Nonlinearities in theory-of-mind development
Blijd-Hoogewys, Els M.A.; Bulgarelli, Daniela; Molina, Paola; van Geert, Paul L.C.

Published in:
European Journal of Developmental Psychology

DOI:
10.1080/17405629.2021.2020636

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:
2022

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

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.

Download date: 27-10-2023
Nonlinearities in theory-of-mind development: New evidence from Dutch and Italian boys and girls

Els M. A Blijd-Hoogewys, Daniela Bulgarelli, Paola Molina & Paul L. C van Geert


To link to this article: https://doi.org/10.1080/17405629.2021.2020636

Published online: 04 Jan 2022.

Article views: 155

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=pedp20
Nonlinearities in theory-of-mind development: New evidence from Dutch and Italian boys and girls

Els M. A Blijd-Hoogewys\textsuperscript{a}, Daniela Bulgarelli\textsuperscript{b,c}, Paola Molina\textsuperscript{c} and Paul L. C van Geert\textsuperscript{d}

\textsuperscript{a}INTER-PSY, Groningen, The Netherlands; \textsuperscript{b}Dipartimento di Psicologia, Università degli Studi di Torino, Torino, Italy; \textsuperscript{c}DIST–Dipartimento Interateneo di Scienze, Progetto e Politiche del Territorio, Università degli Studi di Torino, Torino, Italy; \textsuperscript{d}Faculty of Behavioural and Social Sciences, Department of Clinical and Developmental Psychology, University of Groningen, Groningen, The Netherlands

\textbf{ABSTRACT}
The extent to which Theory of Mind (ToM) performance is influenced by cultural and gender differences remains a subject of debate. A sample of 324 Dutch and 511 Italian children (52\% boys; 2.8–11.7 years; 50\% boys; 2.6–10.3 years; respectively) was administered the ToM Storybooks. Analysis focused on indicators of nonlinearity: moving standard deviations, moving skewness, and moving rate-of-change. Loess curve smoothing was used and showed local peaks in these nonlinear indicators. A first peak was found around the age of 51 months, a local minimum between 70 and 79 months, and a small peak at 85 months. The first peak was statistically significant in all groups (though differing in gender and culture), for two out of three nonlinear indicators phenomena, except for Dutch girls (peak 2 and 3 was significant). These results show a substantially nonlinear development regardless gender and culture, but a different timing in development.

\textbf{ARTICLE HISTORY} Received 25 March 2021; Accepted 13 December 2021

\textbf{KEYWORDS} ToM Storybooks; cross-sectional; loess curve smoothing; culture; gender

\textbf{Introduction}

Theory of Mind (ToM) is the social cognitive ability to attribute mental states – i.e., beliefs, desires, intentions, emotions, and perceptions – and to use these attributions in understanding, predicting, and explaining behaviours of oneself and others (Wellman, 2014). It is crucial for showing socially adequate behaviour (Imuta et al., 2016).

\textbf{Correspondence to} Daniela Bulgarelli daniela.bulgarelli@unito.it Dipartimento di Psicologia, Università degli Studi di Torino, Torino, Italy

The Supplementary material includes (1) an explanation of the Monte Carlo methods used in order to test hypotheses about the probability of the observed nonlinear phenomena if the null hypotheses (of smooth and monotonic change) were true, and (2) figures of all the nonlinear indicators discussed in the article for the total group and the various subgroups (boys, girls, Italy, Netherlands), including tables specifying p-values and confidence intervals for these indicators.

© 2021 Informa UK Limited, trading as Taylor & Francis Group
Previously, we demonstrated that ToM development, as defined in a Dutch cross-sectional sample of subjects of different ages, follows a nonlinear trajectory (Blijd-Hoogewys & van Geert, 2017). Temporary regressions, accelerations and plateaus were found using the ToM Storybooks (Blijd-Hoogewys et al., 2008). The temporary regressions were accompanied by a decrease in growth rate and changes in inter-individual variability in standard deviations, and a change in skewness of the ToM data, all suggesting a developmental shift in ToM understanding. Both boys and girls showed a marked regression around the age of six. However, girls had an earlier growth spurt than boys and the underlying ToM developmental path was more salient in girls than in boys. They evidenced an additional earlier regression (a plateau), around the age of five. On average, girls had slightly higher ToM total scores than boys. This gender difference was significant for the youngest and the oldest group (<54 months and ≥78 months, respectively). Nonlinearities are well-known developmental phenomena. Temporary regressions have been found in several domains, including motor and verbal development, nonverbal symbol learning, creativity, reasoning, face perception and false belief (FB) understanding, a key ToM milestone (Pauls et al., 2013).

Giving the ubiquity of nonlinear phenomena in complex processes such as ToM development, we may assume that nonlinearities occur in the developmental patterns of all cultures, although we do not know how culture might affect their timing and magnitude. Looking at the influence of culture on ToM, FB understanding evidences the same developmental trajectory from below-chance to above-chance performances among several Western and non-Western cultures (Callaghan et al., 2005). There is a similar conceptual ToM development in all countries (Wellman, 2018). However, the sequence of different ToM developmental milestones may differ somewhat (Peterson & Slaughter, 2017; Shahaeian et al., 2011), as well as their specific developmental timing (Linares Pava, 2019). Specificities related to a particular culture might be linked to the mental state lexicon in its language (Liu et al., 2008; Shatz et al., 2003), to features of a collectivistic versus an individualistic society (Ahn & Miller, 2012; Shahaeian et al., 2014), and to schooling (Callaghan et al., 2005; Kuntoro et al., 2013). Studies comparing children’s ToM performances among European countries report
mixed results: British children had higher FB understanding scores than Austrian (Wellman, 2018) and Italian children (Lecce & Hughes, 2010). In turn, Italian children outperformed Finnish children at age 3–4, while the opposite was true at age 6, and at later ages both groups no longer differed (Rosenqvist et al., 2017). Whether ToM performance is culturally influenced remains the subject of debate because the mechanisms that generate ToM differences are unclear (Linares Pava, 2019). Thus, it might be interesting to also look at nonlinear developmental phenomena in ToM performance over culture. The Netherlands and Italy respectively belong to the German and Latin clusters in Europe. Similarities and differences were found between these two countries with respect to implicit cultural models of the child and parenting; for instance, the theme of children’s independence is related to rules and values in Italian parents’ interviews and to the recognition of the child’s own interests in the Dutch parents’ ones (Harkness & Super, 2006). Also, in the Netherlands and Italy, parents attribute different characteristics while describing a temperamentally difficult child (Super et al., 2008). Such differences may impact children’s ToM development (Miller, 2016). In this article, we focus on nonlinear phenomena comparing data from the Netherlands and Italy.

Looking at gender influence on ToM development in children (<12 years), most of the studies did not find differences between boys and girls, nevertheless they often lacked statistical power to demonstrate small gender differences (Charman et al., 2002). However, if we assume that any ToM gender effect would be expected to be a weak effect only, it may be that most previous research has lacked the statistical power to demonstrate such an effect (Barreto et al., 2018; Hughes et al., 2011; Kolodziejczyk & Bosacki, 2015; Kuhnert et al., 2017; Longobardi et al., 2017; Lonigro et al., 2017; O’Hare et al., 2009; for a discussion see, Charman et al., 2002). Exceptions are Charman et al. (2002, N = 1093) and Ibanez et al. (2013, N = 424) who found gender differences, and Ronald et al. (2006, N = 638 twins) who did not. It is noteworthy that often studies pooled together children of different age, and this prevented from detecting differences due to gender that were not stable across the years (Charman et al., 2002). When gender differences were found, usually the studies showed a slight female advantage in preschoolers using FB tasks (Sodian & Kristen-Antonow, 2015;
Walker, 2005) and in school-age children using basic ToM measures (Calero et al., 2013) or advanced ToM tasks (Devine & Hughes, 2013). Since studies differed on ages, countries, and measurements, and the results were inconsistent, it is not possible to report clear conclusions about gender differences in ToM. It is suggested to test gender differences across ages in large-sized samples.

The goal of our study was to test the non-linear ToM development in two cultures, using single cross-sectional measurements. Our main hypothesis was to find indicators of nonlinearities in particular age ranges, in both Dutch and Italian children, and to find differences due to gender in both countries, as our sample sizes should be sensitive enough to detect them.

**Methods**

**Participants**

Typically developing 3- to 11-year-old children from the Netherlands and Italy participated in this research. The Dutch sample consisted of 324 children (52% boys; 2.8–11.7 years; boys: \(M = 72.48, SD = 25.51\); girls: \(M = 75.17, SD = 28.08\) months), the Italian sample consisted of 511 children (50% boys; 2.6–10.3 years; boys: \(M = 72.76, SD = 23.67\); girls: \(M = 69.29, SD = 24.08\) months) (see, Table 1). Despite the different numbers of children in the age groups by country, the percentages showed that the sample was quite balanced, especially in the younger ages that are of our interest, as the paper focuses on basic ToM skills. All children came from kindergartens and elementary schools, from

<table>
<thead>
<tr>
<th>Age (in years)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8–9</th>
<th>10–11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>32 (19%)</td>
<td>31 (19%)</td>
<td>31 (19%)</td>
<td>31 (19%)</td>
<td>15 (9%)</td>
<td>14 (8%)</td>
<td>13 (8%)</td>
<td>167 (100%)</td>
</tr>
<tr>
<td>Girls</td>
<td>29 (18%)</td>
<td>24 (15%)</td>
<td>32 (20%)</td>
<td>26 (17%)</td>
<td>16 (10%)</td>
<td>12 (8%)</td>
<td>18 (11%)</td>
<td>157 (100%)</td>
</tr>
<tr>
<td>All</td>
<td>61 (19%)</td>
<td>55 (17%)</td>
<td>63 (19%)</td>
<td>57 (18%)</td>
<td>31 (10%)</td>
<td>26 (8%)</td>
<td>31 (10%)</td>
<td>324 (100%)</td>
</tr>
<tr>
<td>Italian sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>44 (17%)</td>
<td>47 (18%)</td>
<td>42 (16%)</td>
<td>40 (16%)</td>
<td>39 (15%)</td>
<td>40 (16%)</td>
<td>5 (2%)</td>
<td>257 (100%)</td>
</tr>
<tr>
<td>Girls</td>
<td>51 (20%)</td>
<td>59 (23%)</td>
<td>42 (17%)</td>
<td>34 (13%)</td>
<td>25 (10%)</td>
<td>36 (14%)</td>
<td>7 (3%)</td>
<td>254 (100%)</td>
</tr>
<tr>
<td>All</td>
<td>95 (19%)</td>
<td>106 (21%)</td>
<td>84 (16%)</td>
<td>74 (14%)</td>
<td>64 (13%)</td>
<td>76 (15%)</td>
<td>12 (2%)</td>
<td>511 (100%)</td>
</tr>
</tbody>
</table>
provincial and urban regions. Teachers and/or parents did not report psychological or developmental problems that could have hampered their performance on the tasks.

**Measure**

Hoogewys et al., 2008), a comprehensive test including 34 tasks, measuring several basic ToM components. The tasks concern (1) emotion recognition, (2) distinction between physical and mental entities (real-mental, real-imaginary, and close impostors), (3) understanding that seeing leads to knowing, (4) understanding of desires, and (5) understanding of beliefs (including FB). The tasks are incorporated in six stories on the protagonist, Sam. The books have interesting features for children, e.g., full colour pictures, cuddly patches of fur, toy doors to open, and magnetized emotion faces that can be placed on the characters. The test takes 40 to 50 minutes to administer, including a 5-minute break. The tasks require the children to take Sam’s perspective; for examples see, Figure 1 and Blijd-Hoogewys et al. (2008). ToM questions make a relatively strong appeal to lexical and syntactic knowledge (for a review Bulgarelli et al., 2022). A common variance of 18 to 22% with language was found in a previous article on the ToM Storybooks (Blijd-Hoogewys et al., 2008).

![Figure 1. Example of a task from the ToM Storybooks.](image)
The ToM Storybooks have been translated into different languages (English, Finnish, French, Italian, and Spanish), and have been standardized for the Netherlands (Blijd-Hoogewys et al., 2008) and Italy (Bulgarelli et al., 2015; Molina & Bulgarelli, 2012). Also, an Italian version of the test was adapted for blind children (Bartoli et al., 2019). A maximum total score of 110 points can be obtained. The test has good psychometric qualities (Blijd-Hoogewys et al., 2008). The internal consistency (Cronbach’s α = .90), test-retest reliability (r = .86, p < .001), inter-rater reliability (Cohen’s K = .81-.97), discriminant validity (differentiating children with autism from typically developing children), divergent and convergent validity are good (Blijd-Hoogewys et al., 2008, 2010; Molina et al., 2020). Concerning the content validity, a Principal Component Analysis showed a solution of five components (belief action, emotion recognition, mental physical, belief emotion and desire emotion) as the best theoretical interpretation. The correlations between the components varied from .25 to .45 and the solution accounted for 53.8% of the variance. A confirmatory factor analysis run on 681 Italian children aged 3–8 year also reported five components (emotion, desire, mental–physical, belief and perception knowledge). For a discussion about these two factorial structures, see, Bulgarelli et al. (2015).

**Procedure**

Permission for the study was granted by the ethical committee of the University of Groningen and Università degli Studi di Torino; for each child written consent was obtained in advance from parents. All children were administered the test individually in a quiet room, at school for the Dutch and Italian children, and at home for the Dutch preschoolers (< 4 years). The data collected on children were anonymized and confidentially stored.

**Data analysis and hypotheses**

The study of nonlinearities is based on two types of indicators (Blijd-Hoogewys & van Geert, 2017). One type focuses on central tendencies of scores, namely the average score for a particular age, and the first derivative of the average score, which represents the rate-of-change of the average score over time. When dealing with inter-individual differences, this rate-of-change is in fact a ‘rate-of-difference over time’.
The second type of indicator focuses on typical distributional measures. The first is a measure of dispersion of the scores around the local central value, namely a moving standard deviation. It shows how wide the cloud of scores is for a particular age. The second is a measure of asymmetry, moving skewness.

Our main hypothesis is that the above indicators of nonlinearity, except for the average score, show local peaks in particular age ranges; they will increase towards a local maximum and then decrease towards a later local minimum. Also, since ToM develops over time, basically from a state of ‘absence’ to a state of ‘presence’ in children, it can be expected that the ToM score distributions in both the early and late stages are typically asymmetrical. For, in the early stages, we expect to find a distribution with a tail to the right (i.e., a positive skewness value), where most children have scores below the mean, with a number of ‘forerunners’. In the late stages, we expect a distribution with a tail to the left (i.e., a negative skewness value), as most children will have reached the mastery level of ToM, with a minority of children still lagging and only slowly catching up. A positively skewed distribution suggests a low degree of consolidation of an ability, a negatively skewed distribution a high degree of consolidation (Van Geert & van Dijk, 2002). In summary, over the years, we expect a trend from positive to negative skewness. It is an open question whether this trend is linear, or whether it shows nonlinearity comparable to those of central tendencies and dispersion measures (see above). For our null hypothesis, we expect a monotonic age curve, representing the sequence of central ToM score values over age, and a monotonic and symmetrical dispersion function in the form of a monotonic change in the standard deviation of the ToM scores over age. This null hypothesis can be operationalized by searching for the best possible fitting monotonic curve for the ToM scores and for their moving standard deviations over age.

To acquire insights in nonlinear changes in ToM development, we used a descriptive nonparametric method, namely Loess curve smoothing (Simonoff, 1996). These curves are based upon a locally weighted least squares estimate, with a window of 30 consecutive observations, with a total number of observations that exceeds 800. This was used for the moving standard deviations and the moving skewness. For instance, the standard deviation curve can be found by first calculating the standard deviation over a moving window of observations covering 1–2 months of age, and then fitting a monotonic curve. Afterword, we used random
permutation techniques, and more generally, Monte-Carlo analyses, which are assumption-free techniques (Kroese et al., 2014). Given an age-related function of the expected score and the expected distribution, a Monte-Carlo procedure simulates possible data sets, with a similar number of observations as the number of observations in the data. Differences in sample sizes over data sets are automatically accounted for in the statistical simulation technique. For each of the simulated data sets emulating the null hypothesis, we calculated the statistical indicators of nonlinearity used to specify our data set, namely nonparametrically smoothed growth, rate-of-change, skewness, and distributional curves. By generating a large enough collection of simulated curves (5000), we obtained distributions of these statistical indicators of nonlinearity under the null hypothesis. We could then compare our observed indicators of nonlinearity with the indicators that are possible – with varying probabilities – given the null hypothesis. Such comparisons result in standard confidence intervals and \( p \)-values, which estimate the probability that the observed results could have been found if the null hypothesis were true.

**Results**

**Overall data**

The internal consistency in this study was good (Cronbach’s \( \alpha = .95 \), for both Dutch and Italian items). The ToM total score increased with age, both in boys and girls separately, and in both cultures (see, Table 2). A ceiling effect was present at the older ages: this makes the different numbers of Dutch and Italian children in these groups less noteworthy. None of the children reached the highest test score, mainly due to the qualitative questions: even the older children did not obtain the maximum score as common answers to those questions do not include reference to mental states.

By differencing the Loess smoothed curve of the ToM total scores and the best fitting monotonically rising model, we determined a curve for the rate-of-change (see, Figure 2). This curve shows a salient peak around the age of 51 months, a local minimum between 70 and 79 months, and a small peak at 85 months. The skewness curve (not depicted here) was like the rate-of-change curve, with a peak of positive skew at 47 months,
Table 2. Descriptive statistics of ToM Total score for the Dutch and Italian samples.

<table>
<thead>
<tr>
<th></th>
<th>ToM Total Score</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3 yr</td>
<td>4 yr</td>
<td>5 yr</td>
<td>6 yr</td>
<td>7 yr</td>
<td>8–9 yr</td>
<td>10–11 yr</td>
</tr>
<tr>
<td>Dutch sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td>37.66 (7.03)</td>
<td>60.06 (14.23)</td>
<td>74.74 (9.85)</td>
<td>74.71 (11.72)</td>
<td>86.40 (8.16)</td>
<td>90.07 (8.09)</td>
<td>93.92 (8.12)</td>
</tr>
<tr>
<td>Girls</td>
<td></td>
<td>41.66 (13.13)</td>
<td>64.83 (14.76)</td>
<td>71.56 (13.19)</td>
<td>78.65 (9.95)</td>
<td>88.69 (8.59)</td>
<td>91.42 (7.74)</td>
<td>91.33 (7.32)</td>
</tr>
<tr>
<td>Italian sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td>36.70 (8.28)</td>
<td>51.17 (12.97)</td>
<td>60.74 (14.14)</td>
<td>70.45 (12.52)</td>
<td>78.31 (9.88)</td>
<td>84.75 (12.28)</td>
<td>73.40 (19.44)</td>
</tr>
<tr>
<td>Girls</td>
<td></td>
<td>35.73 (7.16)</td>
<td>48.24 (15.40)</td>
<td>63.24 (12.28)</td>
<td>70.24 (12.49)</td>
<td>76.80 (10.34)</td>
<td>87.97 (7.77)</td>
<td>86.29 (7.11)</td>
</tr>
</tbody>
</table>
a period of zero skew (typical symmetrical distribution) between 60 and 80 months, and then dropped down to negative skew towards 130 months, when most children obtained maximum ToM total scores.

Nonlinear trends were confirmed in the total sample (see, Table 3). First, there was a peak in skewness at 47 months, which was statistically highly significant ($p < .0001$). Second, there was a statistically highly significant peak in the rate-of-change at 51 months. Third, there was a statistically highly significant peak in the dispersion (moving standard deviations) at 56 months and a statistically significant second peak (not depicted in Table 3) around 70 months ($p = .05$). The latter was not accompanied by a statistically significant peak in skewness and rate-of-

**Figure 2.** Loess smoothed rate-of-change curve of ToM total score in the total sample.

**Table 3.** Summary of nonlinearity variables at age peak 1 (ages in months).

<table>
<thead>
<tr>
<th></th>
<th>Skewness</th>
<th>Rate-of-change</th>
<th>Standard deviation</th>
<th>Skew slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>p</td>
<td>Age</td>
<td>p</td>
</tr>
<tr>
<td>All children</td>
<td>47.0</td>
<td>&lt;.0001</td>
<td>51.0</td>
<td>.0022</td>
</tr>
<tr>
<td>- all boys</td>
<td>50.0</td>
<td>.05</td>
<td>51.0</td>
<td>.0004</td>
</tr>
<tr>
<td>- all girls</td>
<td>46.0</td>
<td>&lt;.0001</td>
<td>52.0</td>
<td>.053</td>
</tr>
<tr>
<td>All Dutch children</td>
<td>45.0</td>
<td>.03</td>
<td>51.0</td>
<td>.0288</td>
</tr>
<tr>
<td>- Dutch boys</td>
<td>ns</td>
<td></td>
<td>56.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>- Dutch girls</td>
<td>ns</td>
<td></td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>All Italian children</td>
<td>50.0</td>
<td>.01</td>
<td>50.0</td>
<td>.0388</td>
</tr>
<tr>
<td>- Italian boys</td>
<td>ns</td>
<td></td>
<td>49.0</td>
<td>.07</td>
</tr>
<tr>
<td>- Italian girls</td>
<td>49.0</td>
<td>.01</td>
<td>ns</td>
<td>58.0</td>
</tr>
</tbody>
</table>
change. The probability that the null hypothesis demonstrated peaks in all three nonlinearity conditions of the total sample was below 1/5000, also for two out of three conditions. A considerably weaker result was obtained for the second peak (p < .03, for all 3 conditions).

Looking at the inter-individual differences over time, we found a linear decrease in skewness over age, from a high positive to a high negative skew. We checked whether this decrease in skewness, irrespective of the occurrence of a clear peak in the skewness (i.e., an early increase in the skewness followed by a rapid decrease), was also likely to be found if the null hypothesis were true. The data showed a statistically highly significant intercept (1.76, p < .0001) and negative slope (−0.03, p < .0001). All intercepts under the null hypothesis were smaller than the observed intercept, and all slopes under the null hypothesis were greater than the observed slope (which means, less negative).

**Gender differences**

A similar picture as seen in the total group, is seen for both genders (see, Table 3), though at somewhat different ages, with girls having the first peak in their ToM development earlier than boys (concerning skewness and standard deviation, not rate-of-change). The peak in skewness always preceded the peak of rate-of-change, after which the peak in standard deviation followed. Concerning the peak in rate-of-change, girls showed a statistical trend.

Visual inspection of the Loess smoothed age curves of the ToM total score show that these curves are similar for boys and girls within each culture (see, Figure 3(a) and (b)). The rate-of-change of the ToM total score, operationalized by the Loess smoothed difference curves, shows clear qualitative similarities within gender (see, Figure 4(a) and (b) if the curves are shifted across age: −6 months for Dutch boys and −6 months for Italian girls). Note that on visual inspection, there are three peaks (summits). However, only the first one reached statistical significance. In Dutch boys, only two peaks are visible.

Also, girls and boys appeared to be relatively similar in the beginning (eventually with some delay), but then the cross-sectional ToM growth in girls accelerated in comparison to the ToM growth in boys (something which seemed to start at the age of approximately five years).
Figure 3. (a) and (b). Age curves of ToM total score (Loess smoothed) of Dutch and Italian children.
Figure 4. (a) and (b). Qualitative similarities in the normalized age curves of rate-of-change of ToM total score (Loess smoothed, normalization to values between 0 and 1).
As regards the different countries (see, Table 3), Dutch children had the first peak in their ToM development earlier than Italian children (concerning skewness and standard deviation, not rate-of-change). Looking at change scores, there were pronounced significant differences between the Netherlands and Italy (see, Figure 5), which changed over time. At the beginning Dutch children had higher change scores, later Italian children had higher change scores. In the end, the change scores were no longer statistically different.

All genders over culture showed the first peak in ToM total score, which was accompanied by two out of three different nonlinear developmental phenomena, except for Dutch girls. However, Dutch girls did show a second and third age peak (at 67 and 85 months); Dutch boys also showed a second age peak (at 71 months); all were observed only in the rate-of-change ($p = .01$, $p = .02$ and $p = .04$, respectively). Overall, the Dutch girls had the highest scores (see, Table 2, except for age 5 and 10–
11). At start, Dutch girls were quicker in their ToM development than Italian girls, who were 10 months later (44 and 54 months, respectively). Italian boys were quicker than Dutch boys, who were 6 months later (46 and 52 months, respectively).

**Discussion**

Our results demonstrate nonlinearities in ToM development in Dutch and Italian children. ToM increases with age. As expected, a first peak in ToM ability was found around the age of 51 months in almost all groups. The clearest ToM growth in the preschool years is expected, since the ToM Storybooks measure basic abilities (Blijd-Hoogewys et al., 2008; Bulgarelli et al., 2015). In support of a universal ToM development, a typical pattern was observed in the two countries, with first a peak in the skewness of scores, then a peak in the central tendency of scores (rate-of-change) and finally a peak in the dispersion of scores (standard deviations). Peaks in the dispersion signify that there are temporary increases in the score differences between children, i.e., individual differences are considerable, and they are not evenly distributed across a mean. As expected, at the beginning, there were some forerunners being better in ToM than most of their peers (positive skewness) and at the end some late bloomers lagging in ToM in comparison to their peers (negative skewness). For more information on nonlinearity, see Blijd-Hoogewys & van Geert (2017). Note that none of the children reached the highest test score, mainly due to the qualitative questions. Two explanations are possible here: first, when faced with questions that seem too easy, older children can give wrong answers because they over-interpret the question or, alternatively, they do not explicitly refer to mental states, as a behavioural explanation might seem more common.

Specificities in ToM development due to culture were also found: the Dutch children showed a clear age difference, while the Italian children showed a weaker inflection. The differences between The Netherlands and Italy changed over time: the Dutch children had higher change scores at the age of 51 months, the Italian children at the age of 50 months. At 79 and 71 months respectively, the change scores were no longer statistically different. This result is in line with the literature: British and Finnish children outperformed Italian children in FB tasks at 5 and 6 years, and differences disappeared later (Lecce & Hughes, 2010; Rosenqvist et al., 2017). Our article shows how culture is associated with ToM, comparing
the Netherlands and Italy, though not by what mechanism this association works. Our contribution lies in the use of a comprehensive test and in the nonlinear analysis. Further studies addressing concurrent individual and social variables to explain such differences are needed.

The observed developmental trend (group curve) should be seen as the result of age, individual and cultural factors, and gender. Overall, Dutch girls had the highest ToM scores. Looking at the rate-of-change, Italian boys, Italian girls, and Dutch girls displayed three pronounced peaks. Those are signs for a faster ToM growth at specific moments. Dutch boys only evidenced two peaks. Our more detailed analysis highlights differences that are more difficult to grasp with other methods, as Charman et al. (2002) pointed out. Our study considers a large sample and shows that the nonlinear parameters were similar within gender. Girls and boys showed similar peaks and rate-of-change in their development, but girls developed earlier. Girls seem advantaged in ToM development, at certain ages, and this is the case in both cultures. But the interpretation of these differences is still open. Our study shows that the rhythm of development (not the final level) is different for boys and girls, but since our sample is cross-sectional, that is merely a suggestion.

It follows from the non-ergodicity (alternatively non-homology or ecological fallacy) principle that population-based distributions should not be interpreted as (idealized) developmental curves of an imaginary ‘average’ subject. Development is a process of very often idiosyncratic intra-individual differences over time, and intra- and inter-individual variability structures are in principle different. Nevertheless, knowledge of nonlinearities in development indicated by group measures is particularly important for diagnostical reasons, when individual children are compared with the distribution of scores in their age range. By assuming that the age-specific scores change, in terms of levels and dispersions, diagnostic norms may under- or overestimate true levels if they don’t reckon with eventual non-linearities in the score distributions across age.

A first limitation of this study is the lack of comparison between the two countries concerning the socioeconomic status of the children’s families; the collection of data such as parents’ educational level was not undertaken in the Netherlands. A second limitation is that the ToM Storybooks question only basic ToM skills, typically acquired in preschool years. It would be interesting to include more advanced ToM tasks, that are typically mastered at later ages. A third limitation is the low number of
10–11-year-old Italian children; nevertheless, the focus of this paper is on the development of basic ToM which mainly concern young children. In addition to charting developmental trends across age cohorts, it is necessary to also study ToM development in individual children by means of intensive time-serial designs, to estimate how individual trajectories relate to group-based average trajectories and nonlinearities.

To conclude, the underlying process of ToM development, with its nonlinearities, is quite similar over culture and gender; the specific timing evidences some differences. More research is warranted.

Acknowledgments

The authors thank the children and families who participated in the studies and the kindergartens and schools that took part in the project.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are available on request from the corresponding author, DB. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

ORCID

Daniela Bulgarelli http://orcid.org/0000-0001-5671-3841

References


