

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).

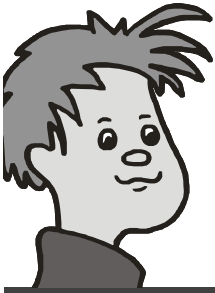
The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

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 4



Discontinuous paths in the development of Theory-of-Mind. A nonlinear dynamic growth modeling approach.

Abstract: Research on Theory of Mind (ToM) has mainly focused on ages of core ToM development. This article follows a quantitative approach focusing on the level of ToM understanding on a measurement scale (the ToM Storybooks) in 324 typically developing children between 3 and 12 years of age. It deals with the eventual occurrence of developmental discontinuities in ToM functioning, using non-linear smoothing techniques, dynamic growth model building and additional indicators (moving skewness, growth rate changes and variability). The ToM sum-scores showed an overall developmental trend that leveled off towards the age of ten years. Within this overall trend two discontinuities were found, one at the age of around 56 months and another at the age of 72-78 months. These temporal decreases in ToM sum-score were accompanied by a decrease in growth rate and variability, and a change in skewness of the ToM data, all suggesting a developmental shift in ToM understanding. The temporal decreases also occurred in the different ToM sub-scores and most clearly so in the core ToM component of beliefs. It was also found that girls had an earlier growth spurt than boys and that the underlying developmental path was more salient in girls than in boys. The consequences of these findings are discussed from various theoretical points of view, with an emphasis on a dynamic systems interpretation of the underlying developmental paths.

This chapter is submitted as: Blijd-Hoogewys, E.M.A., Van Geert, P.L.C., Serra, M., & Minderaa, R.B. (submitted). Discontinuous paths in the development of Theory-of-Mind. A nonlinear dynamic growth modelling approach.

INTRODUCTION

Theory-of-Mind has become one of the most intensively studied subjects in the field of social cognitive development. It is considered an important condition for understanding the social environment and for showing socially adequate behavior (Astington & Jenkins, 1995). Theory-of-Mind (ToM) refers to the ability to attribute mental states -such as beliefs, desires and emotions- to oneself and others and to use these mental states in understanding, predicting and explaining the behavior of oneself and others (Mitchell, 1997; Premack & Woodruff, 1978). For instance, a child comprehends that if Sam wants an ice-cream (a desire) and his mother does not allow him one, Sam will be unhappy (the consecutive emotion). Or, a child comprehends that if Sam thinks his ball is in the garage (a belief) he will look in the garage for his ball (the consecutive action) even though the ball may in reality be in the garden. A four-year-old who is questioned about the actions of Sam and who also knows the true location of the ball will be able to predict the action of Sam correctly. A three-year-old, however, will not be able to do so: he will most likely say that Sam will look in the garden. The three year old cannot distance himself from the knowledge of the true location, and he does not comprehend that others can hold beliefs that do not match reality as he sees it. Four-year-olds, on the other hand, understand false beliefs (for a meta-analysis on false beliefs see Wellman et al., 2001). This difference between three- and four-year-olds has been studied intensively.

Qualitative versus quantitative approaches of development

Literature has elaborated on ToM development mainly in a qualitative manner. By 'qualitative' we mean an approach that focuses on some particular developmental phenomenon, such as understanding false beliefs, and searches for the age at which this phenomenon is established in development. Research has shown that ToM develops at a specific rate and according to a particular sequence. It evolves from a simple desire theory to a complete belief-desire theory, from true beliefs to false beliefs, and from the understanding of first-order beliefs to second-order beliefs (Wellman, 1990). Deviations from this normal developmental path have been used in describing the ToM difficulties of, for instance, children with autism. Research has shown that in order for children with autism to be able to succeed on false belief tasks they need to have a verbal mental age of about

eleven years old (e.g. Happé, 1995). This is a mental age six to seven years older than for normally developing children.

In addition to a qualitative approach of ToM, a more quantitative approach can be chosen. By 'quantitative' we mean an approach that defines a quantitative dimension, for instance the level of false belief understanding on a measurement scale, and then studies the changing level of that dimension over the course of developmental time. Wellman and Liu (2004), for instance, looked in particular into the conceptual changes of different ToM aspects. They argue that the ToM developmental order is not one of addition or substitution, but one of modification or mediation (Wellman & Liu, 2004, pp. 536). Initial insights broaden or generalize into later insights.

One can question how these generalizations come about. Is there a gradual development or are there temporary accelerations, delays or even regressions observable during ToM development? Temporary regressions are well-known characteristics in developmental psychology. After mastering an ability, children can have a temporary relapse before the ability consolidates. This phenomenon is often referred to as U-shaped development (because a U-shaped curve is visible when this kind of development is depicted in a graph), or eventually N-shaped development. Such temporary regressions have been found in a variety of domains, including language development, social cognition, creativity, reasoning, auditory localization and face perception (for an early collection of studies, see Strauss & Stavy, 1982). Recent findings on U-shaped curves include motor and verbal development (Gershkoff-Stowe & Thelen, 2004), non-verbal symbol learning (Namy et al., 2004), and face perception (Cashon & Cohen, 2004).

In addition to temporary regressions, developmental curves may also show accelerations, which are often the hallmark of rapid changes, i.e. developmental transitions (see for instance Fischer and Bidell, 2006). The qualitative developments mentioned earlier, for instance the acquisition of false belief understanding, are likely to be examples of such transitions. Theorists focusing on the dynamics of developmental growth have argued that developmental transitions are likely to be preceded by temporary regressions (Fischer & Bidell, 2006; Van Geert, 1991).

Up to now, ToM research has primarily focused on ages of transition, whereas the long-term forms of conceptual changes in ToM understanding have received less attention than deserved (Wellman et al., 2001). Focusing on such changes may shed an alternative light on ToM development and could make identification of children with ToM problems

easier and more accurate. In this article, we hope to illustrate the additional value of a quantitative approach to ToM development in typically developing children.

In order to be able to observe the more subtle quantitative aspects of ToM-development, such as eventually occurring regressions, temporary accelerations and decelerations in the rate of change, one should take two issues into account. First, ToM development does not solely depend on the development of (false) belief understanding, but also involves other ToM components, like for instance desires (Astington, 2001). As a consequence, research focusing on quantitative ToM development should involve a series of tasks incorporating different core ToM components but also relevant ToM precursors and associated abilities (like for instance being able to make the distinction between mental and physical states). For that purpose, we developed the ToM Storybooks; this instrument contains a variety of tasks on different ToM and ToM-related components (for more details, see section 'method' in this article). Second, it is advisable to test children of a wide age range. Instead of confining itself to testing children from the age of three to six, research should aim at a considerably broader age range, for instance up to ten or even twelve years old. The justification for doing so is that there is no convincing evidence that by the age of six ToM is fully acquired (e.g. Hala & Carpendale, 1997) and stable.

Cross-sectional versus longitudinal: tapping 'first-encounter reaction'

In order to study the quantitative changes in ToM development, one can choose for a cross-sectional or a longitudinal design. In a cross-sectional design, children are tested only once and children of different ages are compared with one another. In a longitudinal, or more preferably a time-serial design, children are repeatedly tested over a longer period and can be compared with themselves.

At first glance, a time-serial design would be superior for the purpose set in this article. This is a design with as many measurements as are needed to capture the temporary and often non-linear forms of change characteristic of a particular developmental phenomenon (van Geert & Steenbeek, 2005; Steenbeek & van Geert, 2002). However, such a method also brings along considerable logistic problems. Children need to be tested repeatedly over an extended period, for instance of eight years (between three to eleven years of age). Since so few research has focused on the dynamics in ToM development, it is hard to predict at what intervals

children should be tested in order to find evidence of developmental phenomena such as accelerations and decelerations, transitions and temporary regressions. A cross-sectional design provides an answer to the question of age-related changes and potential critical points in ToM and is a first step towards future time-serial research.

Next to these logistic problems, there is also a theoretical consideration to prefer a cross-sectional design in this particular case (which is absolutely not to say that we recommend this in every case). Cross-sectional research almost by definition taps a 'first-encounter'-effect. That is, the children are not familiar with the particular test or task and thus form a hypothesis as to what is expected from them, what the test entails, etcetera. From a dynamic testing point of view (Grigorenko & Sternberg, 1998), this first encounter may reveal information that later, repeated test administrations no longer contain, because they are affected by habituation or learning (which is absolutely not to say that such repeated measurements are not or less valid, but they do provide different perspectives on the measured variable, eventually).

Parametric versus non-parametric models

In order to describe quantitative changes in development, different fitting models can be used to represent the general underlying trend. In research, linear or quadratic models are often used. Unfortunately, such models do not sufficiently take local deviations of the distribution of data into account. This may lead to over- and underestimations of the expected average scores in certain age periods, and thus conceal local but real accelerations, decelerations or regressions, if any occur.

In contrast, non-parametric models, like the Loess or Lowess estimate smoothing procedure, follow local distributions of data as reliably as possible. They apply a locally weighted least squares estimate, and are commonly used as smoothing techniques (see for instance Härdle, 1991; Simonoff, 1996). Such non-linear techniques can be of substantial value for testing dynamic changes even when applied to cross-sectional data.

Indicators of developmental transitions

As noted earlier, non-linear phenomena such as accelerations, decelerations and temporary regressions might be the hallmark of underlying

developmental transitions, i.e. transitions from one pattern of understanding to another one. However, the occurrence of such phenomena alone is not sufficient to demonstrate the existence of transitions, and additional statistical indicators are thus required.

In the literature on discontinuous developmental changes, which have been studied in the context of catastrophe (or bifurcation) theory, statistical indicators known as ‘catastrophe flags’ have been used (see for instance Gillmore, 1981; Van der Maas & Molenaar, 1992; Van Dijk & van Geert, 2007; Hartelman et al., 1998; Hosenfeld et al., 1997; van der Maas et al., 2003; Wimmers et al., 1998). Three of these indicators can also be used to demonstrate transitions that are more gradual than the bifurcations referred to in the catastrophe-theory literature. These indicators are changes in the skewness of the distribution, temporal changes in growth rate and change in variability (Van Geert & Van Dijk, 2002; Bassano & van Geert, 2007; Van Dijk & van Geert, 2007). They can be used with time-serial longitudinal data, but also with cross-sectional data, if a first impression of the ages at which eventual transition phenomena occur needs to be obtained.

Changes in the skewness of the distribution over time can be applied to ToM development in the following way. We expect that during the early state, when no ToM growth has taken place yet (at least as measured by the ToM Storybooks), there are random variations around an initial state mean, and this variation is likely to be symmetrical. That is, skewness of the distribution should be near zero. As ToM growth begins to take off, skewness will increase. First a positive skewness will appear, since the majority of subjects are still characterized by a relatively low level of ToM, with a number of more advanced subjects showing a high growth rate and thus considerably higher scores on the ToM Storybooks. As the majority of subjects catches up while the more advanced subjects reach the (eventually temporal) upper level of ToM, the skewness of the distribution first moves towards zero and then eventually to negative values. The resulting negative skewness means that the majority of subjects are now close to the asymptotic developmental level, with some subjects still lagging behind (and thus accounting for the tail to the left). Finally, as all subjects tend to reach the asymptotic developmental value, the skewness moves again towards zero, representing a symmetrical distribution of values around some final-state average (for more explanation, see van Geert & van Dijk, 2002).

A temporal change in growth rate is another indicator of developmental transition. It can be demonstrated in the form of marked oscillations in the first derivative of the developmental curve. The first

derivative at any point of the growth curve represents the rate of growth at that point. A particularly strong instance of change in the growth rate occurs in the form of a temporal regression, where the growth rate temporarily drops down to negative values. Since the changes in skewness over time are related to accelerations in the growth of the developmental phenomenon at issue, we expect to find a certain level of coherence between the first derivative of the non-linear ToM growth curve and the change of skewness over time.

Change in variability, the third indicator of developmental transition discussed in this article, can be observed as intra- and inter-individual variability. A temporal increase in the *intra*-individual variability is considered a strong indicator of a developmental transition (van Geert & van Dijk, 2002). However, such an indicator can only be used in repeated measures designs. *Inter*-individual variability, which is applicable to cross-sectional data and which is expressed in terms of standard deviation over a certain period of time, might also temporally increase during a transition. This increase may be due to a side-effect of changing skewness of the distribution. Changes in the pattern of variability among individuals are known as anomalous variance and are indicators of a coming transition (see Van der Maas & Molenaar 1992).

Research questions

The aim of our study is to describe the form of the curve or pattern of ToM development over a considerable time period, namely from 3 to 12 years, measured by means of a test that comprises a wide variety of ToM components and tasks. We have chosen to focus on non-linear aspects of the developmental curve, including local accelerations, decelerations and regressions.

We expect to find discontinuities in the developmental pattern of ToM. Our research questions focus on the veracity of these discontinuities. Are the discontinuities real or due to artifacts? Are the discontinuities supported by additional indicators of discontinuity? Are the discontinuities not only observable in the ToM total score, but also in the ToM sub-scores and in both boys and girls?

METHOD

The ToM Storybooks

Children's ToM knowledge was tested with the ToM Storybooks, version Sam (a revision of the test used in Serra et al., 2002; see also Blijd-Hoogewys et al., 2008; Blijd-Hoogewys et al., submitted a). The ToM Storybooks is a test that measures a variety of ToM components and associated aspects, based on the work of Wellman (1990).

The ToM tasks are incorporated in short stories. The stories are illustrated with full color pictures and enlivened by the use of caressable patches of fur, toy doors that can be opened, and magnetized emotion faces that can be placed on the characters. The test takes 40 to 50 minutes, including a short break.

In total, there are 34 tasks spread over six storybooks in total. A maximum sum-score of 110 points can be obtained, which can be divided into five sub-scores: 1) emotion recognition (maximum=14 points), 2) distinction between physical and mental entities (real-mental, real-imaginary and close impostors; maximum=44 points), 3) understanding that seeing leads to knowing (maximum=3 points), 4) understanding of desires (maximum=17 points), and 5) understanding of beliefs (maximum=32 points). The latter encompass tasks on standard belief, changed belief, not own belief, explicit false belief, false belief, inferred belief and inferred belief control.

Each task incorporates one to five questions, including both test questions and justification questions (for an overview of questions per task, see Appendix A). There are in total 74 binary test questions and 18 justification questions. The answers to the test questions are coded as correct or incorrect (1 or 0 points; maximum of all test questions=74). The justification questions result in 2, 1 or 0 points, depending on the amount and correctness of the mental state terms spontaneously used by a child (maximum of all justification questions=36). In order to evaluate the justifications, a category system is used. This system is based on the category system used by Rieffe (1998), on various categories from Wellman (1990) and on an exploration of the empirical data. For each justification question, correct answer categories were determined (for a short review of correct answers, see Appendix A, column 'answer').

The ToM Storybooks have good psychometric qualities. The internal consistency, test-retest reliability, inter-rater reliability, divergent and convergent validity are good (see Blijd-Hoogewys et al., 2008).

Subjects and setting

We tested 324 children. The ages ranged from three up to and including eleven years, with approximately the same number of boys and girls per age range (see Table 1 for the age distribution).

The children came from preschools, kindergartens and elementary schools, from both provincial and urban regions in the Netherlands. All children had a Dutch linguistic background, and did not have language acquisition problems that could have hampered their performance on the tasks (for the effect of language on ToM performance see for instance Garfield et al., 2001; Lohmann & Tomasello, 2003). Thirteen percent of the children came from a lower social background, distributed over the whole age range. This percentage corresponds with the percentage as known from the Dutch National Bureau of Statistics.

Table 1: *Age distribution of sample being administered the ToM Storybooks*

	Age (in years)							Total
	3	4	5	6	7	8-9	10-11	
Boys	32	31	31	31	15	14	13	167
Girls	29	24	32	26	16	12	18	157
All	61	55	63	57	31	26	31	324

Procedure

All subjects were individually tested in a quiet room at school. Test administrators were carefully instructed to follow standard procedures. For practical reasons, kindergarten children were tested at home. If necessary, the parent was allowed to be present during testing but was requested not to interfere.

Statistical method

In order to acquire insight in non-linear changes in ToM development, we used a descriptive non-parametric method, namely Loess curve smoothing. Such a model is superior to a monotonic regression model, even over a low-order polynomial model which is the standard in psychology, in revealing temporary regressions (see Blijd-Hoogewys et al., submitted a). Where used, the Loess smoothing was based on a window size of 30%. The Loess procedure evaluates consecutive windows of data (in this case 30% of the

data; thus 0-30%, 1-31%, 2-32%, and so on), each time calculating the average point of the regression and finally resulting in a smooth curve of these points (for an explanation of the Loess smoothing procedure, see Blijd-Hoogewys et al., submitted a).

Next to that, we used random permutation techniques, and more generally, Monte-Carlo analyses, which are assumption-free techniques. Wellman and colleagues (2001) have argued for the use of more assumption-free techniques, such as bootstrap methods, in ToM research. It entails a simulation of the test statistic at issue as based on the null hypothesis, which can be compared to our empirical ToM data (Good, 2001; Manly, 1997; Todman & Dugard, 2001). The test statistic is the measure used to specify a distinction or property of interest, for instance the difference between average scores at different ages. In order to test the statistical significance of non-linear properties of the developmental curve, we used interval differences of moving averages (a moving average is the average of scores for a particular period, for instance two months, which is then moved across the data; an interval difference is the difference between a moving average at time t and the moving average at a later time, $t+i$).

RESULTS

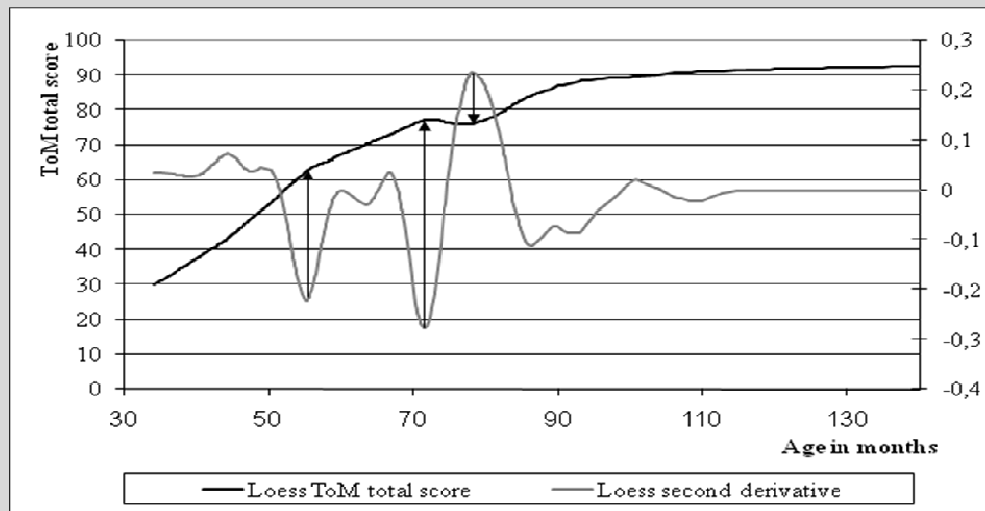
Non-linear phenomena in the curve of ToM scores

In order to reveal non-linear properties, we used Loess curve smoothing. The resulting smoothed curves of the ToM sum-scores of children (maximum=110 points) reveal that there are three points of developmental interest (see Figure 1). In order to determine the exact timing of these points, we inspected the minima and maxima of the second derivative of the developmental curve at issue. The second derivative is the acceleration of the growth and is a better technique to pinpoint the exact timing of these points than the first derivative (remember that the first derivative is the rate of growth). The three points corresponded with the most marked inflection points in the developmental curve.

The first is a bend in the curve at around 56 months (4 years and 8 months). The second is a bend at around 72 months (6 years), followed by a dip or temporal regression of the curve, which shows its deepest point at 78 months (6 years and 6 months). This dip (which one might alternatively call a temporary regression or a local U-shaped development) is the most striking

deviation from monotonicity in the non-linear developmental curve based on the data from boys and girls taken together. More detailed analyses can be found in the sections on gender differences and differences between ToM components.

Figure 1. The Loess fitting curve of the ToM sum-score data plotted versus age displays a discontinuity. Based upon the second derivative of this curve three points of developmental interests were found, namely at 56, 72 and 78 months.



In the remainder of this article, we refer to these non-monotonic changes by means of the more common term ‘discontinuity’. We use this term in its most general meaning, thus referring to a change in a rate or direction of change that is expected to rise monotonically.

Are these discontinuities real?

Since development is tacitly assumed to imply continuous improvement, temporary regressions or discontinuities are seldom welcomed enthusiastically (Siegler, 2004). Nevertheless, before taking the data as valid evidence for the existence of such discontinuities, it should be checked whether the discontinuities are not simply the result of inadequate selection procedures or of statistical anomalies, namely accidental sampling effects or the influence of specific, biased or incompetent testers. In this section, we

confine ourselves to the most pronounced discontinuity in the developmental curve, which is the dip at around 72-78 months.

The first objection against the reality of this observed discontinuity is that we have made a serious error in the selection of children around the ages at which the temporary regressions have been found. However this is highly unlikely, the selection procedure has been carried out with the utmost care and selection criteria were uniform over all ages.

The second objection is that the discontinuities in general and the 78-month dip in particular are statistical artifacts. Consequently, the null hypothesis to be tested is that the generic curve underlying the data is actually a monotonically rising curve and that the dip is due to accidental sampling variations. Accordingly, it was hypothesized that sampling accidents amounted to statistical variations in the representativeness of the local samples (e.g. the sample of subjects around the age of 75 months), resulting in an accidental over-representation of low-scoring individuals. In order to test this null hypothesis, we first specified both duration and temporal position of the empirically found discontinuity (the 72-78 month dip). The best fitting monotonic growth curve (i.e. a curve without regressions) was determined and a regression model for the variances was fitted (the best fitting curve was of the form $y = a + b/x^{1.5}$). Since we had no prior assumption about where a discontinuity in ToM ability should occur, we tested for the accidental occurrence of apparent discontinuities anywhere along the time interval between the lowest and highest ages from our set of data. Because a theoretical expectation about the length of the discontinuity is also lacking, the null hypothesis should also be tested for different time windows (it is understood that such durations must be within the limits of a few months, longer durations would hardly be interpreted as ‘temporary dips’). By means of a Monte Carlo technique - simulating distributions of scores based on the null hypothesis - we calculated the probability that the null hypothesis model yields a discontinuity, comparable to the observed discontinuity, somewhere along the observed age interval. This method is based on bootstrapping samples and is assumption-free (compare Wellman et al., 2001). Moreover, it accounts explicitly for the unbalanced distribution of ages (e.g. fewer older children in the sample). The discontinuity was defined as the maximal difference between moving averages of the sum-scores simulated on the basis of the null hypothesis model, and was calculated for a series of predefined intervals (e.g. from one to three months).

First, we tested for all possible locations and for different durations of discontinuities across the entire age span. The pattern of the p-values supported the conclusion that it is unlikely that the observed discontinuity is an accidental sampling effect of an otherwise continuous, monotonically rising simple curve (Monte Carlo, $p=.01$ through $p=.05$, depending on the length of the tested interval). Thus, it is highly likely that the observed discontinuity is caused by a real discontinuity in the data. When we took another indicator for discontinuity, namely negative slope (also calculated over an interval of variable length), we found converging evidence (Monte Carlo, $p=.02$; the p-value refers to an average duration for the discontinuity, as determined on the basis of the first significance test).

Next, we checked if the data were experimenter-dependent, implying that the discontinuity is in fact caused by anomalous test administrations by a particular test assistant who, for some accidental reason, is associated with the age range at which the dip is found. We defined eight groups of data sets by leaving out the data of one particular tester or a group of similar testers at a time. If the temporary discontinuity is due to an anomalous tester, it should disappear in the dataset from which this particular test assistant is lacking. We repeated the statistical procedure described above for each of the reduced data sets. The resulting p-values showed that the discontinuity remained significant for each of the reduced data sets and was thus not due to an anomalously operating experimenter (Monte Carlo, $<.001$ through $p=.05$).

In summary, neither selection errors, nor accidental sampling errors nor a deficient experimenter can account for the occurrence of the observed discontinuity.

Additional indicators for discontinuity

In order to further corroborate the validity of the observed discontinuities, we can, first, check whether the change in the average sum-scores is backed up by changes in additional statistical indicators, namely changes in skewness, growth rate and variability. Second, we can check whether the observed discontinuities in the sum-scores are due to a particular sub-group, more precisely whether they are due to subjects with either low, average or high sum-scores, or to either boys or girls. Thirdly, we wish to know whether the discontinuities occur in all components of ToM, i.e. in all sub-scores, or whether they are due to particular components (e.g. due to

additional components instead of core ToM components; e.g. due to the sub-score ‘distinction between physical and mental entities’).

What we expect to find is that the discontinuities in the average sum-scores over time are associated with changes in skewness, growth rate and variability that are indicative of an underlying transition, that the discontinuities occur irrespective of the score level (low-average-high) and the subject’s gender, and finally, that they occur in all ToM sub-scores, especially in the core ToM sub-scores like the understanding of desires and beliefs. We begin with a section discussing the association with additional indicators.

Skewness and variability as indicators of an underlying transition

In the introduction, we discussed three qualitative indicators of developmental transition, namely skewness, temporal changes in growth rate and change in variability. Before further analyzing the developmental ToM pattern, we first wished to determine whether the hypothesized properties stated are indeed characteristic of a developmental transition of the kind we now expect to find in the ToM-data. In order to do so, we mathematically simulated a transition model in order to check whether the expected qualitative indicators occur. A good example of a developmental transition is a two-step growth process (for details on how such models can be specified and simulated, see Van Geert, 1991, 1994; a two-step growth process can easily be extended towards a three- and more-step model if needed).

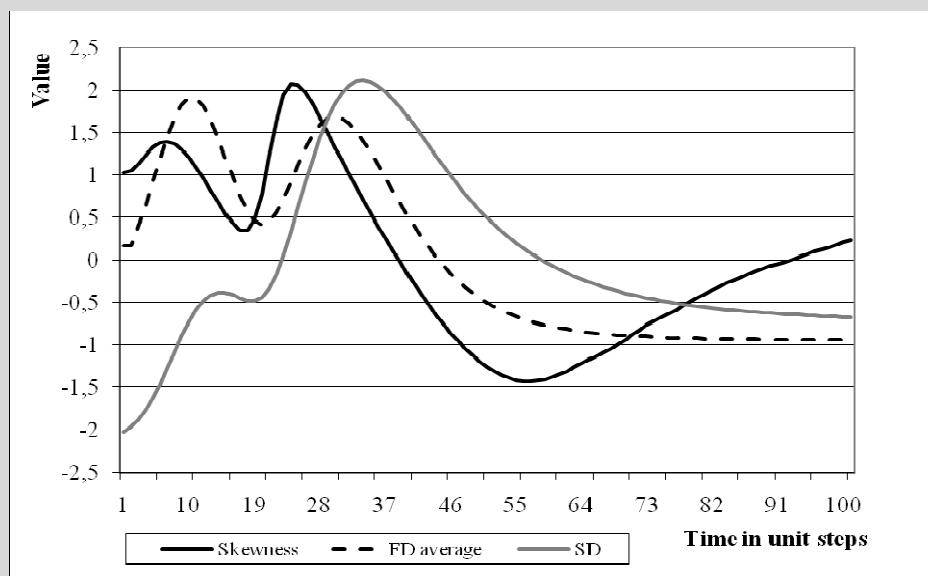
A simulation sample consisting of 600 individual two-step growth curves with randomized initial level, growth rate and final state level was simulated over 100 time steps (meaning that the age range between 3 and 12 was divided into 100 time points). The skewness, growth rate (first derivative) and variability (standard deviation) for each simulated time point of the 600 cases was calculated. Then, time trajectories of skewness, growth rate and variability over the 100 simulation groups could be inspected and compared, with the aim of finding characteristic qualitative properties.

Note that by taking skewness, growth rate and variability measures from each simulated time point (we have 100 time points), we in fact discarded all available information about the individual (longitudinal) trajectories. In fact, by taking a random sample at each point in time from all the simulated cases, we simulated a purely cross-sectional measurement, formally similar to the cross-sectional data collected in our ToM study. The qualitative properties of the time-series of skewness, growth rates and

variability were conserved in these purely cross-sectional samples, thus showing that time-serial longitudinal datasets are not necessarily needed in order to discover ‘flags’ or markers indicating underlying growth patterns.

Figure 2 shows the pattern of changes in the skewness, first derivatives and standard deviation based on a simulation of a two-step growth process. It shows a series of two peaks in the skewness, the first derivative and the standard deviation. The skewness and first derivative peaks largely coincide (covariance of the series is on average 0.7). The skewness peaks precede the growth rate peaks (first derivative) and the latter precede the variability peaks (standard deviation).

Figure 2. Loess curves of the three qualitative indicators of developmental transition based on simulated data. The skewness peaks precede the growth rate peaks and the latter precede the variability peaks.



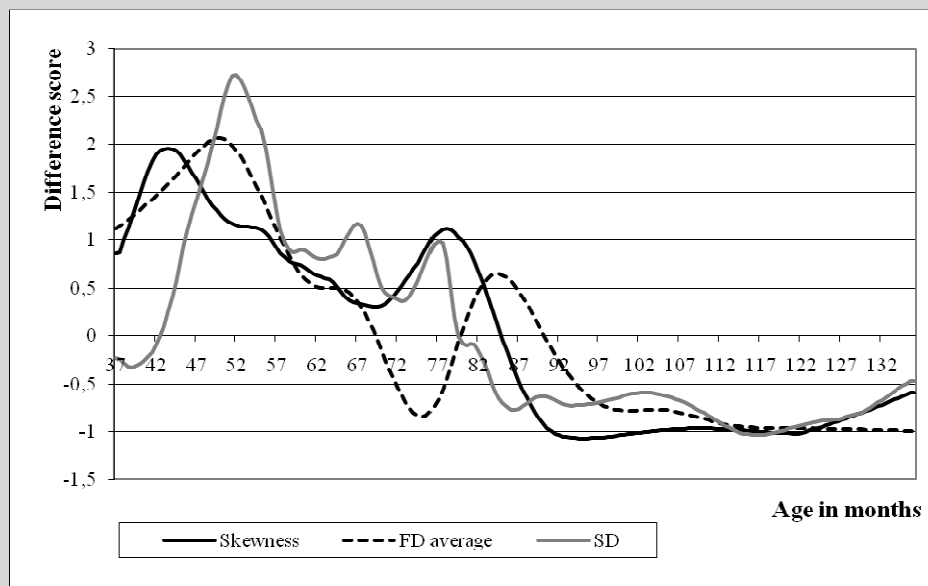
Legend: FD= first derivative (or growth rate), SD= standard deviation (or variability).

In contrast with the simulation data, the empirical data provide only one case for each time point (i.e. each age of a participant). This implies that skewness must be calculated over a moving time window, comprising a sufficient number of cases (we have chosen 21 successive data points, this is averagely a period of 6.34 months). Successive data points are likely to show the progress (or eventually regression) corresponding with their time window (e.g. the period from 34 to 40 months that spans the first time

window). Skewness must differ from the mere change in the variable (if it does not, it adds nothing to our analysis), and thus be defined independent of the change. A good way of doing so is to calculate skewness over the residual values, i.e. the sum-scores minus the corresponding value of the smoothed, non-linear growth curve. Since the nonlinear smoothing procedure adapts itself optimally to the local averages, the residuals should be expected to be symmetrically divided around zero, except if the skewness varies as a function of developmental time, which is what we expected to find.

After calculating the skewness of the residuals, the skewness data were smoothed by means of a Loess smoothing method, with a window size of 30%. First derivatives (i.e. local growth rates) of the developmental curve were then calculated and also smoothed. Lastly, the variability (moving standard deviation) was calculated. The resulting curves are represented in Figure 3 (we used standardized values, so the three criteria can be presented using one scale).

Figure 3. Loess curves of the three qualitative indicators of developmental transition based on empirical data. A mixture between a two-step and a three-step growth process is obvious. There are two peaks, both in the skewness and in the first derivative curve.



Legend: FD= first derivative (growth rate), SD= standard deviation (variability).

Figure 3 shows a mixture between a two-step and a three-step growth process. There are two large peaks, with a smaller peak in between, most clearly observable in the variability measure and less in the other two measurements. The qualitative similarity with the model simulation of a two-step process is striking. There are two peaks, both in the skewness and in the first derivative (i.e. growth rate) curve. As is the case in the simulation, the peaks of skewness largely coincides with those in the first derivative (growth rate), and the skewness peaks come somewhat earlier than those of the first derivative. Note that the model simulation was not meant to model the current data set, but to show which type of co-varying pattern in skewness and growth rate should be expected in a two-step developmental process. The covariance of the series, after standardizing, is 0.88, which is comparable to (and even higher than) the high covariance that the simulation model predicted.

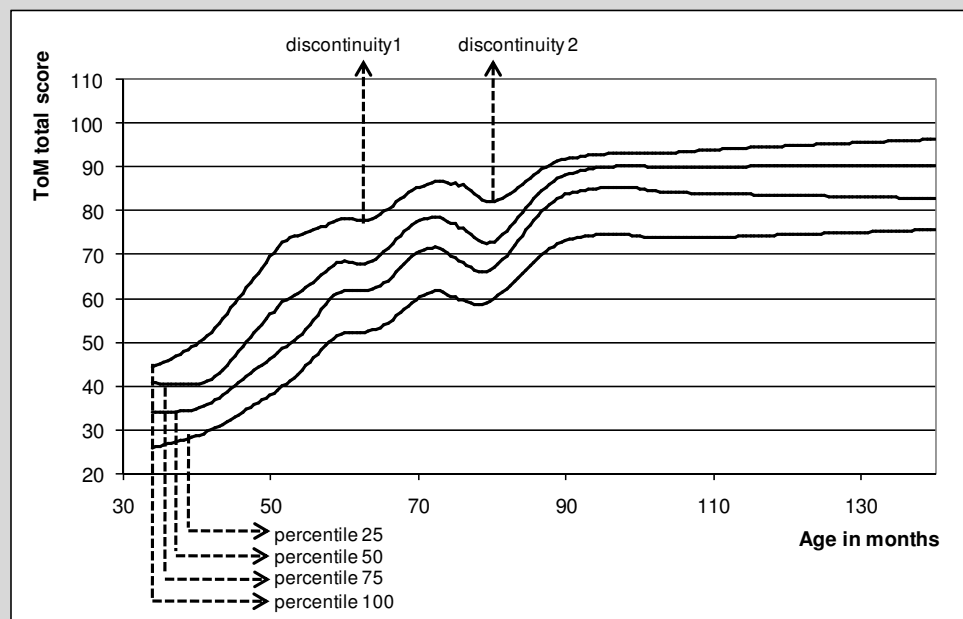
Before concluding that the skewness and first derivative data support the notion of a two-step developmental process, we need to know what the probability is that a similar covariation of skewness and first derivative curves can be obtained if the underlying statistical variation of the sum-scores is in fact symmetrical across age (and not varying systematically, as hypothesized). That is, what is the probability that a null hypothesis model produces the same or higher covariance? According to the null hypothesis model, the distribution of the sum-scores around the predicted value is symmetrical. The predicted value at any age is the value of the non-linear developmental curve at the age point in question. The variability of the sum-scores around each expected point is given by the standard deviation of the observed residuals. The null hypothesis model can be tested by generating random series of sum-scores based on a normal distribution model, with means equal to the successive values of the nonlinear growth curve and standard deviations equal to the observed standard deviation of the residuals. Two hundred such series were generated, Loess smoothed with a 30% window and then standardized. Covariances were calculated between each smoothed series and the first derivative time series. Only 2 out of the 200 series had a covariance greater than or equal to the observed covariance. This means that the p-value of the observed covariance, i.e. the probability that the observed covariance can be explained by the null hypothesis model is approximately $2/200$, i.e. 0.01 ('approximately', because the number of null hypothesis simulations – 200 – is rather limited). We can thus conclude that the skewness data provide further independent evidence for the existence of at least a two-step process in the development of ToM.

Discontinuities in sub-groups and sub-scores

After having checked additional indicators of a transition, we now proceed with the question of whether the observed discontinuities in the sum-scores are due to a particular sub-group (ToM level or gender).

Firstly, we divided the children in four groups, based upon local percentile borders, which divides the group of subjects into four score levels, from low to high. Per 33 children (a tenth of the data set), children were divided in four groups: a group below P25, a group between P25 and P50, a group between P50 and P75, and a group higher than P75. A 30% Loess smoothing was computed. All groups showed a clear temporary regression effect (see Figure 4); the discontinuity was most obvious in the middle two groups.

Figure 4. Loess curves of percentile borders of the ToM sum-scores plotted versus age.

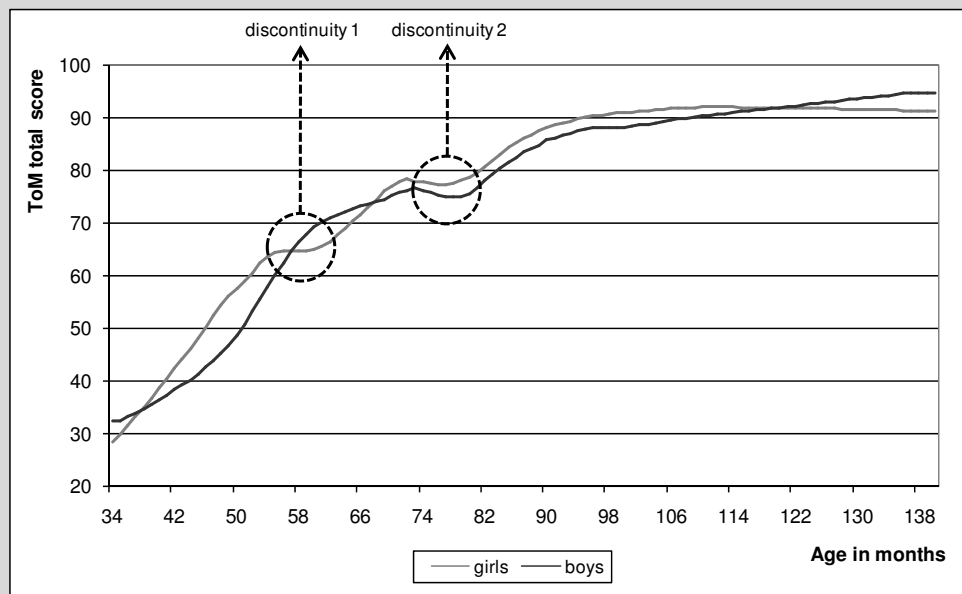


Note: Children were divided in 4 groups: a group below P25, a group between P25 and P50, a group between P50 and P75, and a group higher than P75. All groups showed two clear discontinuities.

Secondly, we compared the curves for both genders. In doing so, some differences became obvious (see Figure 5). The girls showed two deviations from monotonic increase: an increase between the fourth and fifth

year, followed by a plateau (first regression) and then again a growth spurt between the fifth and sixth year, followed by a dip (second regression) and ending with a last growth spurt. The boys showed only one dip, which was more pronounced than the simultaneous dip of the girls. Through slope hunting techniques, we investigated the statistical significance of these dips in the null hypothesis model.

Figure 5. Loess curves of the ToM sum-scores plotted versus age, for boys (black line) and girls (gray line).



Note: The girls show two regressions, the boys only one.

The slope hunting technique went as follows: first, a moving average of age was calculated for windows of 30 consecutive ages in the data set. A corresponding moving slope (i.e. a slope over a moving window of 30 consecutive sum-scores from the sample) was then calculated for the sum-scores, for boys and girls separately. Next, the lowest value of all slopes was determined. At the end, we tested the position of all local slopes minima and compared all combinations by place, amount and distance, with our original Loess curve. The dip of the boys appeared to be significant (Monte Carlo, $p=.007$). The dip of the girls was more flat and appeared not to be significant (Monte Carlo, $p=.20$). However, this dip appeared around three months earlier than in boys. If we reckon with the fact that ToM develops

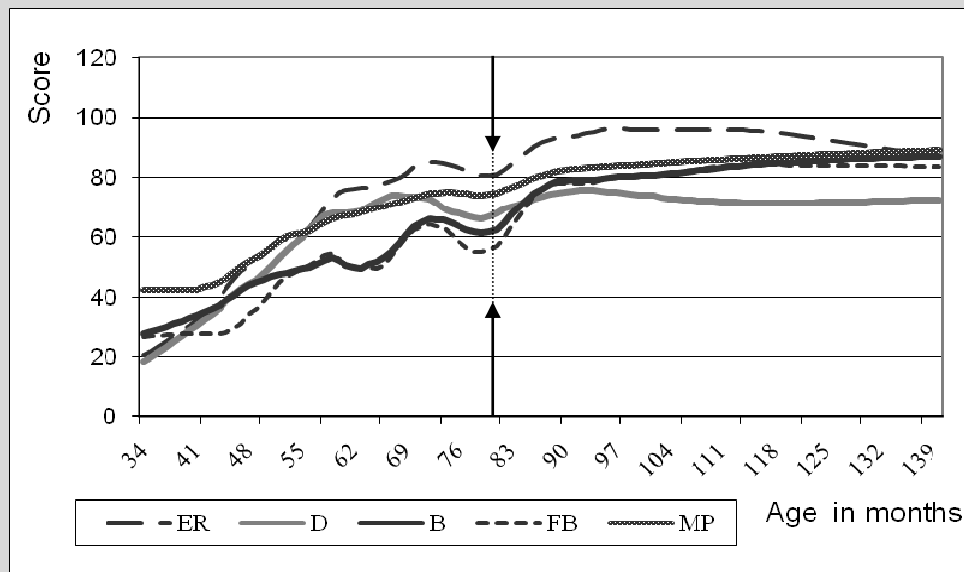
earlier in young girls than in young boys (Charman et al., 2002), the earlier appearance of the dip in the girls seems a meaningful phenomenon. The probability that the occurrence of a dip of this magnitude appearing up to three months earlier but not later than in the boys is about 0.03 (Monte Carlo).

Finally, we wished to check whether the dip is observable in all ToM sub-scores and preferably in the core ToM sub-scores (on desires and beliefs). For this purpose the sub-scores were rescaled, to make comparisons easier. Loesses with a window size of 20% were calculated¹. All ToM sub-scores showed dips at the same ages as the dips found for the ToM sum-score (Monte Carlo, $p=.01$) (see Figure 6). The curves of the core ToM sub-scores showed the same characteristics as that of the ToM sum-score; the sub-score on beliefs and the additional sub-score on only false beliefs (maximum score=9) even displayed the most distinct dips. The deepest point of the dip in the sub-scores was at the same age as the dip based upon the ToM sum-score (at 78 months).

The comparison between sub-scores for various ToM components was also carried out for boys and girls separately. A similar pattern was observable, again with the core ToM sub-scores displaying the most obvious dips. When we looked at the additional sub-score for false beliefs, similar findings were found. However, boys now evidenced two statistically significant discontinuities (Monte Carlo, $p=.006$ and $p=.005$). They appeared at the same age as the discontinuities found for the ToM sum-scores. For girls, the first discontinuity is again not significant ($p=.18$), however the combination of the slopes of both discontinuities is statistically significant (MC, 100 simulations, $p=.05$). Thus, although one discontinuity has a reasonable chance to occur by coincidence, however, two chance discontinuities are very unlikely.

¹ Since the range of the sub-scores is obviously considerably smaller than that of the total test, a window size of 30% as used for the sample of total scores does not have sufficient temporal resolution. For this reason, a smaller window size of 20% was selected.

Figure 6. Loess curves of the ToM sub-scores plotted versus age. All ToM sub-scores showed dips at the same ages as the dips found for the ToM sum-score.



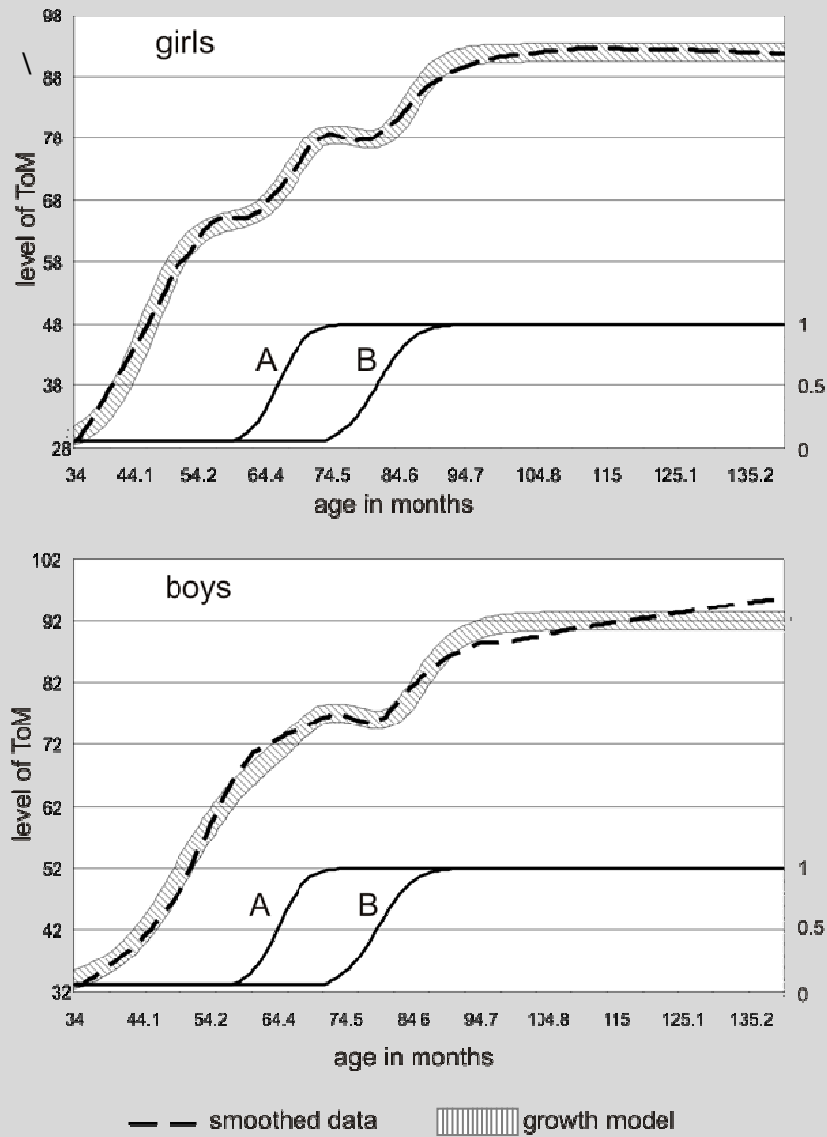
Legend: ER= Emotion Recognition, D= Desire, B= Belief, FB= False Belief, MP= Mental Physical. The deepest point of the dip based upon the ToM sum-score is pointed out with the arrow and dotted line.

Is there a two or three step developmental model?

As we mentioned before, girls evidenced a three-step development and boys more a two step development. However it is highly probable that also boys show a three-step development. It is hypothesized that the first transition is observable only in girls, because of differences in major parameters - in particular the value of the main parameter, which is the growth rate - and not because of differences in the underlying variables affecting the growth of ToM.

In order to show that this interpretation is indeed feasible, we fitted a three-step growth pattern of ToM-knowledge, based on the emergence of two underlying, supportive variables, one around the age of 56 months (A) and another around the age of 72 months (B) (see Figure 7). These supportive variables are hypothetical and may for instance include executive functions, which are known to be important in ToM functioning (for the relation between ToM and executive function see e.g. Carlson et al., 2002).

Figure 7. A three-step growth pattern (broad striped line) fitted over the smoothed ToM sum-score data (interrupted line), taking into account the emergence of two underlying, supportive variables A and B. The top graph shows the fit of girls, the bottom part shows the fit of boys. The underlying variables A and B are of the same magnitude and occur at the same age in both genders.



The growth model that was fitted to the smoothed data is of the type described by Van Geert (1991) and by Fischer and Bidell (2006) and contains positive parameters, i.e. a supportive relationship, for the A and B levels and negative parameters, i.e. a competitive relationship, for the first derivative of the hypothetical A and B levels (which corresponds with the actual change in these levels) (for an explanation of the underlying logic of the model, see Fischer & Bidell, 2006, and Van Geert, 1991, 1994). Figure 7 shows the fit with the smoothed curves of boys and girls separately, based on underlying hypothetical variables A and B which are of the same magnitude and occur at the same age in both sexes. Table 2 shows the values of the model parameters.

Table 2. *Values of the model parameters used for the dynamic systems growth model.*

	Girls	Boys
Growth rate	0.15	0.10
Support from A	0.33	0.10
Support from B	0.24	0.16
Competition from the growth of A	0.00	0.00
Competition from the growth of B	-3.20	-2.50

A striking difference between boys and girls is that the parameter values cause faster growth and more effect of supportive and competitive variables in girls than in boys. The first plateau, which is observable in the girls thanks to their higher growth rate, is in fact concealed in boys, as a consequence of their lower growth rate and lesser effect from the A-variable (which is a hypothetical variable emerging around the age of 4.6 years). The second plateau is observable in both girls and boys. Although the competitive effect of B on ToM is greater in girls than in boys, the observable effect is more salient in boys. This finding may lead to the conclusion that girls evidenced a three-step development and boys only a two step development. However, in dynamic growth models, parameters often show nonlinear co-variations, e.g. competitive effects among variables can be masked by higher growth rates. The dynamic growth model (see Figure 7) showed that the expression of the steps in the form of *observable* plateaus and marked dips may depend on the values of the growth parameters, in particular the value of the main parameter, which is the growth rate. It can be concluded that a dynamic growth model involving the effect of two variables affecting the growth of ToM, one occurring around

the age of 56 months and the other around the age of 72 months, can account for the variety of non-linear phenomena observed in the data, including the differences and similarities, the plateaus and dips between boys and girls.

DISCUSSION

Discontinuity in the development of ToM

Our findings support the general developmental view of ToM. Based on cross-sectional analyses, our test results show that ToM increases with age (with the greatest increase between 42 and 56 months, that is between 3.5 and 4.7 years of age), and that it continues to develop after the age of six. The development before the age of four and a half is evidently continuous. However, after this age discontinuities occur. Two temporary regressions, one around the age of four years and eight months and one at the age of six to six and a half, are found not only in the ToM sum-score curve but also in the ToM sub-score curves.

The regressions can be viewed as indicators of discontinuity in ToM development. We have demonstrated that the probability that the main regression (72-78 months) is either a statistical selection artifact or an experimenter artifact is very small. The application of additional indicators – skewness, growth rate and variability – provided further support for the occurrence of a transition –or two transitions – in the development of ToM. Also, the discontinuity found cannot be accredited to gender differences. Both boys and girls showed a marked regression around the age of six. However, girls also showed evidence for an additional earlier regression, around the age of five.

We have chosen to illustrate that there are discontinuities in ToM development, first with the overall data, and second to go into more detail by taking into account gender and ToM sub-scores. When we looked at the core ToM-components, like desires and beliefs, we found that the two discontinuities (at 56 and at 72-78 months) were statistically obvious in both boys and girls.

There are different views on the manner in which ToM develops in preschoolers. For instance, one view assigns central importance to the occurrence of a conceptual change. This change takes place between the age of three and four/five (Wellman et al., 2001; Gopnik, 1993; Perner, 1991). A

second view implies continuous increases in ToM-related processing abilities rather than radical conceptual shifts in understanding mental states (e.g. German & Leslie, 2000; Carlson & Moses, 2001; Birsch & Bloom, 2004). Our data show evidence in favor of both views. There is a pattern of overall continuous change, with a steep growth of ToM knowledge around the age of four, followed by a more continuous increase of ToM knowledge leveling off towards the age of five and interrupted by a temporary regression around the age of six, which occurs in boys and girls alike. Differences between these two views seem to cohere with the difference between a qualitative and quantitative approach of development. In this article, we have demonstrated the advantages of a quantitative approach if it is used to focus not only on general monotonic trends but also on eventual non-linearities in the data, such as temporary regressions, peaks in skewness etcetera.

Overall, we found boys and girls to follow the same developmental path. However, we also found some gender differences in ToM development. Up to now, gender differences have seldom been reported in ToM research. In fact, most studies find no statistically significant differences between boys and girls, which might be due to the use of tests that are insufficiently capable of capturing subtle individual ToM differences (Baron-Cohen et al., 1997a), or have insufficient statistical power. Similar to the study of Charman and colleagues (2002), our study included a more extensive sample than the majority of current studies do. In addition, we employed statistical techniques that are sensitive to more subtle developmental patterns. Under such methodological conditions, eventual gender differences are more easily recorded from the data, not only in the appearance of ToM skills but also in the rate of ToM development. The early ToM-growth in girls was more rapid than that of boys. Gender difference in the rate of ToM development has been hypothesized before by Baron-Cohen and colleagues (Baron-Cohen et al., 1997a) and by Charman and colleagues (2002) who found that young girls have a ToM-advantage, which disappears as children get older. Such a higher early rate of growth results in a greater likelihood of a later temporary standstill (van Geert, 1994), which has indeed been demonstrated in our data, for girls showed two discontinuities and boys only one. The more rapid ToM growth in girls might be due to the fact that, from the beginning, girls are more focused on sociability. For instance, already in one day old neonates, a definite sexual dimorphism is observable (Connellan, et al., 2000). Next to that, girls also have better verbal abilities than boys (Halpern, 2000), stronger syntactic

abilities and a larger amount of social experiences (Charman et al., 2002). Language is considered an important factor in ToM functioning (e.g. Garfield et al., 2001; de Villiers & Pyers, 2002). Finally, there is evidence that females show more pronounced responses of the mirror neuron system than males (Cheng et al., 2006), and the mirror neuron system has been hypothesized to directly relate to ToM-related abilities in both adults and children (for a review see Oberman & Ramachandran, 2007; and see also further in this article).

Potential explanations for the observed temporary regressions

In this article we reported the discovery of one or two temporary regressions, indicative of either a two- or three-step development. The literature on U-shaped growth and non-linear growth curves in general provides some hints on possible explanations.

The first explanation is that the discontinuities reflect a temporary conflict between competence and performance (Marcus, 2004). According to this view, the development of ToM-competence follows in reality a monotonically rising function, but for some accidental reason, performance on ToM-tests gets a little worse around the age of six, maybe because a particular performance component interferes negatively. The question is of course what this performance factor is. In addition, one may question whether this competence-performance distinction is relevant on the level of testable psychological functions. Dynamic systems theory, as advocated by the late Esther Thelen and her collaborators, makes no distinction between these two levels, and sees a temporary regression as a direct consequence of dynamic interactions between components that are responsible for the production of answers to ToM-questions (Gershkoff-Stowe & Thelen, 2004). According to this view, there is no Theory-of-Mind in the sense of an identifiable, internal conceptual structure. All behavior is soft-assembled, and temporary regressions reflect the ‘continuous changes in the collective dynamics of multiple, contingent processes’ (Gershkoff-Stowe & Thelen, 2004, page 11).

Another point that we wish to re-emphasize is that, from a dynamical point of view, cross-sectional data provide ‘first-encounter’-results (see section Cross-sectional versus longitudinal research), which might be more sensitive to the factors causing the temporary regression than results from frequently repeated measurements. The possibility that the regression effect is, among others, sensitive to testing conditions, does not

reduce its developmental significance. The question is of course which continuously changing aspect or aspects of ToM-related behavior is or are responsible for the observed discontinuities.

According to Brainerd (2004), temporary regressions in performance occur if a particular performance class – for instance the class of ToM-related questions – is served by opposing strategies, or dual processes. It is conceivable that up to the age of six, the child has employed an intuitive and direct solution to ToM-problems, while at around the age of six a new approach begins to emerge, which is more cognitive and reflective in nature (we will discuss this point further in our overview of the mirror neuron system and its potential relation to ToM). The emergence of a second strategy requires a form of reorganization of components responsible for ToM-performance, and the observed discontinuities are likely to reflect this reorganization (Werker et al., 2004; Feldman & Benjamin, 2004; Friend, 2004; Marcovitch & Lewkowicz, 2004; Rogers et al., 2004). That such discontinuities indeed occur as a consequence of continuous, long-term growth in a developing system has been demonstrated by modeling development, either by means of connectionist networks (Rogers et al., 2004) or by means of dynamic systems models of the type advocated by Van Geert, Fischer and others (see Demetriou & Raftopoulos, 2004, for a discussion regarding U-shaped growth). In these models, long-term development is context-specific and dependent on dynamic interactions among many components – biological, cognitive, emotional, behavioral – that constitute the developing system (Van Geert, 1991, 1994, 1998; Fischer & Bidell, 2006; Fischer & Rose, 1994). Relationships between the multiple components in a system can be supportive, competitive, conditional or neutral. The dynamics of these relationships over time explain the emergence of phenomena such as accelerations, decelerations and regressions.

Based on dynamic modeling and indirect evidence from brain development, neo-Piagetian theory predicts relatively major shifts in development around the age of six years, dependent on the context or content of the developmental function (Fischer, 1980; Fischer & Bidell, 2006; Case, 1991). The shift is broadly associated with a marked increase in more reflexive, coordinated ways of thinking in contrast with the more intuitive, uni-dimensional ways of thinking that precede it. Although the application is purely speculative, it might be so that around the age of six the intuitive ToM judgment, which is considered to be largely based on biologically founded forms of empathy (Preston & de Waal, 2002) is

supplemented by a more reflective, cognitive form of ToM-reasoning. In this regard, it has been shown that six-year-olds have little trouble assigning false beliefs to others, but only arrive at a truly interpretive ToM at the age of seven (Carpendale & Chandler, 1996; Lalonde & Chandler, 2002). As predicted by the theories discussed earlier, this emergence of a new ToM-specific strategy might explain the temporary regression found in our data.

The previous explanations all rely on the notion of distinctive, developmentally ordered strategies for solving ToM-problems. In fact, there is supportive but indirect evidence of two 'approaches' to ToM: an intuitive (or automatic) and a reflective (or controlled) route (Lieberman, 2007). Indirect evidence for an intuitive, neuro-physiologically-based understanding of ToM-related properties of other persons comes from the rapidly growing literature on the neuronal systems that underlie the spontaneous understanding of human actions and psychological states of others. An example of such systems is the mirror neuron system (for the relationship between the mirror neuron system and autism, see for instance Gallese, 2006; Hadjikhani et al., 2006; Iacoboni & Dapretto, 2006; Lepage & Théoret, 2006; Oberman & Ramachandran, 2007; Williams et al., 2006b; but see Hamilton et al., 2007 for critical remarks). There is neuropsychological evidence that specific parts of the brain, such as the medial prefrontal cortex and the temporal-parietal junction are involved in the processing of ToM-related information (Frith & Frith, 2003; Kobayashi et al., 2007; Lieberman, 2007; Saxe & Wexler, 2005).

The literature cited shows that the severity of neurophysiological dysfunction in specific areas of the brain is related to the severity of problems in the social domain of autistic children and adults. The ToM of people with autism is thus likely to rest heavily on reflective representation. It is hypothesized that through cognitively mediated routes people with autism are able to compensate for the lack of an intuitive ToM (Baron-Cohen et al., 1993; Dissanayake & Macintosh, 2003; Eisenmajer & Prior, 1991). It is a strategy they can only master if a verbal mental age of eleven years is attained (e.g. Happé, 1995). Typically developing subjects, on the other hand, use the direct biology-based routes as well as the more cognitive ones. Their understanding of ToM is a combination of approaches and strategies (Lieberman, 2007), the combination of which changes across development (Kobayashi et al., 2007). It is not unlikely that the regressions found in our study reflect a major reorganization in the composition of this complex of strategies, but the necessary neurocognitive evidence to support such hypothesis is lacking for ages under six.

It should be noted that the discontinuities found in our data need not reflect a difference in ToM understanding per se but could reflect a developmental difference in other factors necessary for the task. For instance, attention, inhibition and ‘curse of knowledge’ may play a role (e.g. Birch & Bloom, 2004; Carlson & Moses, 2001; German & Leslie, 2000). At the age of six, the development of executive functions undergoes its first active stage of maturation (Brocki & Bohlin, 2004). It is not unthinkable that this development also has consequences for the ToM development of children (Carlson et al., 2002). In the future, paediatric structural and functional neuroimaging might give more insight into spurts of cognitive functions that (positively or negatively) influence ToM performance.

Finally, data recently collected on children with PDD-NOS (Blijd et al., submitted c) show a highly comparable dip in ToM-scores. However, in accordance with the developmental delay in ToM typical of such children, the dip occurs at a slightly later age than in the typically developing children. The delay in the dip supports the conclusion that the dip is a genuine phenomenon of ToM-development, and not of interference with some other non-ToM factor, which is not necessarily delayed in children with PDD-NOS.

A three step developmental model

Visual inspection of the graphs revealed that girls showed two plateaus and boys only one. The dip of the boys coincided with the second (more shallow) dip of the girls. The dynamic growth model showed that the observable effects of the parameters that are likely to govern the occurrence of the dips depend on their covariation with the growth rate parameter. In short, the proposed dynamic growth model explains the non-linear phenomena observed in the data, including the differences and similarities, the plateaus and dips between boys and girls.

Limitations of the research and prospects for further study

One of the restrictions of this research is that it had fewer children in the older age range (from 8 years on), which implies a reduction in reliability at the older ages. Also, the test was probably too easy for the older children since we did not include more advanced ToM tasks that are typically mastered at later ages. Perhaps additional regressions would have been

found at the older ages if second-order belief tasks (Perner & Wimmer, 1985) or more complex emotional constructs would have been used. However, not having included such tasks does not change anything to our main message, that there are discontinuities in ToM development.

Since this research is based upon a new test administered only in Dutch children, one could wonder to what extent cultural and linguistic differences and the choice of particular stories and tasks affected the individual age limits. However, meta-analyses have shown that such factors have only little effect upon the general development of ToM skills in normally developing children (Wellman et al., 2001). Concerning false belief tasks, it is illustrated that researchers can vary the tasks over an extensive set of possibilities without seriously influencing the group performance of children. We trust that the medium, in which the ToM tasks are presented here, the ToM Storybooks, has provided valid results.

The major limitation of our research is that the growth curve of ToM is based on cross-sectional data and should be seen only as a first attempt to describe non-linear aspects of ToM development. But as has been argued in this article, cross-sectional data address the child's 'first encounter'-reaction to explicit ToM questions. This first-encounter is likely to reveal properties that a repeated measure design might eventually conceal, due to repetition and learning effects. However, in order to obtain a better account of non-linear phenomena in ToM, further time-serial longitudinal research is required. After all, individual developmental routes can deviate considerably from developmental pathways based on averages of cross-sectional measurements (see conservation curves of Van der Maas, 1993). Only a longitudinal design provides the opportunity to investigate long-term changes and short-term stability of ToM and its components. The findings from the current article can serve as a guideline for the age periods in which more extensive measurements are needed to find evidence of discontinuities in ToM development. In addition, there is already some converging evidence from longitudinal ToM research both in typically developing children (Serra et al. 2002) and children with PDD-NOS (Blijd-Hoogewys et al., submitted c), further supporting the robustness of this developmental phenomenon.

Further longitudinal research looking into developmental discontinuities in ToM development is recommended. It is essential that such studies comprise a number of repeated measurements that is sufficient to capture the expected non-linearities and that growth models are fitted that are sensitive to such eventual non-linearities.

In sum, this article has explored the existence of temporary regressions in ToM development. Because little is known about the dynamics in ToM development, a cross-sectional design was applied in combination with non-linear fitting methods. Data from the ToM Storybooks, a comprehensive measurement of ToM, showed that a two or three step developmental model can be distilled. There are discontinuities: one at the age of four years and eight months, and one between the ages of six to six and a half. These non-linear phenomena could not be explained as accidental sampling effects and were supported by additional indicators of discontinuity, namely changes in skewness, growth rate and variability. The discontinuities were observable not only in the ToM total score, but also in the ToM sub-scores and in both boys and girls. Finally, the dynamic growth models presented in this article might serve as a starting point for the formulation of a theory of ToM in a broader developmental context.

