measurement, they had significantly lower scores on tasks concerning mental physical distinctions (RM, RI & CI), belief-action, belief-emotion and desire-action. No significant differences were found for tasks concerning emotion recognition and desire-emotion.

We can conclude that the PDD-NOS group showed a significant quantitative difference on ToM total scores and most ToM sub-scores. The question then arose whether this difference corresponded either with a delay or a deviance. The latter would be visible as a qualitative difference in the ToM score profile. In order to check this, we compared the differences in the shape of ToM sub-score profiles. If these are parallel, there is no evidence for any qualitative difference.

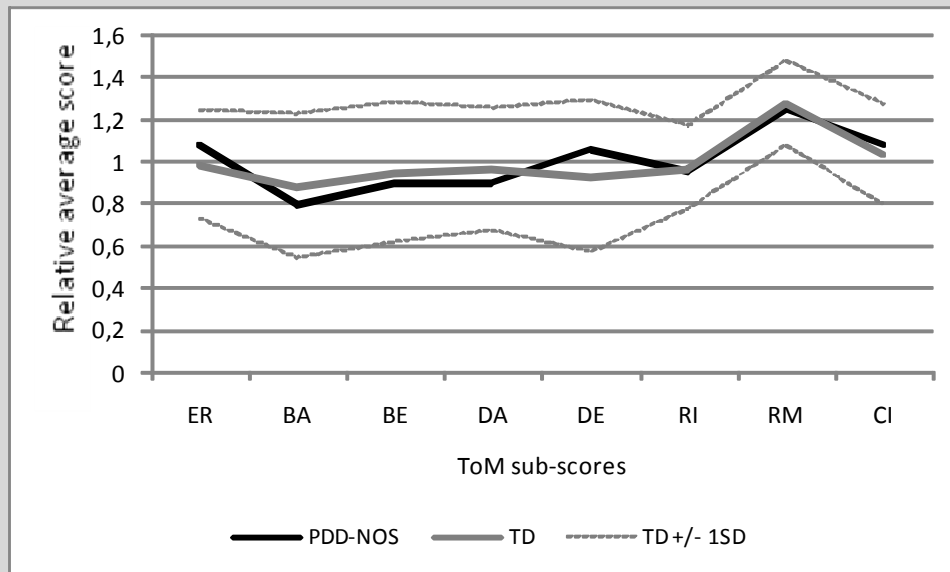
Table 4: ToM results at measurement 1 and measurement 5

	Measurement 1		Measurement 5		M1 vs. M5 PDD-NOS
	PDD-NOS	Control A	PDD-NOS	Control B	
Total score:					
Total ToM-score (0-110)	67.60 (18.23)	77.70 (12.14)	85.23 (12.97)	85.85 (8.03)	MC, p<.001 (ns, p=.41)
ToM-Q score	85.10 (21.28)	100.00 (6.76)	98.97 (20.92)	99.73 (5.67)	MC, p<.001 (ns, p=.43)
Subscores:					
<i>Emotion Recognition (0-14):</i>	6.81 (2.03)	7.04 (1.21)	7.75 (1.13)	7.79 (0.54)	(ns, p=.21) (ns, p=.18) T, p=.05
<i>Mental Physical:</i>					
Real-mental items (0-24)	6.01 (1.59)	6.96 (1.12)	7.82 (1.04)	7.77 (0.64)	MC, p<.001 (ns, p=.58) T, p<.001
Real-imaginary items (0-8)	7.88 (2.05)	9.04 (1.07)	9.46 (0.97)	9.63 (0.44)	MC, p<.001 (ns, p=.18) T, p<.001
Close impostors (0-12)	6.78 (2.31)	7.34 (1.26)	7.83 (1.76)	8.27 (0.75)	MC, p=.07 (ns, p=.11) T, p=.05
<i>Desires:</i>					
Predicting action (0-5)	5.67 (2.35)	6.62 (1.13)	7.27 (1.62)	7.15 (0.65)	MC, p=.01 (ns, p=.65) T, p=.01
Predicting emotion (0-12)	6.69 (2.18)	7.19 (1.22)	7.72 (2.37)	7.24 (0.84)	(ns, p=.14) (ns, p=.86) T, p=.05
<i>Beliefs:</i>					
Predicting action (0-26)	5.01(2.70)	6.35 (1.45)	7.28 (2.54)	7.27 (1.32)	MC, p=.01 (ns, p=.51) T, p<.001
Predicting emotion (0-6)	5.67 (2.54)	6.91 (1.34)	7.28 (2.21)	7.97 (1.34)	MC, p<.001 MC, p=.05 T, p=.005

Legend. M1=Measurement 1, M5=Measurement 5, MC=Monte Carlo analyses, T=F-test.

In order to compare the profiles, we normalized the scores: we set the average total score of both groups to one, so as to compensate for the difference in averages, conserving the eventual differences between the sub-scores. Having done so, Figure 2 shows that all profile differences of children with PDD-NOS are considerably smaller than one standard deviation from the sub-scores profile of control group A, which leads us to conclude that there is no reason to assume that the ToM profiles of the PDD-NOS group are qualitatively different or deviant. Their absolute scores (remember that the scores in Figure 2 are normalized scores) of course are lower, indicative of a delay.

Figure 2: *Relative average sub-scores of the ToM Storybooks*



Legend. PDD-NOS= Pervasive Developmental Disorder Not Otherwise Specified. TD= Typically Developing children. SD= Standard Deviation. ER= Emotion Recognition, BA= Belief Action, BE = Belief Emotion, DA= Desire Action, DE= Desire Emotion, RI= Real vs. Imaginary, RM= Real vs. Mental, and CI= Close Impostors.

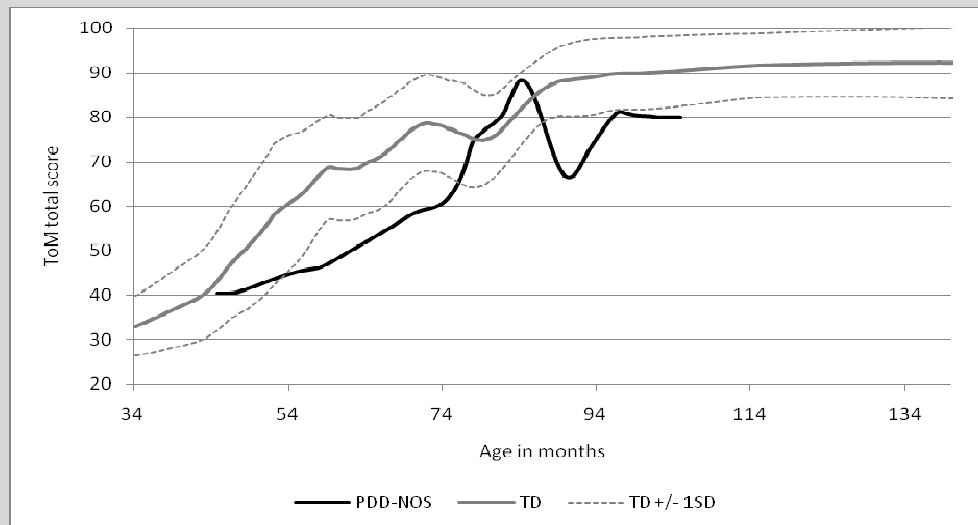
Age trends in ToM skills of children with PDD-NOS

The second question focuses on whether ToM problems of the PDD-NOS group are subject to age trends. That is, are there ages in which the differences with the TD group are more distinctive than in others? In order

to answer this question, we looked at smoothed curves of ToM total scores and ToM sub-scores. We did not consider ToM-Q scores, since by definition no age effects can be expected, since quotient scores are corrected for age.

The data were smoothed using a Loess smoothing technique (locally weighted least squares estimate; Härdle, 1991; Simonoff, 1996); a Loess 30% window model was applied. Figure 3 shows the smoothed curves of the ToM total scores for the PDD-NOS group and the total pool of TD group (N=324). Considering the latter, also the curves of plus and minus one standard deviation are depicted. The figure clearly illustrates that the PDD-NOS group had lower ToM total scores. However, the difference with TD varied with age. Until about 70 months of age, the difference was about 1.5 standard deviation from the average TD score (this is not so for the earliest ages), which is a considerable difference (Cohen, 1988; Parker & Hagan-Burke, 2007; Rice & Harris, 2005). Then the average PDD-NOS score rose above the average TD score, at 90 months it showed a sharp drop to below 2.5 standard deviations from the typical average, and then moved back to about a standard deviation less than the typical average.

Figure 3: *Smoothed curves of ToM total scores for both children with PDD-NOS and the typically developing norm group*



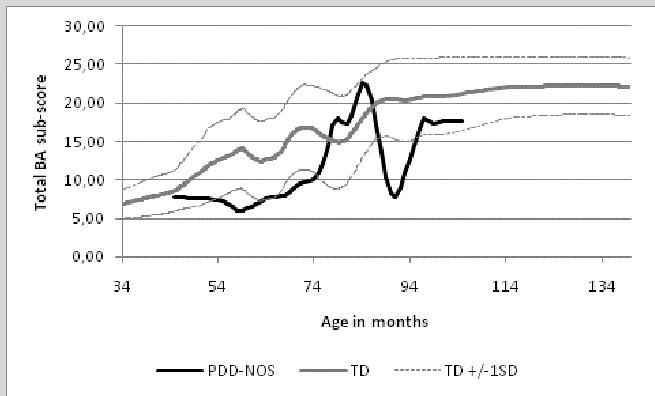
Legend. PDD-NOS= Pervasive Developmental Disorder Not Otherwise Specified. TD= Typically Developing children. SD= Standard Deviation.

Taken in isolation, these rapid fluctuations could easily be attributed to random sampling fluctuations. However, in a previous article, we could shown that the anomaly in TD - the pattern of a temporary rise followed by a dip - was neither a statistical artifact of sampling-based fluctuations nor the effect of tester-bias (Blijd et al., submitted b). Interestingly, the anomaly in TD occurred somewhat earlier (deepest regression point at 78 months), consistent with the finding that children with PDD-NOS are delayed in ToM-development in comparison to their normally developing peers.

The same was found for the ToM sub-scores. We report the graphs for four sub-scores (see Figure 4: 4A, 4B, 4C and 4D).

Figure 4: *Smoothed curves of four ToM sub-scores for both children with PDD-NOS and the typically developing norm group*

4.A. *Belief Action*



4.B. *Belief Emotion*

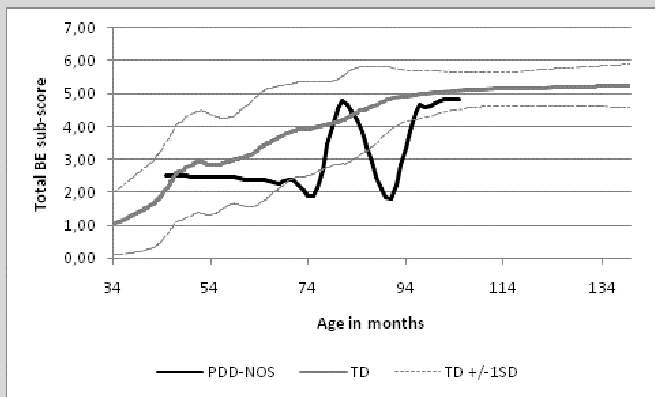
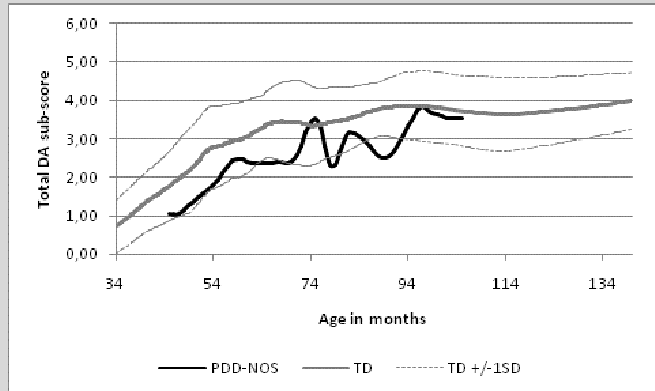
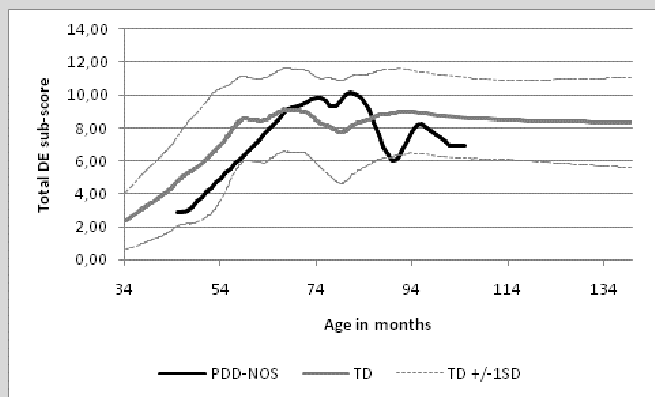


Figure 4: *Smoothed curves of four ToM sub-scores for both children with PDD-NOS and the typically developing norm group*

4.C. *Desire Action*



4.B. *Desire Emotion*



Legend. BA=Belief Action, BE=Belief Emotion, DA=Desire Action, DE=Desire Emotion, PDD-NOS= Pervasive Developmental Disorder Not Otherwise Specified. TD= Typically Developing children. SD= Standard Deviation.

Developmental progress of ToM skills in children with PDD-NOS

The third question, whether the ToM functioning of children with PDD-NOS shows a developmental progress can be answered by investigating whether they have significantly higher scores at the last measurement in

comparison with the first. We compared ToM total scores, ToM-Q scores and ToM sub-scores of the PDD-NOS group on the first and the fifth measurement with the expected scores from the two control groups (group A and B). The time interval between the first and fifth measurement is 16 months.

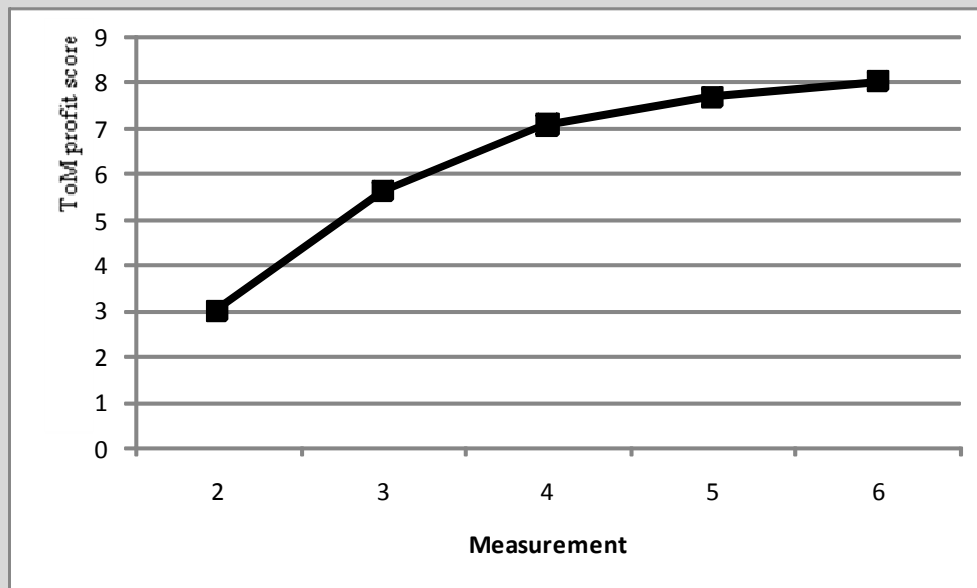
The average ToM total score of the PDD-NOS group at measurement five was significantly higher than at measurement one (MC, increase: $M=17.63$, $sd=12.17$, $p<.001$). The same was found for ToM-Q scores (MC, increase: $M=13.87$, $sd=17.12$, $p<.001$). The increases were of such an extent that at measurement five, there were no longer significant differences between the PDD-NOS group and TD group (group B) in ToM total scores (MC, $M=85.23$, $sd=12.97$ versus $M=85.85$, $sd=8.03$) and ToM-Q scores (MC, $M=98.97$, $sd=20.92$ versus $M=99.73$, $sd=5.67$) (see Table 4: column measurement 5). Also, the ToM sub-scores at measurement five were significantly higher than those at measurement one. Significant differences between the PDD-NOS group and TD group at measurement one (compared with group A) were no longer significant at measurement five (compared with group B), except for the tasks concerning belief-emotion.

There seems to be a larger ToM increase than can be expected due to mere aging, since their ToM increase is higher than that seen in the TD group. It can be assumed that the additional growth effect is due to learning effects induced by test repetition. This leads to a supplementary question: Did the repeated ToM measurements lead to learning effects?

This question is answered in the following way. We first computed percentile scores (P_{10} , P_{20} , ..., P_{100}) for the ToM total scores of the TD group ($N=324$). Then, the percentile score at measurement one of each child with PDD-NOS was determined. Taking into account this percentile score, the expected ToM total scores of the following measurements were computed and compared with the empirically found ToM total scores. Since these analyses were limited to comparisons within the PDD-NOS group, we used the total time interval of 20 months (instead of 16 months). Profit scores - the additional gain in comparison to the expected gain on basis of the cross-sectionally determined norm scores - for the second until the sixth measurement were calculated. On the basis of these profit scores, we visualized the developmental progress in the PDD-NOS group (see Figure 5). They showed a classical learning curve: at the beginning they profited more due to repeated measurements, however, the longer they were tested,

the less additional learning effect it had on their ToM scores. The additional gain in ToM total scores was statistically significant (MC, $p < .001$).

Figure 5: *Developmental progress of percentile ToM scores in children with PDD-NOS, in addition to expected progress merely due to age.*



Age trends in ToM developmental progress in children with PDD-NOS

Concerning the fourth question, whether the developmental progress is subject to age changes, we looked at the change over time in the ToM total scores and ToM sub-scores. More precisely, we looked at the extent and acceleration of ToM development at the different ages. This was defined by means of the slope of the developmental curve of each individual child, specified by a simple linear regression. Since only the PDD-NOS group was concerned, again the time interval of 20 months was considered (measurement 1 through 6).

The average slope of the ToM total scores in the PDD-NOS group was significant (see Table 5). The relationships between the slopes – the amount of increase – and age was negative, as expected (MC, $r = -.63$, $R^2 = .40$, $p < .001$). The older children were, the higher their starting score, and the lower their developmental increase possibilities. The slopes of half of

the ToM sub-scores were found to be significantly correlated with age; the slopes for the belief sub-scores did not correlate at all.

Table 5: Average slopes of ToM scores in children with PDD-NOS

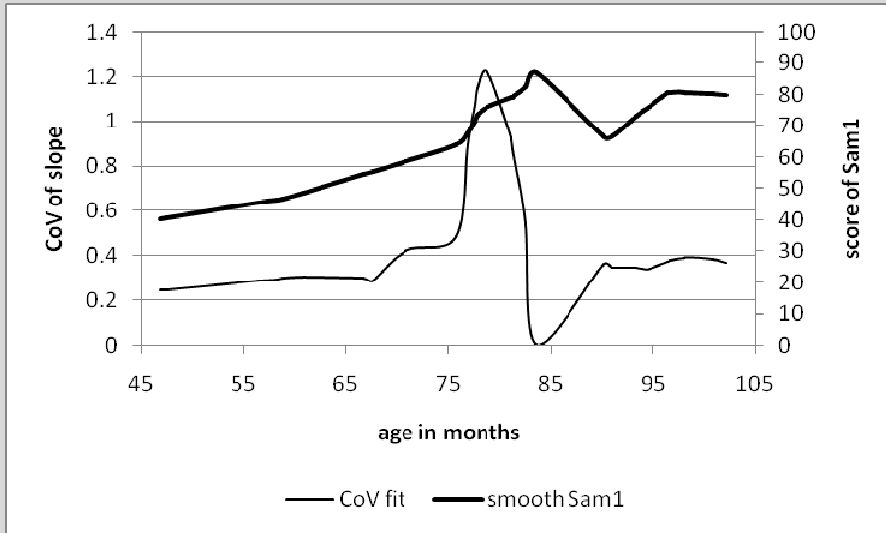
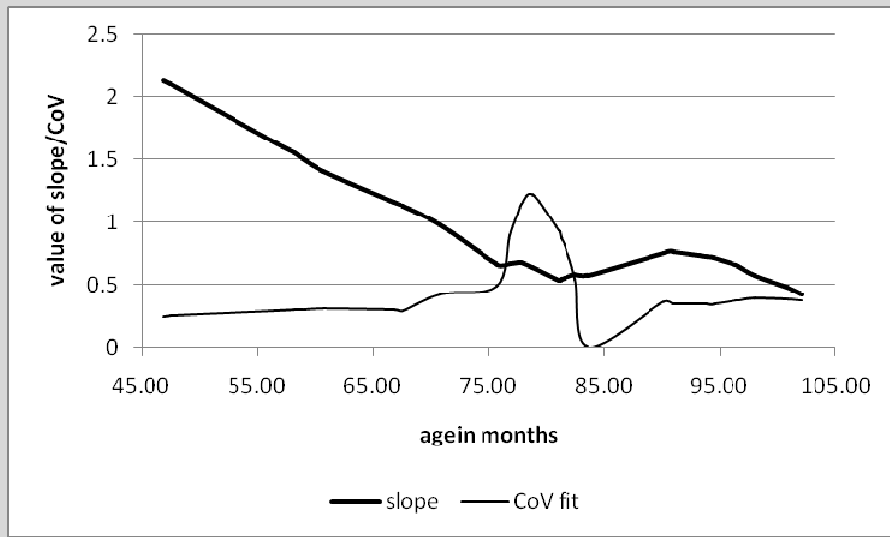
	M (SD)	r	R ²	sign.
Total score:				
Total ToM-score	0.90 (0.64)	-0.63	0.40	p<.001
Subscores:				
<i>Emotion Recognition</i>	0.06 (0.15)	-0.62	0.38	p<.001
<i>Mental Physical</i>				
Real-mental items	0.21 (0.14)	-0.44	0.19	p=.01
Real-imaginary items	0.07 (0.07)	-0.28	0.08	p=.13
Close impostors	0.06 (0.13)	-0.33	0.11	p=.08
<i>Desires</i>				
Predicting action	0.03 (0.06)	-0.28	0.08	p=.14
Predicting emotion	0.05 (0.18)	-0.59	0.35	p<.001
<i>Beliefs</i>				
Predicting action	0.34 (0.30)	-0.09	0.01	p=.65
Predicting emotion	0.05 (0.18)	-0.18	0.03	p=.33

A Loess smoothing procedure was applied to the slopes (amount of linear growth) of the ToM total scores over age (at the first measurement). The smoothed trend showed a clear decline in expected growth up to the age of 75 months, followed by a period in which the expected slopes remain more or less similar. Visual inspection of the residuals (difference between observed slope and slope as predicted by the non-linear trend curve) showed a striking increase in the residuals around the age of 75 months. This phenomenon is statistically independent of the fact that the temporal trend of the slopes flattens around that time, and could thus provide an additional indication of an anomaly in ToM-development in children with PDD-NOS.

In order to check whether the temporary increase in residuals is statistically significant (i.e. not explainable by random fluctuation of the residuals, as expected under the null hypothesis), we proceeded as follows. We first fitted a non-linear model of the squared residuals (Loess smoothing) in order to obtain a model of the temporal change in the

variance. The variance model was used to calculate the temporal evolution of the standard deviation (square root of the variance), and then to calculate the time evolution of the coefficient of variation (CoV or standard deviation divided by expected mean; see Van Geert & Van Dijk, 2002). Ideally, the coefficient of variation should be a constant value independent of the average slope and thus independent of age. Figure 6 shows that the slope decreases with age and that the coefficient of variation has a marked anomaly around the age of 80 months, which just precedes the dip in the ToM scores. The coefficient of variation anomaly is a form of anomalous variance, which is considered a 'flag' for a discontinuous change (e.g. van der Maas & Molenaar, 1992; Van Dijk & Van Geert, 2007) (for other indicators of developmental transition, see Chapter 4 of this dissertation). The peak in the coefficient of variation was significant (MC, $p < .05$).

Figure 6: *Peak in coefficient of variation*



Upper figure: Relationship between slope and Coefficient of Variation (CoV). Slope (as a measure of growth over 20 months) decreases with age. CoV is for the most part stable (as expected), but peaks around the age of 80 months. The peak is statistically significant ($p=.02$).

Lower figure: The peak in CoV just precedes the dip in the scores of the first measurement. The anomalous variance, as expressed by the sudden peak in CoV, precedes the dip, which is consistent with the assumption that the dip corresponds with a discontinuity in ToM development.

ToM skills in children with PDD-NOS: link with daily social skills

For the last question – whether there is a meaningful relationship between ToM and everyday social behavior– the correlations of ToM scores and VABS scores are calculated, and it is investigated whether gains in one correspond with gains in the other.

We found ToM Total scores to correlate with VABS raw scores, both at the first measurement (Communication, $r=.72$; Daily Living Skills, $r=.55$; Socialization, $r=.24$) and the sixth measurement (Communication, $r=.44$; Daily Living Skills, $r=.43$; Socialization, $r=.16$). Note that the correlations at the last measurement were considerably lower.

We have shown that the children displayed a significant increase in ToM scores over time (third question). Oppositely, the VABS scores were rather stable (see Table 6). Only Play/Leisure of the Socialization sub-domain and Daily Living Skills of the Community sub-domain showed a significant improvement in age equivalent score and in VABS quotient scores at the sixth measurement (QQ-plots showed normal distributions, therefore paired samples t -tests were used) (Play/Leisure: $M_2=.89$, $sd_2=.31$ versus $M_1=.60$, $sd_1=.18$, $p<.001$; Daily Living Skills: $M_2=.86$, $sd_2=.20$ versus $M_1=.79$, $sd_1=.19$, $p=.05$). However, the growth in these VABS sub-scores did not correlate with the slope of ToM total scores. The amount of VABS growth did not correspond with the amount of ToM growth, which could already be suspected through the lower correlations between ToM and VABS at the last measurement in comparison with those at the first measurement.

Table 6: Mean Scores (Standard Deviations) for VABS sub-domains

	Developmental Quotient at measurement 1	Developmental Quotient at measurement 6
Communication		
Receptive + Expressive language	0.81 (0.22)	0.80 (0.23)
Community		
Daily living skills	0.79 (0.19)	0.86 (0.20)*
Socialization		
Interpersonal relationships	0.63 (0.28)	0.62 (0.20)
Play and leisure time	0.60 (0.18)	0.89 (0.31)**
Coping Skills	0.70 (0.22)	0.64 (0.21)

Note. ** $p\leq.001$, two-tailed; * $p<.05$, two-tailed.

DISCUSSION

In summary, the following results were found. At the first measurement, children with PDD-NOS had significantly lower levels of ToM-knowledge in comparison with TD children. This was observable in their ToM total scores, ToM quotient scores and ToM sub-scores. At the end of the research period, however, children with PDD-NOS no longer showed significant differences in their ToM total scores and ToM quotient scores, in comparison with TD children. Their scores had ameliorated considerably over time.

Closer inspection of the ToM sub-scores showed that some ToM aspects were not problematic from the start (emotion recognition and desire-emotion tasks), while others did form a problem but ceased to do so at the end of the testing period (belief-action, desire-action and mental physical tasks). One ToM aspect remained problematic until the end, namely predicting emotions on the basis of beliefs. Note that false belief understanding (a belief-action task) was more difficult at the first measurement, but had improved considerably at the last measurement. These findings largely agree with the findings from Serra and colleagues (2002) who conducted a longitudinal study in children with PDD-NOS using a previous version of the ToM storybooks. In the current research, children with PDD-NOS had problems with beliefs, both in predicting behaviors and emotions. In addition, they had problems with desire-action, real-imaginary, real-mental, and close impostor tasks. Some of these ToM problems were less pronounced and therefore not significant in the study of Serra and colleagues (2002). The only contradictory finding is that Serra and colleagues found their PDD-NOS group not to have difficulties in using beliefs to predict actions, while the current PDD-NOS group did evidence such problems, which is more consistent with clinical expectations. The somewhat different findings in the current study may be attributed to the larger sample size used for this study ($n=30$ instead of $n=11$), the wider age range aimed at (3-8 years instead of 4-5 years), and the longer period of time children were followed (20 months instead of 6 months). To summarize, despite the obvious differences in design, the findings from the present study and the study of Serra and colleagues largely agree: Children with PDD-NOS evidence ToM problems in comparison with control children. So, research question number one “Do children with PDD-NOS evidence ToM problems?” can be answered affirmatively.

Further inspection of the ToM differences between the PDD-NOS and the TD group at the first measurement, suggested that these differences were subject to age trends. So, also research question number two “Are their ToM problems more distinct at certain ages?” can be answered affirmatively. At the age of 78-85 months, no differences were found between both groups. This was found in both the ToM total scores and sub-scores. This finding is not surprising, since around that time both groups underwent temporary regressions in their ToM functioning, which greatly obscured the comparison between the groups. The TD group showed an anomaly at the age of 72-78 months (which is the difference between the preceding peak and the following dip in the scores). The PDD-NOS group showed an anomaly somewhat later at the age of 85-90 months. As explained in the introduction, the occurrence of temporary regressions is a well-known characteristic in developmental psychology. Serra and colleagues (2002) reported regressions in ToM development at 56 months in TD, and somewhat later, at 60 months, in PDD-NOS. In a previous study in TD (Blijd-Hoogewys et al., submitted b), we recorded two regressions: a small one at the age of 56 months (in accordance with Serra et al., 2002) and a more profound one at the age of 72-78 months (which Serra et al., 2002 could not assess, because their research group was younger). In the current study aimed at children with PDD-NOS, the second, more profound temporary regression was found, again at a somewhat later age, namely at 85-90 months. This anomaly was supported by an additional temporary increase in the coefficient of variation found in the longitudinal data, indicative of a true discontinuous change (Van Geert & Van Dijk, 2002). This anomalous variance preceded the dip and supports the assumption that the observed dip is an expression of an underlying developmental discontinuity, for instance a change in the way children understand and answer ToM questions. To recapitulate, the TD group entered into the regression period before the PDD-NOS group did. As a result, for a certain period the PDD-NOS group appeared to be better at their ToM, at least in comparison with their peers. But then, in their turn, they entered into the regression period, while the TD group was already catching up with their normal developmental trend. As a result, the PDD-NOS group temporarily had a considerably lower ToM as expressed in their ToM scores. After a while however, the difference between both groups at the first measurement stabilized again.

The former findings can be considered the first indications of a delayed development of ToM in children with PDD-NOS. To unravel

whether there is truly a delay or a deviance, we looked at their ToM sub-score profiles. When we corrected for the delay in their scores, we found no differences between the shape of their ToM sub-scores and that of the TD group. Put differently, there were no qualitative differences, which is in accordance with the prediction based upon the two-dimensional ToM continuum. The qualitative similarity of the score profiles suggests that their development is not deviant. All scores are delayed in a similar way. The delay in the development of ToM in children with PDD-NOS was mainly obvious in the preschool phase (between 4.5 and 6 years of age); their ToM scores were 1.5 standard deviations lower than in TD children.

The scores of the PDD-NOS group ameliorated as the research period prolonged. So also research question three “Do children with PDD-NOS evidence a developmental progress in ToM functioning?” can be answered affirmatively. It appeared that the developmental increase in ToM was not only due to aging, but that there was also an additional effect due to learning caused by the repeated measurements, this notwithstanding the precaution taken to use parallel versions of the ToM Storybooks every four months. The children with PDD-NOS evidenced a statistically significant classical learning curve on top of the predicted age-related increase. This is in accordance with dynamic testing theory, which states that repeated measurements may lead to learning effects beyond trivial habituation effects (Grigorenko & Sternberg, 1998). In intelligence tests, for instance, quotient scores can be influenced by practice effects (Kaufman & Kaufman, 1983; Wechsler, 1974; Tellegen et al., 2003), with even a rise of 10 points for Performance IQ. We found an increase of 14 ToM-Q points, i.e. about one standard deviation, over a period of 16 months, in children being tested every four months. Learning effects have also been demonstrated in ToM research (Muris et al., 1999) and with the ToM Storybooks (Blijd-Hoogewys et al., 2008). A single repeated measurement after two weeks sufficed to induce a significant score increase in TD children of three to seven years old. One explanation for this learning effect over time in the current research is that having heard and being questioned about mental states so often, has lead to practice effects in the ToM of these children. This effect has been hypothesized earlier: reading storybooks to children might benefit their development of mental states (Dyer et al., 2000). The current research affirms this finding in children with PDD-NOS.

We also tested whether there were age trends in the developmental progress of ToM (research question four “Is the nature of the ToM developmental progress subject to age changes?”). Older children with

PDD-NOS had higher starting scores, which reduces their potential developmental range, due to ceiling effects. This was obvious in both ToM total scores and ToM sub-scores. However, the relationship between the slopes of the growth curves and average age was nonlinear, with the negative relationship between slope and age leveling off at around 80 months, which occurs just before the anomaly in ToM total scores sets in.

In addition to observing development in the ToM functioning of children with PDD-NOS, we also found a development in their socialization skills, as measured with the VABS interview (research question five “Is there a meaningful relationship between their ToM scores and behavioral scores?”). At the end of the research period they displayed a significant increase in their score on Play/Leisure and on Daily Living Skills. However, the growth in both VABS sub-domains did not correlate with the slope of ToM total scores and only the correlation of ToM with Daily Living Skills could be judged as sufficient. This supports the hypothesis that children with PDD-NOS show social cognitive progress, but less corresponding behavioral progress. At the one side, this could be attributed to the hypothesis that their ToM is more cognitively mediated rather than intuitively reflected. On the other side, additional information processing problems, as known in autism, may also play a substantial role in catching up with their ToM delay, such as executive function problems, a weak central coherence and generalization problems.

In the introduction of this chapter, a two-dimensional ToM continuum was presented resulting in four quadrants of characteristic ToM functioning. It was assumed that this continuum correlates with severity of autism diagnosis. It was hypothesized that children with PDD-NOS belong to one quadrant evidencing mainly quantitative problems, not qualitative problems in their ToM. This hypothesis was verified, providing evidence of criterion validity of the ToM storybooks: the ToM Storybooks are able to discriminate between children from quadrant A and quadrant D. From this study, no conclusions can be drawn about the other quadrants or other pervasive developmental disorders, like children with AD and AS. Longitudinal research, based upon repeated measurements, involving such groups is desirable to further test the hypothesized ToM quadrants. We expect that children with AD and AS show some spontaneous ToM development. After all, a subgroup of adolescents and adults with such a diagnosis are known to succeed on basic ToM aspects (Kleinman et al., 2001). However, we expect more qualitative problems in their development. It has been suggested that individuals with HFA and AS who succeed on

ToM tasks do so by using cognitively mediated routes, also referred to as the hacking hypothesis (see Dissanayake & Macintosh, 2003). Such a ToM is not considered an intuitive skill to use mental states (Abell et al., 2000), but more a compensation strategy that is less appropriate in complex and rapidly evolving social situations (Frith & Happé, 1999). One can imagine that the development of such a ToM shows a deviant profile in comparison with that in TD, although a deviant mechanism should not automatically be expected to lead to deviant ToM scores. It could as well lead to a ToM delay.

Going back to the two-dimensional ToM continuum, could it be that children in quadrant B and C (AD and AS respectively) merely use cognition in order to understand mental states of others, and that individuals in quadrant C (AS and possibly also HFA, an AD of normal intelligence) master this cognitive skill more profoundly? Is it possible that children in quadrant A and D (PDD-NOS and TD respectively) use a more intuitive form of ToM? But is that sufficient for children with PDD-NOS? We have illustrated that the ToM development in PDD-NOS is delayed. One can imagine that, due to the delay, ToM is not sufficiently developed before the window of opportunity closes, before it can become a fully automatic, encapsulated ability to mind-read (e.g. Papp, 2006). In that case, children with PDD-NOS will never catch up with their TD peers, and will also to a certain extent, have to rely on cognitively mediated ToM-strategies, a more consciously ToM. If this is so, the distinction between delayed and deviant development of ToM is likely to become blurred, and development that was once delayed may become deviant. Note that also TD children are suspected to use a more cognitive and reflective way of ToM when they age. The dip found in TD children might reflect the emergence of a second strategy (see Blijd-Hoogewys et al., submitted b), perhaps indicating the closing of the window of opportunity, after which development of ToM becomes considerably slower or is less easily influenced by practice. If this is so, it would be highly advisable to stimulate children with PDD-NOS in their ToM development before this age, for instance by mere testing them repeatedly with a ToM test. For, this study showed that repeated testing with the ToM Storybooks stimulated ToM development. If application in daily live is aimed at, training of context-specific applications of ToM seems more advisable, like with the training of Steerneman and colleagues or with specialized social skill training programs.

The spontaneous ToM development found in children with PDD-NOS might shed a different light on ToM training studies. Such studies do

not habitually incorporate natural development. For instance, a recent ToM training study (Gevers, et al., 2006) reported an average increase of 9.90 total ToM points, which is an increase of 1.49 standard deviation (on the ToM test, a similar test, with a maximum total score of 72; Muris et al., 1999), after a training period of seven months. The authors reported a rise in first-order beliefs, but no progress in the recognition of emotions, distinction of physical-mental items and false beliefs. Furthermore, significant progress was found in all Socialization sub-domains of the VABS. Our study followed a group over a period three times as long (20 months) without ToM training and resulted in an average increase of 15.63 total ToM points, which is an increase of 1.11 standard deviation (on the ToM Storybooks, with a maximum total score of 110), only slightly lower than the one reported in the former mentioned ToM training study. We found a similar rise in ToM scores.

To summarize, this article found that children with PDD-NOS evidence ToM problems, that their ToM development is delayed, that their growth trajectory is nonlinear (there is an anomaly at 85-90 months), and that this nonlinearity is delayed in comparison to their TD peers (evidencing an anomaly at 72-78 months). The ToM problems of children with PDD-NOS were quantitative and not qualitative of nature. These results are in accordance with what we predicted based upon the two-dimensional ToM continuum.

Further research is required to validate or invalidate the hypothesized ToM quadrants, in particular longitudinal research in children with AD and AS. Concerning such research, it is recommended to add more complex ToM tasks (like second-order belief tasks) in order to tap a broader developmental range in ToM abilities, and to explore the influence of other factors, like for instance language ability and executive functions. Further exploration of the quantitative, qualitative and nonlinear aspects of ToM development in TD children as well as children with different types of ToM problems may be expected to provide further insight into the complexities of disorders affecting social interaction and communication.