Second-Order False Beliefs and Recursive Complements in Children with ASD

Irina Polyanskaya, Torben Braüner, and Patrick Blackburn

1. Introduction

Much is known about how Theory of Mind (ToM) is acquired and the factors which facilitate its development; it is now well established that children’s understanding of other minds depends on a number of language skills and working memory development. But most previous research has focused on first-order (FO) false belief (FB) understanding, using different versions of FO unexpected contents (Smarties style) and transfer (Sally-Anne style) tasks. Second-order (SO) false belief understanding, however, is far less studied and relatively little is known about its development (Miller, 2012), particularly when it comes to children with Autism Spectrum Disorder (ASD).

While there is agreement that second-order reasoning is more complex than first-order reasoning – there is one new mental state in the reasoning chain, and it is a belief about another agent’s belief - the cognitive status of the shift from first- to second-order FB understanding remains unclear. Echoing the divide between the Vygotskian and Piagetian traditions, two main theoretical positions have been proposed: conceptual change and complexity-only (Miller, 2012). According to the conceptual change position, gaining SO FB competency is a qualitative transformation of the underlying thought system, one requiring acquisition of new conceptual resources. According to the complexity-only position, on the other hand, SO FB development does not involve significant conceptual change: rather, once FO FB conceptual abilities have been acquired, performance on SO FB tests is related to higher information processing skills. Some autism research findings seem to support this position (Tager-Flusberg & Sullivan, 1994).

Ever since the seminal study of Baron-Cohen and his colleagues in 1985, the Theory of Mind deficit account has been an influential explanation of the difficulties that children with ASD experience, namely delayed or different development of the ability to represent other minds (Tager-Flusberg, 2007). Even though some studies report findings that do not support the ToM deficit account
(Bowler et al., 2005), and other significant theoretical accounts have been put forward (Rajendran & Mitchell, 2007), the Theory of Mind deficit account remains influential, and is applied in both theoretical and clinical settings. For example, a recent study with Danish-speaking children, comparing Theory of Mind, Executive Function and Local Bias accounts, reports that ToM deficits differentiate children with ASD from those with typical development at baseline and across time (Cantio, 2016). Thus, a deeper understanding of ToM is relevant to understanding a key theoretical approach to ASD; and exploring the cognitive factors that are important for SO FB development in ASD may turn out to be clinically and educationally valuable.

Both language and working memory seem of special relevance for the ASD population. On the one hand, language has been viewed as providing compensatory techniques for developing understanding of other minds - some studies say that children with ASD use language to “hack out” the solutions to FB tasks (Lind & Bowler, 2009), and other studies say that they use language as a “scaffolding” in developing the capacity to understand mental states (Tager-Flusberg & Joseph, 2005). A recent training study with Danish-speaking children with ASD has confirmed that language can support FO false belief understanding (Boeg Thomsen, 2016). On the other hand, several studies have reported that executive functions - including working memory - correlate with and predict FO FB in children with ASD (Boeg Thomsen 2016; Pellicano, 2007).

FB reasoning is involved in everyday communicational and social situations, often combined with pragmatic and affective elements. In the FO case, Wellman & Liu’s (2004) investigation indicated a clear ascending order of difficulty in understanding desires, beliefs, perceptual access to knowledge, false beliefs and hidden emotions for typically developing children. Using the same methodology, Peterson, Wellman & Liu’s (2005) study with children with ASD has illustrated that the same order of difficulty applies for desires, diverse beliefs and access to knowledge, but that false beliefs are harder than hidden emotions; indeed, they suggest that individuals with ASD may have a distinctive difficulty with the type of mental state understanding that is needed for false belief tasks. All this suggests that it may be valuable to move beyond the FO case and investigate SO FB tasks in detail for children with ASD.

2. Objectives

In our study we investigated possible developmental links between second-order false belief understanding, linguistic recursion and working memory, and our participants were Danish-speaking children with ASD. By linguistic recursion, we mean “the embedding of a constituent inside a constituent of the same category” (Pinker & Jackendoff, 2005). This is relevant to FB tasks because it is linguistic recursion that enables us to build sentences to talk about beliefs about beliefs (Anne believes that Sally believes that the ball is in the basket) and indeed, about beliefs about beliefs about beliefs, and so on. As mentioned above, Tager-Flusberg & Joseph (2005) suggested (in the setting of FO FB tasks) that
sentential complements might provide a “scaffolding” for ToM development in children with ASD; one of our goals was to investigate this idea in a SO FB setting, applied to recursive sentential complements. Moreover, de Villiers, Hobbs, & Hollebrandse (2014) have hypothesized that mastery of recursive sentential complements is necessary for recursive false belief reasoning. Our own previous logical analysis of SO FB tasks highlighted the importance of recursion: it shows that SO reasoning can be viewed as the recursive embedding of first-order reasoning about different agents - but as recursive logics of belief are more complex than those required to analyze FO FB tasks, processing issues are also relevant and should be experimentally investigated (Braüner, Blackburn, & Polyanskaya, 2016b). This brings us to the other cognitive competency that we are investigating: working memory. In addition to the earlier mentioned findings regarding working memory’s predictive role for FO FB (both for children with ASD and typically developing), a recent study has identified working memory as a significant predictor of SO FB competencies in typically developing population (Arslan, Hohenberger, & Verbrugge, 2017).

3. Methods and materials
3.1. Subjects

There were 62 children in our sample. Their age range was 6-16 years old, with four most frequent age groups: 10 year-olds (19%), 11-year-olds (16%), 12-year-olds (11%) and 13-year-olds (14.5%). Most participants were recruited from schools for children with special educational needs in the Sjælland region of Denmark, though five children from the sample were recruited via direct contact with parents. The children had to satisfy the following criteria in order to be initially included: parental consent had to be obtained, they had to be diagnosed with ASD (based on a formal evaluation by a specialist), they had to be aged 6-16 and they could have no medical treatment affecting cognitive performance. Moreover, they had to have Danish as their native language, be monolingual, have no learning difficulties or language delays at time of recruitment, and (based on a teacher’s assessment) had to have the emotional readiness to undergo a testing situation and a training program.

3.2. Procedure

In the first step, children were given standard tasks from verbal comprehension and working memory indices (WISC-IV) as well as general grammar comprehension (TROG-2). The results of these tests served as quantitative inclusion criteria: we only selected children with working memory and language skills within the age norms specified by these tests (the standard score for VC and WM had to be at least 80 for a child to be included to the study). Children who met all the above-mentioned criteria were included in the correlation study, and in the next step they were given SO FB tasks and recursive complement tasks. Teachers were asked to complete a Social Responsiveness
Scale (SRS) questionnaire, a validated quantitative measure of autistic traits developed by Constantino et. al. (2003) and suitable for use in research studies of ASD. The testing sessions were carried out in schools or in private homes, in an environment familiar to children. All children were tested individually over two or three sessions, which usually lasted between 45 and 60 minutes each, depending on the child.

4. Second-order false-belief tasks

We made use of all four different patterns of SO FB reasoning that we are aware of (Braüner, Blackburn, & Polyanskaya, 2016a). The four SO FB tasks were randomized in order of presentation across children. The four tests are Danish translations of standard SO FB stories: Ice cream (Perner & Wimmer, 1985), Puppy (Sullivan, Zaitchik, & Tager-Flusberg, 1994), SO Sally Anne (Baron-Cohen et. al., 1999), and Bake Sale (Hollebrandse, Van Hout, Hendriks, 2014). Analysis of the logical structure of the four tasks shows that they are all distinct (Braüner, Blackburn, & Polyanskaya, 2016a).

In all four tasks, obtaining the correct answers requires the child to understand/represent what one protagonist of the story believes that another protagonist believes, where the first protagonist’s belief is false. Two of the stories (Ice-Cream and Puppy) were taken from the Danish version of the validated Dutch ToM Storybook Frank (Blijd-Hoogewys et. al., 2008), translated and validated for Danish by Clemmensen et. al. (2016). Here is the ice-cream story: Frederik, Katrine and the ice-cream man are in the park. Katrine goes home to get money to buy ice-cream. While she is gone, the ice-cream man tells Frederik that he is going over to the market square to sell ice-cream there. As he drives over to the market square, Katrine happens to see him and he tells her the same thing. But Frederik does not know this. Later on Frederik goes to Katrine’s house and her mother tells him that Katrine has gone out to buy ice-cream. So Frederik runs off to look for Katrine. The child participants are asked ‘‘Where does Frederik think Katrine has gone?’’ (judgement question) and they are asked to justify their answer.

None of the test questions in any of the four stories we used contained any recursive sentential complements. That is, participants did not need to have mastered this form of linguistic recursion to understand the questions. In all the stories children were asked memory or first-order false belief questions to ensure that they understood the scenarios. In total, children were asked 10 SO FB questions (6 judgement and 4 justification) and 11 control questions (4 first-order and 7 memory). The score range for judgement answers was 0-6 (two of the tests have 2 judgement questions) while for justifications it was 0-12, yielding our composite score range of 0-18. A justification score of 0 was given for incorrect explanation. A justification score of 1 was given for correct reasoning without any references to mental states (Because that is what the ice-cream man told him), a score of 2 was given for correct reasoning using one mental state verb (Because
he does not know that she saw the ice-cream man), and 3 was given for using two or more mental verbs (Because he thinks that she does not know that the ice-cream man left the park). Participants were only credited with points for SO FB questions if they had passed the control questions.

5. Recursive embedding tool (RET)

We found no validated test in Danish to measure the ability to comprehend and/or produce recursive embeddings, so we developed one for the purposes of this study and validated it based on responses from 240 typically developing Danish-speaking children and 15 adults. RET initially consisted of two subsections, possessive nouns and sentential complements, but since there was a clear ceiling effect for the possessive nouns subsection, we only used the sentential complements subsection in the present study. The subsection on sentential complements consists of five items. The items for this subsection were shaped as short stories concluded by questions; participants received 1 point for a correct answer and 0 for a wrong one. The stories and accompanying images are inspired by the hypothesis that in order to recognize genuine sentential complement recursion, children need truth-value contrasts between clauses (De Villiers, Hobbs, & Hollebrandse, 2014), thus the stories include a statement that does not correspond to the story-world’s reality to prompt the recursive reading. Such considerations lie behind the following RET example (note the truth-value contrast between the rain outside the window and Mom’s statement that the sun is shining):

Lilja is talking to her mom in the kitchen. Mom says: “The sun is shining outside. Go get your brother from his room and let’s go out and enjoy the sun”. Lilja goes to her brother’s room and they are looking out of the window. It is raining outside. Lilja tells brother that mother says that the sun is shining outside. What does mom say? What does Lilja say to her brother? (Translation from the Danish original.)
6. Results

In our first analysis, we inspected the interrelationships among composite second-order false belief scores, recursive complements, general grammar comprehension, working memory, age in years and autistic severity symptoms (SRS). Not all variables were normally distributed, which is why we used a nonparametric measure of the strength and direction of association, namely Kendall's tau-b ($\tau_b$). Table 1 shows the correlation coefficients and strength among study variables:

<table>
<thead>
<tr>
<th></th>
<th>SOFB</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RET (1)</td>
<td>.28</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROG (2)</td>
<td>.27</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM (3)</td>
<td>.36</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (4)</td>
<td>.36</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRS</td>
<td>-.16</td>
<td></td>
<td>-.03</td>
<td>-.002</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 1 (** means significant at the 0.01 level)

As can be seen from Table 1, SOFB scores correlated significantly and positively with all variables except SRS, where – as expected – there was a negative correlation, but it was not significant.

In our second analysis, we used hierarchical regression analysis to test whether mastery of linguistic recursion (RET) predicted SOFB understanding, when controlled for age, working memory (WM) and grammar comprehension. The SOFB score was the dependent variable and age, working memory and TROG were entered first, followed by RET scores. All relevant statistical assumptions were checked. With age, WM and TROG, the model explained 29.4% of the variance in SOFB understanding, $F(3,54)=7.50$, $p<.001$. With RET scores, the model explained an additional 5.6% of the variance, $F(4,53)=7.15$, $p<.001$. Regression coefficients, standard errors and p-values for the full model can be found in Table 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.50</td>
<td>.24</td>
<td>2.05</td>
<td>.046</td>
</tr>
<tr>
<td>WM</td>
<td>.22</td>
<td>.13</td>
<td>1.74</td>
<td>.088</td>
</tr>
<tr>
<td>TROG</td>
<td>.21</td>
<td>.35</td>
<td>.58</td>
<td>.561</td>
</tr>
<tr>
<td>RET</td>
<td>.66</td>
<td>.31</td>
<td>2.14</td>
<td>.037</td>
</tr>
</tbody>
</table>
As Table 2 shows, as well as recursive complements mastery (p<.05), age was also a significant predictor (p<.05), but working memory was only marginally significant (p=.088), and grammar comprehension was not significant at all.

In our third analysis, we divided up the total working memory score used in the previous analysis into simple and complex memory scores. The difference is that simple memory tasks tap what is usually referred to as short-term memory, while complex memory tasks involve the ability to manipulate information kept in short-term memory. More precisely, in order to investigate the contribution of simple and/or complex memory scores, we performed a multiple regression like the full model above, but with the total working memory divided up into simple and complex memory scores. We excluded grammar comprehension as this variable was not significant in the previous model (our second analysis). The relevant statistical assumptions were checked, and two values had to be filtered from the analysis because of the high leverage. The multiple regression model statistically significantly predicted SOFB scores, $F(4,54)=7.16$, p<.001, with 34.6% of the variance explained. Regression coefficients, standard errors and p-values for the resulting model can be found in Table 3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.56</td>
<td>.23</td>
<td>2.44</td>
<td>.018</td>
</tr>
<tr>
<td>Simple WM</td>
<td>-.18</td>
<td>.43</td>
<td>-.43</td>
<td>.668</td>
</tr>
<tr>
<td>Complex WM</td>
<td>.37</td>
<td>.18</td>
<td>2.07</td>
<td>.043</td>
</tr>
<tr>
<td>RET</td>
<td>.66</td>
<td>.30</td>
<td>2.19</td>
<td>.033</td>
</tr>
</tbody>
</table>

As can be seen in Table 3, three variables were significant (p<.05), namely age, complex working memory scores, and recursive complements, however simple working memory scores were not. Thus, splitting the marginally significant variable total WM indicated the predictive role of complex WM tasks, but not of simple WM tasks.

Our last analysis was inspired by visual inspection of distribution of SOFB scores. As can be seen in Figure 1, there are two clusters of composite scores (at around 5 and 13 points) which suggests that total SOFB can be investigated on two levels when analyzing impact of the independent variables: lower SOFB mean and higher SOFB mean. We ran cluster analysis (an explorative analysis that tries to identify structures within the data) on the composite SOFB score, which confirmed that the data could meaningfully be split into two clusters: one with mean SOFB of 5.30 (SD=2.15, range 1-9) the other with mean SOFB of 12.76 (SD=1.92, range 10-17). Both clusters consisted of almost equal numbers of children (n=29 and n=33), and were both normally distributed, according to a
Shapiro-Wilk test. One-way ANOVA showed that the means were statistically
different, F(1,60)=204.058, p<.001.

As we did in our second analysis, we then conducted hierarchical regression
on each cluster to test whether RET, the independent variable we are most
interested in, could be regarded as predictor for high as well as low SOFB
performance. Age and working memory were the independent variables we
controlled for, but grammar comprehension was not included this time because
of the small sample size. The relevant statistical assumptions were checked, and
two values had to be filtered from each cluster in the analysis because of the high
leverage. Intriguingly, the results for the cluster with mean 5.35 (n=31) and for
the cluster with mean 12.70 (n=27) were not the same.

In the lower SOFB mean case, the model with age and WM accounted for
34% of the variance in SOFB understanding, F(2,27)=6.94, p<.005. By adding
RET scores, the model accounted for an additional 13.9% of the variance,
F(3,26)=7.95, p<.005, and the change was significant. Regression coefficients,
standard errors and p-values for the full model can be found in Table 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.55</td>
<td>.16</td>
<td>3.55</td>
<td>.001</td>
</tr>
<tr>
<td>Total WM</td>
<td>-.04</td>
<td>.11</td>
<td>-.38</td>
<td>.704</td>
</tr>
<tr>
<td>RET</td>
<td>.54</td>
<td>.21</td>
<td>2.63</td>
<td>.014</td>
</tr>
</tbody>
</table>
As can be seen in Table 4, age and RET were significant predictors of SOFB in the cluster with lower SOFB mean. In the higher SOFB mean case, on the other hand, the model with age and WM accounted for 29.3% of the variance in SOFB understanding, F(2,24)=4.98, p<.05. Furthermore, adding RET scores did not add any significant change to the model. Regression coefficients, standard errors and p-values for the full model can be found in Table 5.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.28</td>
<td>.17</td>
<td>-1.60</td>
<td>.124</td>
</tr>
<tr>
<td>Total WM</td>
<td>.26</td>
<td>.09</td>
<td>2.84</td>
<td>.009</td>
</tr>
<tr>
<td>RET</td>
<td>-.05</td>
<td>.28</td>
<td>-.17</td>
<td>.864</td>
</tr>
</tbody>
</table>

As can be seen from Table 5, RET and age were not significant predictors, and only working memory was significant for the cluster with higher SOFB mean. The results clearly demonstrate that different levels of mastery in SOFB reasoning are predicted by different independent variables.

7. Discussion

Our findings indicated that mastery of recursive complements was a strong, significant predictor of second-order false belief understanding, even after accounting for age, general grammar comprehension, and working memory. Furthermore, our results highlighted the importance of the complex memory tasks, that is, the tasks that require ability to both keep information in memory and manipulate it in a given short period of time.

Additionally, our results clearly demonstrate that different levels of mastery in SOFB reasoning are predicted by different independent variables. When we switch from the lower SOFB mean cluster (Table 4) to the higher SOFB mean cluster (Table 5), the significant predictors switch from the age and RET variables to the working memory variable. Thus, it seems that when children have not (yet) developed SOFB mastery, being able to handle recursive complements in language may scaffold the required reasoning, but it is working memory that scaffolds it once children have become better at it.

These results raise many further questions. What does RET scaffold exactly in SOFB reasoning (or in what way) – ability to judge false beliefs or ability to justify judgement/ability to explain reasoning? Would results regarding RET’s being a predictor be different for typically developing children? And are there differences here among the four logically distinct SOFB tasks? But rather than attempt to answer such questions here, we will briefly comment on another aspect of our work, namely its training component.
As we mentioned earlier, our initial inclusion criteria included a teacher's positive assessment of emotional readiness to undergo a testing situation and a training program, as we wanted not only to investigate whether there was a correlation between SO FB mastery and recursive sentential complementation, but also to provide experimental evidence that mastery of syntactic recursion plays a causal role in the development of SO FB understanding.

Our training used materials similar in style to the RET testing material, and aimed to train participants to grasp the following four principles:

1. That several linguistic constituents of the same type may be combined together.
2. That these constituents may be embedded one inside another.
3. That changing the order of embedding changes the meaning.
4. That the number of embedded constituents is potentially unlimited.

For a more details on the design of the training program see (Polyanskaya, Blackburn, & Braüner, 2017). Here we'll merely remark that we divided participants into three groups: the control (interaction only) group, the WM group (who received working memory training via three games on a PC) and those who received the RET-style training. Our results showed the effectiveness of our recursive embedding training and we are currently preparing a paper on this topic. Irina Polyanskaya's forthcoming (late 2018) PhD thesis contains chapters on both the correlation and training aspect of the study reported here.

Acknowledgments

The authors acknowledge the funding received from the VELUX FOUNDATION for the project Hybrid-Logical Proofs at Work in Cognitive Psychology (VELUX 33305). We would like to thank Jens Ulrik Hansen for help with the statistical analysis and Ida Aaen for help with the testing. Irina Polyanskaya would also like to thank Inge-Marie Eigsti for hosting her at the University of Connecticut, and for her extremely helpful discussion on all aspects of the work reported here.

References


