Preprint

This is the submitted version of a chapter published in *LITMUS in Action: Cross-comparison studies across Europe*.

Citation for the original published chapter:

MAIN story comprehension: What can we expect of a typically developing child?
In: Armon-Lotem, Sharon & Grohmann, Kleanthes (ed.), *LITMUS in Action: Cross-comparison studies across Europe* Amsterdam: John Benjamins Publishing Company
Trends in Language Acquisition Research (TILAR)

N.B. When citing this work, cite the original published chapter.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-392551
MAIN story comprehension: What can we expect of a typically developing child?

Ute Bohnacker & Josefin Lindgren, Uppsala University


Abstract
This study investigates story comprehension in 124 typically-developing children aged 4–6 growing up in Sweden, using the LITMUS-MAIN (Multilingual Assessment instrument for Narratives, Gagarina et al. 2012). 72 Swedish monolinguals and 52 Swedish-English bilinguals from similar socio-economic backgrounds told two MAIN stories each and answered standardised questions that probe inferencing of goals and emotions of story characters. Overall comprehension scores were calculated per narrative, and responses to individual questions were analysed. Results were compared across age groups, languages, and tasks. Already at age 4, and certainly at age 6, overall comprehension scores were high, indicating good general understanding of the stories. Scores differed significantly between tasks, suggesting that comprehension of MAIN Cat/Dog is easier than Baby Birds/Baby Goats. For both tasks, protagonists’ goals were understood well at age 4, approaching ceiling (>90%) at age 5–6. Internal states as initiating events were also relatively well understood. However, many children struggled to infer emotional reactions of protagonists when this necessitated taking the whole plotline into account. We suggest that the processing abilities of the average 6-year-old are not yet developed enough to master every aspect of inferential understanding probed on the MAIN. Results held across languages and mono-/bilinguals, suggesting that they generalize.

Keywords (max 5)
emotion, inference, MAIN, story comprehension, Swedish

1. Introduction
Oral narratives yield rich data about a child’s multiple linguistic abilities and give an index of the child’s cognitive and social abilities, which makes them valuable to speech-language pathologists and therapists, educators, psychologists and linguists. Yet, research into and assessment of the narrative abilities in multilingual children has long been hampered by the fact that many elicitation and assessment tools are designed and norm-referenced for monolinguals, that they are not always cross-culturally robust, and that they are often available for only one of the child’s languages (Bohnacker 2016, pp. 19–20). To overcome these shortcomings, a new picture-based narrative assessment tool was developed during COST Action IS0804 2009–2013 (Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment), the LITMUS-MAIN (Multilingual Assessment Instrument for Narratives, Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters 2012, 2015). The MAIN contains 4 carefully designed, structurally parallel, fictional picture sequences and a standardised protocol to assess production and comprehension of macrostructure (i.e. the global, hierarchical organisation of content in the narrative), for more than 20 languages. The picture sequences (Baby Birds, Baby Goats, Cat,
Dog) depict stories with different protagonists and events but identical overall story and episodic structure, organised around protagonists’ goals. The characters’ actions are clearly depicted, but their internal states must be inferred. The child tells the story, with minimal prompting by the experimenter, and is then asked set comprehension questions. This assessment focuses on the children’s understanding of story events, causality, and the goal-directed behaviour of story protagonists, their thoughts, intentions and emotions (internal states). Indeed, understanding such underlying event relations, narrative schemata, and internal states of story protagonists is generally seen as a prerequisite for being able to produce narrative content in a well-formed manner (cf. Stein & Glenn 1979; Shapiro & Hudson 1991, p. 115; Trabasso & Nickels 1992; Trabasso & Rodkin 1994; Astington & Pelletier 2005, p. 327).

Over the past few years, an increasing number of publications, including an Applied Psycholinguistics special issue 2016 on narrative development in monolingual and bilingual children, have investigated story production with the MAIN. These studies cover a range of languages and language combinations: Finnish, Finnish–Swedish, German–Russian, German–Turkish, Greek, Hebrew–English, Italian–English, Slovak–English and Swedish–English (Kunnari, Välimaa & Laukkanen-Nevala 2016; Gagarina 2016; Maviş, Tunçer, & Gagarina 2016; Tsimpli, Peristeri & Andreou 2016; Altman, Armon-Lotem, Fichman & Walters 2016; Roch, Florit & Levorato 2016; Kapalková, Polišenská, Marková & Fenton 2016; Bohnacker 2016), and they document effects of age and elicitation mode (telling vs. retelling). A general finding has been that macrostructure, as measured on the MAIN, develops with age in relatively similar ways in children across languages, and in line with results from other types of narrative elicitation instruments. While these studies all include measures of narrative production, only few have reported results for narrative comprehension (e.g. Bohnacker 2016; Kapalková et al. 2016; Maviş et al. 2016; Roch et al. 2016; Lindgren 2018), and rarely in any detail.

Yet the assessment of story comprehension is an important complement to story production, because it provides additional opportunities for the child to demonstrate understanding of the main components of story structure (i.e. macrostructure) and the internal states of story characters. Bilingual children with limited proficiency in one language may score low on production measures but, when probed with focused questions, may demonstrate an understanding of narrative schemata and core macrostructure elements (Gagarina et al. 2015, p. 256; Bohnacker 2016). Moreover, some typically developing (TD) children and children with language impairment (LI) may show similar profiles in story production but differ from each other in story comprehension (Bishop & Adams 1992; Liles 1993; Gagarina et al. 2015, p. 249). Comprehension of picture-based or orally presented narratives can also predict reading comprehension (Paris & Paris 2003; Westby 2012).

Children’s performance on story comprehension measures may thus be informative and important, and therefore the present paper focuses on this less explored part of the MAIN. By investigating bilingual and monolingual children with similar socio-economic backgrounds across different ages, we aim to pinpoint what can be expected – and what cannot be expected – of a typically developing child aged 4–6. As the MAIN is increasingly being used around the world, we believe that benchmarks for story comprehension are needed, and the current study is a step in this direction.

The paper proceeds as follows. Section 2 provides some background on story comprehension and inferencing, first more generally and then specifically for the MAIN. In Section 3 (method), we describe our participants, materials, testing procedure, coding and analyses. Section 4 presents the results: a quantitative investigation of overall MAIN comprehension scores for the monolinguals and bilinguals in relation to age, followed by an explorative qualitative analysis of child responses to specific comprehension questions. In
Section 5 we sum up our findings, discuss them in light of earlier work, and offer some conclusions.

2. Background

2.1. Children’s narrative production, comprehension and inferencing

Narrative competence comprises the ability to convey a sequence of actions or events that are temporally and causally related, in a coherent manner and with clear reference (who does what to whom), with “depictions of rich and lifelike characters, whose internal states are represented and coordinated with those of other characters, to create rich and complex social dramas” (Nicolopoulou & Richner 2007, p. 412). Narrative competence thus involves the mastery of both ‘landscape of action’ and ‘landscape of consciousness’ (Bruner 1986, p. 14). This means that children must not only learn to narrate action sequences, but also to convey the thoughts, motives, intentions, emotional reactions and other internal states of story characters that underlie and cause these actions and events to the listener: How does this character feel? Why does s/he carry out an action? What’s her/his goal? How does the character feel in reaction to what’s just happened? How does s/he feel at the end of the plotline and why? In order to interpret these internal (or mental) states, theory of mind (ToM) is needed, since internal states, in contrast to overt actions and physical events, are generally not depicted in the stimulus materials (such as picture books, picture sequences).

Internal states of fictional story characters can therefore not simply be observed or ‘read off’; rather, they must be inferred.

Theory of mind is the awareness of mental states and “the ability to use this awareness in interpreting, explaining and predicting the behavior of one self and others” (Aston 

2005, p. 313). Inferencing is an internal process that integrates information from several sources and results into a conclusion (Levinson 1983, pp. 21–22; Sperber & Wilson 1995). When making inferences about the internal states of others (including fictional story characters), the only evidence of the inference is the child’s verbalisation of it (Lett & Leinonen 2001, p. 308).

Many studies have found that children below the age of 8–9 rarely explicitly mention the internal states of fictional characters when telling stories elicited by picture books, picture sequences or other stimuli, in contrast to older children and adults, who do overtly express these internal states (see for the Frog story *Frog where are you?* (Mayer 1969), e.g., Bamberg & Damrad-Frye 1991; Trabasso, Stein, Rodkin, Munger & Baughn 1992; Berman & Slobin 1994; for other stories, see Stein & Glenn 1979; Kemper 1984, pp. 114–118; Shapiro & Hudson 1991; Ukrainetz, Justice, Kaderavek, Eisenberg, Gillam & Harm 2005, p. 1372; Soodla 2011; Iluz-Cohen & Walters 2012; Tompkins, Guo & Justice 2013, p. 414; Mäkinen 2014; Bohnacker 2016). For this lack of explicit mentions of internal states, different explanations have been advanced, such as deficits in linguistic proficiency (e.g., has the child acquired the lexical items necessary to encode internal states?), or deficits in narrative form, as “young children need to master the formal demands of narrative before they can fully express their interpersonal understanding in their storytelling” (Nicolopoulou & Richner 2007, p. 426). Alternative explanations include complex demands of processing, and/or immature inferencing/theory of mind. For instance, concerning the Frog story, Berman & Slobin (1994) found that whilst children at age 9 were able to attribute inner states to story

---

1 Cartoon-like thought bubbles (Pelletier & Astington 2004) are sometimes added to pictures in an attempt to more directly show certain internal states, with varying degrees of success. Most narrative materials for children do not include thought bubbles. The MAIN does not include them either.
characters and also explicitly did so, 4-to-5-year olds were unable to do so. Berman & Slobin concluded that “evaluative commentary attributing inner states to the protagonists demands a level of inference abstracted from what is shown in the pictures beyond the abilities of the younger children in our sample” (Berman & Slobin 1994, p. 73). Relatedly, van den Broek (1997, p. 339) states: “With age, children may become more adept at inferring internal states, either because they have developed stronger inferential skills or because they have gained familiarity with internal states”.

Some researchers have argued that the late (age 8–9) appearance of explicitly mentioned internal states of fictional story characters may in part be due to methodology. Nicolopoulou & Richner (2007), for instance, hypothesised that often-used narrative elicitation formats with story stems, story topics, wordless picture books or picture sequences may be restrictive and “do not fully capture young children’s abilities (and preferences) in character representation” (p. 415). They speculated that if children were to tell stories initiated and invented by themselves, they might be more inclined or able to express the mental states of fictional characters at a younger age. In a special longitudinal intervention programme designed to foster narrative competence by frequent tell-and-dictate-story-to-teacher and story act-out activities for preschoolers (age 3–5, N=30, TD, English), Nicolopoulou & Richner (2007) found that children’s explicit mentions of mental states increased both over the course of the intervention and with age. Whilst their 3-year-old group virtually never mentioned the mental states of fictional characters, some of the 4-year-olds included rudimentary mentions of mental states. The 5-year-old group expressed the mental states of fictional characters in nearly half of the stories told (47%), especially towards the end of the intervention. Age 5 is considerably earlier than what is commonly observed in other studies (age 8–9, see above). Yet whether this age discrepancy is the result of different elicitation methods or mainly a beneficial effect of participating in a special intervention programme, remains an open question, which is also noted by Nicolopoulou & Richner themselves (2007, p. 427). In any event, Nicolopoulou & Richner’s (2007) study indicates that children already at age 5 are able to construe fictional story characters as mental agents.²

Another way of investigating this issue would have been narrative comprehension, in the form of answers to probe questions that require inferencing on the part of the child. This has been done in some studies in the past, and is also done in the MAIN. When such probe questions are put to children, an age discrepancy emerges: Comprehension of internal states is ahead of production.

For instance, Stein & Glenn (1979) compared story retellings with the answers to story comprehension questions (many of these targeted inferences) by 24 monolingual English-speaking upper middle-class 6- and 10-year-olds in the United States. The children listened to four fictional stories, retold them and answered comprehension questions without any pictorial support.³ The 6-year-olds rarely expressed intentions and motives (goals), feelings and thoughts (internal states) of main story characters in their retellings, but showed relatively

---

² The exact age at which children are able to infer the internal states of others is debated, as it is affected by many factors. Internal states are often seen in relationship with general cognitive development (cf. van den Broek 1997; Tomasello 2003; Astington & Pelletier 2005). Outside the narrative field, social cognition research has shown via experimental false belief tasks (Wimmer & Perner 1983; Perner, Leekam & Wimmer 1987) that 4-year-olds are able to understand both themselves and others as mental agents, i.e., that they can construe the internal mental processes of others and use them to understand, explain and predict actions.

³ Retell/recall without pictorial support increases attentional and verbal memory demands on children and therefore does not purely measure their inferential abilities.
good comprehension of these components when explicitly probed.\textsuperscript{4} Also more recent studies investigate children’s goal inferences in stories they have listened to, via retellings and/or probe questions. Monolingual English children have thus repeatedly been shown to have a basic inferential understanding of fictional story characters’ goals at age 4–6 (e.g. Lynch & van den Broek 2007; Tompkins et al. 2013; for story character’s thoughts, see Pelletier & Astington 2004).

Outside a retell/recall context, fewer studies have explored inferential understanding of story characters’ goals and internal states. Trabasso, Stein, Rodkin, Munger & Baughn (1992) found that 30 monolingual US English middle-class 4-year-olds (3;9–5;0) did not spontaneously produce goals of story characters in their Frog story narratives. When the same children told the story a second time, again with the pictures present, but this time explicitly prompted with \textit{why} questions to explain a character’s action (e.g. \textit{Why is the boy calling? > calling for the frog}), they did express the purpose of the action 73\% of the time. From this, Trabasso et al. (1992) concluded that 4-year-olds can infer fictional character’s goals.

Since few studies outside a retell context have investigated young children’s story comprehension via inferential questions that directly probe goals and emotions, it is difficult to predict from the existing literature what we can expect of a child at age 4–6. Trabasso et al.’s (1992) study suggests that 4-year-olds should be able to infer and verbalise at least some goals of fictional protagonists (in the context of the Frog story), when explicitly prompted/probed. We are not aware of any norm-referenced data in this regard. The above studies indicate that children’s comprehension of narrative structure develops earlier than their ability to verbalise these structures in their narratives. This suggests that research that only measures narrative production may underestimate children’s inferencing abilities of goals and other internal states, which are necessary for narrative competence. The present study therefore focuses on story comprehension as measured by answers to probe questions using the MAIN.

\subsection{2.2. Narrative comprehension in the MAIN}

Only a few studies so far have reported results for the MAIN comprehension questions (Boerma, Leseman, Timmermeister, Wijnen & Blom 2016; Bohnacker 2016; Kapalková et al. 2016; Maviş et al. 2016; Roch et al. 2016; Rodina 2017; Lindgren 2018; cf. also Gagarina et al. 2015, p. 264). The understanding of macrostructural elements (e.g. goals, intentions, emotions and reactions of story characters) appears to be ahead of production, but as yet, very few studies have investigated this in detail.

In the MAIN, a set of ten standardised questions is asked, as exemplified in Table 1 for the Baby Goats story (for the complete materials, see Gagarina et al. 2012). The MAIN explicitly queries goals and other internal states by asking comprehension questions of the child after s/he has told the story. So as not to tax memory unduly, the MAIN pictures are fully visible for both child and experimenter when the comprehension questions are asked. All ten questions require inferencing on the part of the child. Three questions elicit goal statements (D1, D4, D7), e.g. \textit{Why was the mother goat in the water?} (expected answer: \textit{to rescue/help the baby goat}). Another six questions (D2/D3, D5/D6, D8/D9) elicit internal states (IST) connected either to the initiating event or to reactions, e.g. D5 \textit{How does the fox feel?} (expected answer: negative emotion + rationale). When the child does not spontaneously provide an explanation or rationale for her/his answer, a follow-up question is asked, e.g. D6 \textit{Why do you think that the fox is feeling...?}

\footnote{Stein & Glenn’s results (1979, pp. 106–113) are not fully transparent here. Trabasso, Secco & van den Broek (1984) reanalysed Stein & Glenn’s data and found that while the 6-year-olds did include internal states in their story retellings, they did so relatively less often than the older 10-year-olds.}
These questions require the child to make inferences from the pictures (see Fig. 1) and describe how s/he has come to a certain conclusion. The questions thus assess inferencing, i.e. how well the child is able to interpret physical and psychological (emotional, motivational) cause-effect relationships and recognise characters’ goals, the reasons for these goals and reactions following attempts to reach the goals (van den Broek 1997, p. 324; Gagarina et al. 2015, p. 249). D8/D9 moreover probes the child’s understanding of internal reactions in a hypothetical what if scenario (*Imagine that X sees Y. How would X feel? Why would X feel ...?*). The last question, D10, probes whether the child can infer meaning about the story as a whole, e.g. *Who does the mother goat like best, the fox or the bird? Why?* (expected answer: *the bird, because it rescued/helped the goat*). The MAIN comprehension questions were designed to have different levels of difficulty, so a child is not expected to answer them all equally well.

**Table 1. Overview of the** MAIN **comprehension questions, exemplified for Baby Goats.**

<table>
<thead>
<tr>
<th>Question</th>
<th>Example (<em>Baby Goats</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal</td>
<td>Why was the mother goat in the water? (picture 1–2)</td>
</tr>
<tr>
<td>D2. Episode 1 IST as initiating event</td>
<td>How does the baby goat feel? (baby goat in water, picture 1)</td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>Why does the baby goat feel …? [insert answer D2]</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>Why does the fox leap forward? (picture 3)</td>
</tr>
<tr>
<td>D5. Episode 2 IST as reaction</td>
<td>How does the fox feel? (picture 5–6)</td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>Why does the fox feel …? [insert answer D5]</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>Why does the bird bite the fox’s tail? (picture 5)</td>
</tr>
<tr>
<td>D8. Episode 3 IST as reaction</td>
<td>Imagine that the bird sees the goats. How does the bird feel? (picture 6)</td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale</td>
<td>Why would the bird feel …? [insert answer D8]</td>
</tr>
<tr>
<td>D10. Overall plotline question</td>
<td>Who does the mother goat like best, the fox or the crow? Why?</td>
</tr>
</tbody>
</table>

**Figure 1.** The MAIN Baby Goats picture stimuli (originals are 9 x 9 cm coloured pictures in a folded strip).

Answers to the 10 comprehension questions are scored one point each, with a maximum of 10 points per story. This is the overall (or composite) story comprehension score.

It should be acknowledged that what is being assessed here is not only the child’s ability to understand the story plotline from the pictures and make inferences regarding the characters’ internal states, but also the child’s ability to formulate her/his understanding of the story into an answer that is both comprehensible and correct. The MAIN comprehension questions thus probe the child’s ability to verbalise her/his understanding. Answering the questions requires at least a basic level of linguistic production skill in the language of testing. Thus, although the comprehension questions are less demanding than storytelling, they are not necessarily easy for the child.\(^5\)

\(^5\) Since the MAIN assesses the understanding of goals and internal states (i.e. story structure
Studies that have investigated comprehension of the MAIN stories and compared it with production (i.e. telling of the corresponding story), have found a noteworthy gap (Bohnacker 2016; Kapalková et al. 2016; Roch et al. 2016; Lindgren 2018): Children generally reach high overall comprehension scores for narrative macrostructure quite early, i.e. already at a point (or age) when their overall scores for narrative production are still low. Goals are rarely expressed but appear to be comprehended well; internal states as initiating event or reaction are also rarely expressed but appear to be relatively well comprehended overall (Bohnacker 2016, pp. 30–40; Kapalková et al. 2016, p. 155–156; Rodina 2017). This gap is reminiscent of the discrepancy reported in studies using other instruments than the MAIN (as discussed above). Thus, when probed with questions, children aged 4–7 exhibit better understanding of the goals and internal states of story characters on the MAIN than their narrative productions would suggest. Consider for instance the following, quite typical, example (1) of a 6-year-old child telling the Baby Goats story, without any explicit mention of goals and emotions whatsoever. Compare this to the same child’s results on the comprehension questions in (2), where the child demonstrates very good, in fact perfect, understanding of the goals and internal states queried.

(1) Swedish–English bilingual child 6;6 years, telling the MAIN Baby Goats story to a listener who cannot see the pictures:  
*The goats were in the meadow eating grass and the little one fell in. And the big one pushed it up. But the wolf saw them (.) and the wolf came and took the (.) and took one of the wolfs but (.) then the bird came and bited its tail and made him go away.*

(2) The same child’s answers to the comprehension questions, MAIN Baby Goats.

<table>
<thead>
<tr>
<th>Probe question</th>
<th>Child’s answer</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Why was the mother goat in the water? (picture 1–2)</td>
<td>to help the little one up</td>
<td>Correct goal</td>
</tr>
<tr>
<td>D2. How does the baby goat feel? (baby goat in water, picture 1)</td>
<td>scared</td>
<td>Correct IST</td>
</tr>
<tr>
<td>D3. Why does the baby goat feel scared? [answer D2 inserted]</td>
<td>because he can’t swim</td>
<td>Correct rationale for IST</td>
</tr>
<tr>
<td>D4. Why does the fox leap forward? (picture 3)</td>
<td>because he wants to eat the goat</td>
<td>Correct goal</td>
</tr>
<tr>
<td>D5. How does the fox feel? (picture 5–6)</td>
<td>scared</td>
<td>Correct IST</td>
</tr>
<tr>
<td>D6. Why does the fox feel scared? [answer D5 inserted]</td>
<td>because the bird or the eagle did chase it away</td>
<td>Correct rationale for IST</td>
</tr>
<tr>
<td>D7. Why does the bird bite the fox’s tail? (picture 5)</td>
<td>because he was trying to help the goats</td>
<td>Correct goal</td>
</tr>
<tr>
<td>D8. Imagine that the bird sees the goats. How does the bird feel? (picture 6)</td>
<td>good</td>
<td>Correct IST</td>
</tr>
<tr>
<td>D9. Why would the bird feel good? [answer D8 inserted]</td>
<td>because it did help the goats</td>
<td>Correct rationale for IST</td>
</tr>
</tbody>
</table>

**Note.** D10 was not asked.

components that are *not* depicted in the stimuli but that must be inferred), the MAIN author group, when developing the materials and protocol, did not think it feasible to probe story comprehension in a purely nonverbal manner.
In what follows, we investigate bilingual and monolingual children (N=124) with similar socio-economic backgrounds across different ages, to find out what can be expected – and what cannot be expected – of a typically developing child aged 4–6. Comprehension will be studied from different angles: overall comprehension scores, but also the performance on individual questions or component types, and how this develops with age. The performance on individual questions may be more telling and unearth differences in inferencing abilities that are not visible on just an overall composite score. Answers to certain MAIN comprehension questions may prove particularly informative with regard to children’s age development.

3. Method
The participants were 124 children aged 4–7. All told two MAIN stories each and subsequently answered standardised comprehension questions on them. For each narrative, overall story comprehension scores were calculated. Additionally, the responses to individual comprehension questions and story components were analysed. Results were compared across languages and age groups.

3.1. Participants
72 monolingual Swedish children aged 4;0–6;11 and 52 Swedish–English bilinguals aged 4;11–7;8, all growing up in Sweden, were included in the study. The children had no history of hearing loss, language impairment or cognitive disorder.

The monolinguals were recruited from 10 (pre)schools and the bilinguals from 20 (pre)schools in urban areas in central Sweden. As is common in Sweden, the participants had attended preschool from age 1 or 2, mostly full-time. Preschool in Sweden is normally attended until age 5 or 6, after which children go through one year of preparatory schooling (called förskoleklass ‘preschool class’) to prepare them for school proper. At age 7 (sometimes already at age 6), children start grade 1 of primary school. In the current study, all 4- and 5-year-olds attended preschool at the time of data collection. The older children were attending preparatory class or had just started first grade in primary school. All children came from mid- to high-SES backgrounds, where the level of parental education was taken as a proxy for socio-economic status (SES). With the exception of three of the monolingual parents, all parents had completed upper secondary school. All children regularly encountered book reading and storytelling activities.

According to parental report, the bilinguals spoke both English and Swedish at home, but to varying degrees. The majority of parents considered their children to be slightly stronger in Swedish than in English or equally strong in both languages. (For further details on the language background and language use of the bilingual participants, see Bohnacker 2016, pp. 25–26.) All bilinguals had a minimum of two years of regular and frequent exposure to both languages and were able to complete the narrative tasks and answer the comprehension questions in both languages.

Table 2 gives an overview of the participants. There were no significant differences in age between the mono- and bilinguals, neither for the 5-year-olds (t(41) = 1.116, p = .271), nor for the 6-year-olds (t(49.084) = –1.586, p = .119), although the spread in age (SD and age range) was larger in the bilingual 6-year-olds. Since there were no significant differences in age between the mono- and bilinguals, for simplicity, the older bilingual group are referred to as 6-year-olds, even though they also include seven 7-year-olds.

6 Production scores for these children are reported elsewhere (Bohnacker 2016; Lindgren 2018: Ch. 7).
Table 2. Participants.

<table>
<thead>
<tr>
<th></th>
<th><strong>Monolinguals</strong></th>
<th></th>
<th><strong>Bilinguals</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4yrs</td>
<td>5yrs</td>
<td>6yrs</td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>4;5 (0;3)</td>
<td>5;6 (0;3)</td>
<td>6;5 (0;3)</td>
</tr>
<tr>
<td>Age range</td>
<td>4;0–4;10</td>
<td>5;0–5;11</td>
<td>6;0–6;10</td>
</tr>
</tbody>
</table>

The Swedish monolinguals are the same children as the monolingual participants in the second author’s recent PhD thesis on the acquisition of narrative competence (Lindgren 2018); the Swedish–English bilingual participants are the same children as in Bohnacker (2016). The present study compares these groups for the first time and explores new aspects of the children’s narrative abilities, focusing on comprehension.

3.2. Materials

The material used was the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al. 2012, 2015). The Swedish version of the MAIN was used with both the monolinguals and bilinguals, and the English version with the bilinguals only. The MAIN consists of four picture sequences depicting a story with different characters and events but identical overall story and episodic structure. In the current study, all four stories (Cat, Dog, Baby Birds, Baby Goats) were used (see Procedure for details).

In addition to standardised procedures for administering the storytelling tasks and a scoring protocol for macrostructure in narrative production, the MAIN contains ten inferential why and how questions targeting characters’ goals and internal states (recall Table 1). The children’s answers to these comprehension questions constitute the data analysed. Only answers to the first nine questions are included in the current study, as the bilingual 6-year-olds did not receive the tenth question. (The tenth question (D10) was a later addition in the development of the MAIN materials.)

3.3. Procedure

All children were tested in a quiet room at their (pre)schools. Each session started with a warm-up phase in which the experimenters asked some questions or played with the child. The sessions were video- and audio-recorded. Three envelopes (all containing the same story) were placed on the table and the child was told to choose an envelope. The experimenter acted as if the stories were unknown to her. All children first looked at the pictures and told the story, after which the comprehension questions were administered. Children told every story in the telling mode (i.e. there was no model story that the children listened to beforehand, and no retelling). The MAIN protocol (Gagarina et al. 2012) was strictly adhered to. During the storytelling, the experimenter could not see the pictures, but when the comprehension questions were asked, the pictures were visible to both child and experimenter. Note that the comprehension task thus does not probe story recall (memory), but the ability to draw inferences from the pictures concerning the thoughts, intentions and feelings of story characters.

Each child told two stories and answered the comprehension questions on these stories. The monolingual children answered questions on Cat/Dog as well as on Baby Birds/Baby Goats, as part of a larger battery of tasks (see Lindgren 2018) within the same session, where Cat/Dog preceded Baby Birds/Baby Goats. All 72 monolinguals were tested by the second author.

The 52 bilingual children were tested by different research assistants who were native or near-native speakers of the respective language. Each session was conducted entirely in one
language. The bilingual children did not do Cat/Dog, but were tested on both Baby Birds and Baby Goats, in separate sessions. They told Baby Birds or Baby Goats in one language (Swedish) and answered the comprehension questions on that story in the same language (Swedish). In another session, they told the other story in the other language (English) and answered the comprehension questions on that story in that language (English). Half of the bilingual children were tested in Swedish first and the other half in English first. For further details concerning the bilinguals, see Bohnacker (2016).

3.4. Coding and Analysis
Responses to the comprehension questions were transcribed and coded as correct or incorrect. Each child received an overall (composite) comprehension score (max = 9) for each narrative, as well as a score (0 or 1 point) for each of the nine questions.

The Swedish responses of the monolingual children were scored by the second author, who is a native speaker of Swedish. The Swedish and English responses of the bilingual children were at first scored by the research assistants who had tested the children, and later checked by the first author in 2014. Child responses showed a lot of variation though, and since the original MAIN scoring protocol (Gagarina et al. 2012) only covers a limited number of typical cases, the need arose for more comprehensive and homogenised scoring across languages. Detailed scoring guidelines, based on the MAIN manual (Gagarina et al. 2012) and including authentic examples from English, German, Swedish and Turkish, were developed during 2015–2017. These scoring guidelines were based on discussions of a large number of unclear cases from a project that includes the authors of the current study and other (near-)native speakers, all trained linguists well-versed in the MAIN. Members of the original core author group for the MAIN were consulted as well. All unclear cases were discussed until agreement was reached. Finally, in 2017, both authors carefully checked all responses by the monolinguals and bilinguals for consistency and adherence to the scoring guidelines.7

The total number of responses to the comprehension questions to be analysed was 1,764 for Swedish (72 monolinguals x 9 questions x 2 stories, plus 52 bilinguals x 9 questions x 1 story), and 468 for English (52 children x 9 questions x 1 story). In the (relatively few) cases where the experimenter had forgotten to ask a comprehension question, following common practice when there is not more than 2% missing data (Widaman 2006), sample mean substitution was used.8 For the Swedish monolinguals, there was 0.2% missing data (3/1,296). For the bilinguals, there was 1.9% Swedish data missing (9/468) and 1.7% English data missing (8/468).

The following statistical analyses were carried out. For the monolinguals, a paired-samples t-test was used to test whether the children performed differently on Cat/Dog and Baby Birds/Baby Goats, after which two one-way ANOVAs, one for each narrative task, were carried out to test whether there were any differences between the age groups. For the bilinguals (who did not do Cat/Dog but both Baby Birds and Baby Goats), scores in Swedish and English were compared using a paired-samples t-test, and differences between the age groups were tested using independent-samples t-tests, one for the Swedish and one for the English scores. For the 5- and 6-year-olds separately, the bilinguals’ scores in Swedish were

---

7 This process meant that a small number of responses by the Swedish–English bilinguals are scored differently in the present study than originally by Bohnacker (2016).

8 This means that for the missing question, the child was given a score identical to the mean for that particular question for the particular age and language group that the child belonged to. For instance, the monolingual 5-year-old MoSwe5-07 was not asked question D7. This child was given the mean score on D7 for the monolingual 5-year-olds as a substitute D7-score.
compared to the monolinguals’ scores on Baby Birds/Baby Goats using independent-samples t-tests.\(^9\)

In addition to the statistical analyses carried out on the overall comprehension scores, response accuracies for specific comprehension questions were investigated. Questions with low response accuracy were investigated further, by categorising children’s incorrect answers into subtypes, exploring them qualitatively and investigating their development with age. This was done to pinpoint which types of inferential questions on the MAIN are mastered early, and which are mastered late by the children.

4. Results

4.1. Overall comprehension scores

For each of the age groups, the means of the overall comprehension scores are given in Table 3, for Cat/Dog for the Swedish monolinguals, and for Baby Birds/Baby Goats for both mono- and bilinguals. The maximum score was 9.\(^{10}\)

<table>
<thead>
<tr>
<th></th>
<th>Swedish monolinguals</th>
<th>Swedish–English bilinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat/Dog</td>
<td>BB/BG</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>6.9 (2.4)</td>
<td>5.1 (2.1)</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>7.2 (2.0)</td>
<td>6.0 (1.8)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>8.6 (0.9)</td>
<td>7.4 (1.7)</td>
</tr>
</tbody>
</table>

*Note.* Max = 9 points. BB = Baby Birds, BG = Baby Goats. Standard deviations in parentheses.

Concerning comprehension of the different stories, there was no significant difference between the monolinguals’ scores on Cat (M = 7.8, SD = 1.9) and Dog (M = 7.3, SD = 2.1) (t(70) = .953, p = .344). However, there was a significant difference in the monolinguals’ scores for Baby Birds/Baby Goats (t(70) = –2.170, p = .033); the monolinguals performed significantly better on Baby Goats (M = 6.6, SD = 1.9), than on Baby Birds (M = 5.6, SD = 2.2). For the bilinguals in Swedish, there was no difference in scores between the stories (t(50) = –.831, p = .410), the children performed equally well on Baby Birds (M = 6.3, SD = 2.1) and on Baby Goats (M = 6.8, SD = 1.8). However, in English, the bilinguals performed significantly better on Baby Goats (M = 6.7, SD = 2.0) than on Baby Birds (M = 5.5, SD = 1.8) (t(50) = –1.126, p = .038). The difference between Baby Birds and Baby Goats should be investigated further in future studies (cf. Lindgren 2018, Ch. 7).

Interestingly, the (monolingual) children performed significantly better on Cat/Dog than on Baby Birds/Baby Goats comprehension (t(71) = 5.273, p < .001). This is an important and unexpected finding, since the four MAIN stories were designed to be parallel in overall structure and complexity, and story production should therefore be comparable across them (Gagarina et al. 2015, p. 256). The comprehension questions were also designed to be parallel. The MAIN authors assumed that the four stories would then also be equally easy/difficult concerning comprehension, even though this assumption was not empirically tested during materials development (2009–2013). Rather, Cat/Dog were primarily aimed for retelling (Gagarina et al. 2012, 2015, p. 256) and then also used for retelling, and Baby Birds/Baby

---

\(^9\) Independent-samples t-tests were also carried out to test for differences between Cat vs. Dog, and between Baby Birds vs. Baby Goats.

\(^{10}\) Only nine questions (D1–D9) were included in the analysis, as the bilingual 6-year-olds were not asked D10. For scores on all ten questions in the monolingual data, see Lindgren (2018, Ch. 7).
Goats were primarily aimed for and used for telling/story generation (Gagarina et al. 2012). A few studies that have investigated comprehension for both modalities have found that the comprehension scores on Cat/Dog in the retelling mode are higher than the comprehension scores on Baby Birds/Baby Goats in the telling mode (e.g. Maviş et al. 2016, pp. 81–82; Roch et al. 2016, pp. 60–61). Differences in comprehension scores between Cat/Dog and Baby Birds/Baby Goats were then simply attributed to the different story elicitation modes that preceded the comprehension questions. When a child has already listened to the story (as in the retell condition), it is not surprising that s/he might do better on inferential comprehension questions on that story, as the child has already heard the goals and internal states explicitly mentioned in the story script. When comprehension questions are asked after the child has listened to the story, inferencing and verbal memory are thus inextricably intertwined, and the comprehension score may be boosted by what the child has just heard explicitly mentioned. By contrast, in the telling mode, when comprehension questions are administered without the child having heard the story before, the questions truly assess the child’s inferencing abilities.

In the present study, the mean comprehension scores for Cat/Dog are significantly higher than for Baby Birds/Baby Goats, and this holds for 4-, 5- and 6-year-olds alike (Table 3). But crucially, this is not to do with retelling vs. telling. Rather, our results show, for the first time, that comprehension for Cat/Dog is higher than for Baby Birds/Baby Goats even when the stories are all administered in the same mode (telling). So even though the Swedish-speaking children had not listened to Cat/Dog but generated the story from the picture stimuli only, they still performed significantly higher on Cat/Dog than on Baby Birds/Baby Goats. This suggests that comprehension on Cat/Dog is in fact easier than on Baby Birds/Baby Goats.

Let us now move on to comprehension scores and age. There was a significant difference in scores between the monolingual age groups for Cat/Dog (F(2, 69) = 5.696, p = .005). Subsequent post-hoc tests (Bonferroni) showed that the 6-year-olds scored significantly higher than both 4-year-olds (p = .007) and 5-year-olds (p = .038), but that there was no difference between 4- and 5-year-olds (p = 1.00).

Significant differences between the monolingual age groups were not only found for Cat/Dog, but also for Baby Birds/Baby Goats (F(2, 69) = 9.343, p < .001). Again, the 6-year-olds performed better than both the 4- (p < .001) and 5-year-olds (p = .031), but with no difference between the two younger groups (p = .314).

Let us now compare the monolinguals to the bilingual children for comprehension on Baby Birds/Baby Goats. There was no difference between the monolinguals’ scores and the bilinguals’ Swedish scores, neither for the 5-year-olds (t(41) = −.808, p = .424) nor for the 6-year-olds (t(55) = 1.700, p = .095). Interestingly though, the bilinguals did not show exactly the same development with age as the monolinguals. As mentioned above, there were significant differences between the monolingual age groups for Baby Birds/Baby Goats, with the 6-year-olds performing better than both the 4- and 5-year-olds. However, in neither language was there any significant difference between the bilingual age groups in comprehension scores (Swedish: t(50) = −.139, p = .890; English: t(50) = −1.949, p = .057). Paired samples t-tests also showed no difference between the bilinguals’ scores in Swedish and English, neither for the 5-year-olds (t(18) = 1.434, p = .169), nor for the 6-year-olds (t(32) = −.667, p = .510).

---

11 Retelling means that the child first listens to the story script (e.g. via headphones) and then (re)tells the story to an experimenter. Telling/story generation means that the child is not exposed to such a model but generates the story by him/herself from the pictures only, without the benefit of having listened to a prior script (Gagarina et al. 2015, pp. 250–251).

12 In some cases, these differences were significant and sometimes they were not.
In general, the children’s overall comprehension scores of narrative macrostructure are relatively high, indicating a good general understanding of the stories.\(^\text{13}\)

### 4.2. Performance on specific comprehension questions

As the comprehension score on the MAIN is a composite score, it may hide interesting differences in the performance on different types of comprehension questions. The present section explores this further. Tables 4 and 5 show response accuracies (percent correct) for each individual comprehension question. Percent correct refers to the percentage of children who answered the respective question correctly. Response accuracies for Cat/Dog (only available for the monolinguals, see Table 4) are generally somewhat higher than for Baby Birds/Baby Goats (Table 5), confirming what we already know from the significant differences in the means of the overall comprehension scores between these story pairs (Table 3).

#### Table 4. MAIN Cat/Dog, comprehension, percent correct (%) answers on the 9 individual questions, Swedish monolinguals aged 4–6.

<table>
<thead>
<tr>
<th>Question</th>
<th>Monolinguals (N = 72)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal</td>
<td>85</td>
</tr>
<tr>
<td>D2. Episode 1 IST as initiating event</td>
<td>81</td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>78</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>86</td>
</tr>
<tr>
<td>D5. Episode 2 IST as reaction</td>
<td>92</td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>82</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>93</td>
</tr>
<tr>
<td>D8. Episode 3 IST as reaction</td>
<td>82</td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale</td>
<td>78</td>
</tr>
</tbody>
</table>

#### Table 5. MAIN Baby Birds/Baby Goats comprehension, percent correct (%) answers on the 9 individual questions, Swedish monolinguals aged 4–6 and Swedish–English bilinguals aged 5–6.

<table>
<thead>
<tr>
<th>Question</th>
<th>Monolinguals (N = 72)</th>
<th>Bilinguals–Swe (N = 52)</th>
<th>Bilinguals–Eng (N = 52)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal</td>
<td>75</td>
<td>88</td>
<td>71</td>
</tr>
<tr>
<td>D2. Episode 1 IST as initiating event</td>
<td>74</td>
<td>67</td>
<td>83</td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>63</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>89</td>
<td>94</td>
<td>96</td>
</tr>
<tr>
<td>D5. Episode 2 IST as reaction</td>
<td>94</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>86</td>
<td>84</td>
<td>76</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>64</td>
<td>79</td>
<td>81</td>
</tr>
<tr>
<td>D8. Episode 3 IST as reaction</td>
<td>42</td>
<td>50</td>
<td>46</td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale</td>
<td>26</td>
<td>37</td>
<td>37</td>
</tr>
</tbody>
</table>

\(^{13}\) This is at a point when their overall scores for story structure production are still low (see Bohnacker (2016) for the Swedish–English bilinguals, and Lindgren (2018) for the Swedish monolinguals). For instance, on Baby Birds/Baby Goats, the 5-year-old monolinguals scored on average 6 out of 17 points for production (Lindgren 2018, p. 185), and the 5-year-old bilinguals scored 5 out of 17 (Bohnacker 2016, pp. 28–29).
Interestingly, we see similar response accuracies for monolinguals and bilinguals, and also very similar response accuracies for the two languages of the bilinguals, for the individual comprehension questions on Baby Birds/Baby Goats (Table 5).

Concerning questions that probe the comprehension of goals of story characters (D1, D4, D7), children’s responses are mostly correct, and this holds for both Cat/Dog and Baby Birds/Baby Goats, and for both monolinguals and bilinguals irrespective of language.

To investigate this further, we have broken down the responses to a particular question targeting understanding of goals for the different age groups (4-, 5- and 6-years), and illustrate this for question D4, episode 2 of Baby Birds/Baby Goats (e.g. Why does the fox leap forward? (expected answer: because he wants to catch/eat the goat)). As shown in Fig. 2, comprehension of goals on D4 is very good already at age 4 and 5. A full 79% of the 4-year-olds answer this question correctly, and accuracy is close to ceiling for the 5- and 6-year-olds for both monolinguals and bilinguals (92%–96%). For English, the corresponding figures are 100% correct for the bilingual 5-year-olds and 94% correct for the bilingual 6-year-olds (not shown in Fig. 2). Percentages correct are similar for the other comprehension questions targeting goals (cf. Tables 4–5). Comprehension of protagonists’ goals, as measured on the MAIN, can thus be said to be fully acquired by age 5.

Yet response accuracies are not at the same (high) level for every comprehension question. For internal states, results are more mixed. In fact, for Baby Birds/Baby Goats, response
accuracy is generally very low on the internal state as reaction questions (D8/D9) asked in connection with the final picture of the story. D8: Imagine that the bird sees the goats. [point to picture 6] How would the bird feel?. D9: Why would the bird feel …? [insert expression used by child in response to D8]. As Table 5 shows, only 42%–50% of the children’s answers to D8 and only 26%–37% of the answers to D9 are correct. This is much lower than the near-ceiling performance concerning comprehension of goals.

Figure 3 shows these very low response accuracies for D8/D9 in Baby Birds/Baby Goats for the different age groups separately. The expected answer is a positive emotion, e.g. the crow/dog feels good, proud, like a hero, relieved, pleased, satisfied, glad, happy because s/he helped/ saved the goats/birds, because the goats are safe now. Very few 4-year-olds (8%) and 5-year-olds (13%, 26%) answer these questions in the expected way in Swedish, and only about half of the 6-year-olds do so (58%, 42%). Results are equally low for the bilinguals’ answers in English, with 11% correct answers for the 5-year-olds and 52% correct answers for the 6-year-olds (not shown in Fig. 3). Accuracy improves with age for both monolinguals and bilinguals, as indicated by Fig. 3, but response accuracy at age 6 is still low.

Figure 3. Comprehension Baby Birds/Baby Goats Episode 3 internal state as reaction + rationale, percent (%) correct answers to question D9, Swedish, by age group.

For D8/D9, the majority of the 4–6-year-old monolingual Swedish and bilingual Swedish–English children thus do not give the expected answer. So what do they answer instead? The incorrect answers on question D8 can be classified into a handful of subtypes. The proportions
of these subtypes are reported in Table 6. What kind of answers are these? And what do they tell us about the children’s story comprehension?

Table 6. Answers to Baby Birds/Baby Goats question D8, episode 3 internal state as reaction broken down by subtype, in percent (%).

<table>
<thead>
<tr>
<th></th>
<th>Swedish monolinguals</th>
<th>Swe.–Eng. bilinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4yrs (n=24)</td>
<td>5yrs (n=24)</td>
</tr>
<tr>
<td>Correct answer</td>
<td>29.2</td>
<td>29.2</td>
</tr>
<tr>
<td>Incorrect answer</td>
<td>70.8</td>
<td>70.8</td>
</tr>
<tr>
<td>- Clearly negative emotion (angry/cross, sad, not good/happy)</td>
<td>45.8</td>
<td>54.2</td>
</tr>
<tr>
<td>- Other emotion (funny, confused, cannot feel, a bit stressed)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>- Physical state (hungry/wanting to eat)</td>
<td>8.3</td>
<td>4.2</td>
</tr>
<tr>
<td>- Action (fly and bite, should chase, take the birds, flying up in the tree)</td>
<td>12.5</td>
<td>4.2</td>
</tr>
<tr>
<td>- ‘I don’t know’ (incl. no answer)</td>
<td>4.2</td>
<td>8.3</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note. n = number of answers. For the bilinguals, Swedish and English answers are combined.

Question D8, Baby Birds: Imagine that the dog sees the birds. How would the dog feel?

Question D8, Baby Goats: Imagine that the bird sees the goats. How would the bird feel?

Correct answer: Good, proud, like a hero.

Most of the children in the present study who are not able to answer D8 correctly answer D8 with a clearly negative emotional state term, saying that the dog/crow feels angry, cross, not good, not happy etc. This type of answer is particularly common among the younger children (4- and 5-year-olds: 34%–54%). Some, but fewer, children answer with a physical internal state term (saying that the dog/crow is hungry), or an unexpected emotion (e.g. saying that the dog/crow feels funny or confused), or even answer with an action (e.g. biting or flying up the tree). Roughly ten percent of the older children, particularly the bilinguals, answer that they do not know what the character would feel. Aside from these I don’t know answers, the majority of the incorrect answers suggest that the children home in on the character’s facial and bodily depictions in one individual picture and infer internal states from them.14

Figure 4 shows the final picture (picture 6) of Baby Birds and Baby Goats, in which the dog/bird (i.e. the ‘saviour’ or helper/hero) is chasing away the cat/fox (i.e. the ‘baddie’).

---

14 As the bilinguals’ incorrect answers on D8 in Swedish and English were similar in type, they were combined for both languages in Table 6. The bilinguals’ incorrect answers break down as follows. 5-years (Swe. n=11, Eng. n=14): clearly negative emotions Swe. 21.1%, Eng. 47.4%; other emotions Swe. 0%, Eng. 5.3%; physical state (hunger) Swe. 10.5%, Eng. 5.3%; actions Swe. 5.3%, Eng. 0%; don’t know Swe. 21.1%, Eng. 15.8%. 6-years (Swe. n=15, Eng. n=14): clearly negative emotions Swe. 15.2%, Eng. 21.2%; other emotions Swe. 9.1%, Eng. 9.1%; physical state Swe. 6.1%, Eng. 0%; actions Swe. 3.0%, Eng. 3.0%; don’t know Swe. 12.1%, Eng. 9.1%.
When the children answer D8 with a negative emotion term (e.g. angry, not good, not happy; Swe. arg ‘angry’), which is the most frequent type of incorrect answer for all age groups in our data (12.5%–54.2%), they focus on what is graphically shown to be happening in the final picture and verbalise this content as a negative emotion. For instance, the dog/bird has a fierce/angry expression when chasing the cat/fox away, so therefore children may make the inference that the dog/bird would feel angry. Such an answer indicates however that the children do not attend to the preamble to the question (i.e. Imagine that the dog [point to dog] sees the birds [point to birds]. How would the dog feel?). Nor do they take the whole plotline of the story into account when answering. They appear to have forgotten that the dog/bird is acting as the saviour of the babies and therefore, when seeing the birds/goats at the end of the story, should presumably feel happy about the result of its intervention (cf. Bohnacker 2016).

Other types of incorrect answers on D8 indicate that the child has actually misunderstood the role of the dog/crow in the story. This is the case, for instance, when the child answers that the dog/crow is hungry, wants to eat up the cat or the baby birds, or should take the baby birds/goats. Such answers are less common (see Table 6), but their existence suggests that some children have not understood that the dog/crow is helping, protecting or rescuing the babies. If only considering the angry/fierce look of the dog/crow in the last picture, the child may erroneously infer that the dog/crow is hungry and may want to eat the family. These results show that many children may still have difficulties with inferring internal states as reactions beyond the purely physical and clearly depicted, even at age 6.

One might think that the low response accuracy on D8/D9 in Baby Birds/Baby Goats is linked to difficulties with understanding a question about a hypothetical scenario. Recall that the prompt asks the child to imagine that the dog/crow sees the birds/goats and then to tell the experimenter how the dog/crow would feel in this hypothetical scenario, i.e. if the dog/crow were to turn around to face and see the (bird/goat) family.

Here, a comparison with the MAIN Cat/Dog stories is illuminating, as they also contain questions about a hypothetical scenario (picture 6), D8: Imagine that the boy sees the cat/dog. How does/would the boy feel? (expected answer: negative emotion, e.g. bad/angry/mad/sad), and D9: Why do you think the boy would feel …? [insert expression used by child in response to D8] (expected answer: rationale for the negative emotion (bad/angry/mad/sad) because the cat/dog has taken/is eating the boy’s fish/sausages). Interestingly, the overall response accuracy for D9 on Cat/Dog in our data is 78%, as was shown in Table 4 (monolinguals age 4–6, N=72). This is three times higher than the response accuracy for D9 on Baby Birds/Baby Goats, which, as will be recalled, was 26% (Table 5, monolinguals aged 4–6, monolinguals age 4–6, N=72).

---

15 Results are reported for Swedish monolinguals only, as the bilinguals were not tested on Cat/Dog.
Fig. 5 shows that the accuracy on D9 for Cat/Dog is already very high for the 4- and 5-year-olds (67%, 71%), and that by age 6, the monolingual children are at ceiling (96%).

![Figure 5](image)

**Figure 5.** Comprehension Cat/Dog Episode 3 internal state as reaction + rationale, percent (%) correct answers to question D9, Swedish, monolinguals.

Thus, the more or less identically worded inferential comprehension question about a hypothetical scenario is much easier in Cat/Dog than in Baby Birds/Baby Goats. Only 16 out of 72 monolingual children (7 four-year-olds, 8 five-year-olds, 1 six-year-old) did not answer D8+D9 correctly on Cat/Dog. Out of these 16, only two children answered D8 with a positive emotion (Swe. glad ‘happy’). Four children said they did not know what the boy would feel, and four children failed to give an adequate explanation for the boy’s bad feeling. Six children gave an action instead of an emotion in answer to D8.

Why, then, is performance so much better on Cat/Dog for similar comprehension questions (D8/D9) concerning a hypothetical scenario? Fig. 6 shows the final pictures for Cat and Dog; compare these to the final pictures for Baby Birds/Baby Goats in Fig. 4.

The expected answer to the inferential question D8 (*Imagine that X sees Y. How does/would X feel?*) on Cat/Dog is a negative emotion, whereas for Baby Birds/Baby Goats, it is a positive emotion. There is however another, and in our opinion, crucial difference between Cat/Dog and Baby Birds/Baby Goats concerning D8. In Cat/Dog, the final picture shows what the cat/dog is doing (i.e. eating fish/sausages), so the child can infer the targeted internal state, i.e. the boy’s reaction, from that very picture, by imagining that the boy turns around and sees the cat/dog eating the fish/sausages. The child only needs to know that the

N=72).
fish/sausages belong to the boy. In order to score correct on D8 in Cat/Dog, it is thus enough to understand the final picture, the child does not have to take into account the other episodes and the whole plotline.

**Figure 6.** The final picture (picture 6) in the MAIN Cat (left) and Dog (right).

By contrast, for Baby Birds/Baby Goats, the child cannot score correct on D8 if s/he only considers the final picture without paying attention to and correctly interpreting the entire plotline, where the cat/fox is the baddie, wanting to catch/eat a baby bird/goat, and where the dog/crow thwarts this attempt by intervening and rescuing the baby bird(s)/goat. We suggest that this underlying difference between Cat/Dog and Baby Birds/Baby Goats is the reason for the considerable differences found in children’s response accuracies on seemingly identical inferential questions (D8/D9).

5. **Discussion and conclusions**

This study has examined how 72 monolingual Swedish and 52 bilingual Swedish–English children aged 4–6, growing up with similar socio-economic backgrounds, develop their understanding of story characters’ goals and emotions. Story comprehension was investigated by analysing the children’s answers to standardised inferential probe questions as part of the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al. 2012). We have focused on this somewhat neglected part of the MAIN, since the assessment of story comprehension is an important complement to story production: It provides the child with additional opportunities to demonstrate understanding of major components of story structure and of the internal states of story characters, even when scores on production measures may still be low. We set out with the aim of providing an (at least provisional) answer to the question of what can be expected – and what cannot be expected – of a typically developing child aged 4–6 for comprehension on the MAIN.

The results have shown that the overall comprehension scores are relatively high for our participants already at age 4, and certainly at age 6, indicating a good general understanding of the stories. The 4–6-year-olds were able to answer the majority of the MAIN comprehension questions correctly. However, only for some groups was there any significant development with age (i.e. improvement) of the overall scores over this age period; this may well be to do with the fact that comprehension scores are generally already quite high. Another possibility is that some particular aspects of inferential understanding, as probed by the MAIN, do not fully develop until after age 6 (see also below).
An interesting and new finding were the significant differences between the overall scores for Cat/Dog vs. for Baby Birds/Baby Goats, despite the fact that all stories had been administered in exactly the same mode (i.e. telling). For instance, the 4-year-old Swedish monolinguals answered on average 7 out of 9 comprehension questions correctly on Cat/Dog, but only 5 out of 9 correctly on Baby Birds/Baby Goats. At age 6, there was still a significant difference between the story pairs. This suggests that comprehension on Cat/Dog is in fact easier than on Baby Birds/Baby Goats. The differences between Cat/Dog and Baby Birds/Baby Goats should be investigated further for other languages.

When exploring the children’s performance on different types of comprehension questions, another finding emerged: Response accuracy on certain questions was more or less equivalent for Cat/Dog and Baby Birds/Baby Goats, whereas on some other questions (especially D8/D9), it was much lower for Baby Birds/Goats than for Cat/Dog. Thus, the overall comprehension scores smooth over or in fact mask telltale differences in the performance on specific questions, particularly so for Baby Birds/Baby Goats (and less so for Cat/Dog, where response accuracy is generally quite high).

Questions targeting protagonists’ goals were understood very well already at age 4, for both Cat/Dog and Baby Birds/Baby Goats. By age 5 and 6, understanding of goals was close to ceiling for both tasks (92%–96%). The bilinguals did not differ from monolinguals in this regard, and the bilinguals understood goals equally well in both their languages. The children possess the necessary cognitive skills to make appropriate goal inferences, i.e. to fill in non-explicit information, as well as the linguistic skills to verbalise them. Thus, one of our conclusions was that goals on this task can be expected to be mastered by typically developing children by age 4–5, irrespective of story. The children’s good goal comprehension is particularly interesting in light of the fact that they, as well as other children aged 4–6, very rarely overtly express goals when telling the MAIN stories (Bohnacker 2016, pp. 30–34; Lindgren 2018, Ch. 7). Our results thus confirm those of other studies that also document good comprehension of story protagonists’ goals when probed, despite rare explicit mention of goals in story production; this holds both for the MAIN (Kapalková et al. 2016, pp. 155–156; Rodina 2017), as well as earlier studies using different materials (e.g. Stein & Glenn 1979; Trabasso et al. 1992; Lynch & van den Broek 2007).

Questions targeting the understanding of other internal states of protagonists, including their emotions as reactions to an event, presented a somewhat more mixed picture. Whilst response accuracy on these questions was again generally high for Cat/Dog, for Baby Birds/Baby Goats, response accuracy, especially on D8/D9, was strikingly low. We argued that it is unlikely that these differences are due to the way the questions are formulated or to the fact that a hypothetical scenario is queried, as both of these aspects are identical for Cat/Dog and Baby Birds/Baby Goats. Instead, we have suggested that the differences in response accuracies between the two story pairs have to do with whether the necessary inference involves only one picture (short-range) or the whole plot and sequence of six pictures (long-range).

The majority of children in the present study did not correctly answer D8/D9 Baby Birds/Baby Goats, and of the younger children, hardly anyone did (monolingual 4-year-olds: 8%, monolingual 5-year-olds 13%, bilingual 5-year-olds 26%). A qualitative analysis of the incorrect responses on D8/D9 revealed that most children homed in on the facial and bodily depictions of protagonists in an individual picture (e.g. a fierce and angry-looking dog) and inferred feelings, thoughts and emotions from these (e.g. being angry). Only some children (mainly older ones) went beyond such short-range interpretations and gave a correct answer, taking into account the whole plotline to infer the internal state of the protagonist.

Importantly, these results held for both monolinguals and bilinguals alike, and across languages (Swedish, English), suggesting that they may generalise.
Even though response accuracies on D8/D9 at age 6 were higher than at age 4 and 5, many children still had difficulties in inferring internal states as reactions beyond the purely physical and clearly depicted. They processed information locally rather than globally, i.e. within the picture rather than across the entire picture sequence.\textsuperscript{16}

We therefore suggest that the processing abilities of the average 6-year-old are not yet developed enough to be able to master every aspect of inferential understanding probed on the MAIN. Note that this is not meant to imply that the children lack Theory of mind or are unable to make inferences. Theory of mind is not the issue here. In fact, the responses showed that our participants, including the 4-year-olds, were able to construe fictional story characters as mental agents: they were able to put themselves into a fictional character’s shoes and infer their internal state. The present results are in line with earlier narrative studies of monolingual English-speaking 4–5-year-olds that employed other methods and materials (e.g. Pelletier & Astington 2004; Nicolopoulou & Richner 2007; Tompkins et al. 2013), but also go beyond them: When the inference necessitated going beyond facial and bodily expression of the character, or beyond the event(s) depicted in only one or two pictures, many children in the present study ran into problems and answered in unexpected ways. Their behaviour is reminiscent of what van den Broek (1997), in the context of story recall, describes as a gradual developmental “shift from focus on within-episode connections only to a focus on between-episode connections” (p. 335), which may not be completed until age 10–11.

On the basis of these results, what, then, can be expected of a typically developing child aged 4–6 for comprehension on the MAIN?

- \textit{Overall comprehension scores} on Cat/Dog should be high already at 4 years, and very high at 6 years, at least 7 out of 9 points (in our study, mean scores were 8.6 out of 9 for the 6-year-old Swedish monolinguals, corresponding to 95% correct). By contrast, overall comprehension scores on Baby Birds/Baby Goats cannot be expected to be as high as on Cat/Dog, neither at age 4 nor at age 6.

\textsuperscript{16} Bohnacker (2016, p. 40) makes a related argument concerning short-range inferences for a different comprehension question, D5/D6, which queries the internal state of the “baddie” cat/fox as reaction in the final picture in Baby Birds/Baby Goats. Here, a negative internal state is expected (D5: \textit{How does the cat/fox feel?} > bad, sad, angry, mad, scared, hurt, still hungry, disappointed). 92% of the Swedish–English bilingual children also answered that way; even though the younger ones tended to do so in very general terms (bad, or not good). When probed further (D6: \textit{Why do you think that X feels... ?}), nearly all 5-year-olds related the cat’s/fox’s feeling bad to it being physically hurt (e.g. \textit{its tail hurts because it is bitten by the dog/crow}). They thus explained their choice of internal reaction from within the picture (short-range). No 5-year-old mentioned that the cat/fox might feel bad, disappointed, or still hungry because its attempt at killing its prey had been thwarted. Most 6-year-olds described the internal state of the cat/fox as scared, less commonly as bad or not good. In contrast to the younger children, few of them attributed feeling bad to the cat/fox being physically in pain; rather, they explained that it felt scared because it was being chased or bitten. Despite this shift from a physical state (pain, hurt) to an emotional state (scared), the 6-year-olds still gave short-range explanations, as they focused on what the dog/crow physically did to the cat/fox and how this made it feel. Very few 6-year-olds (4 out of 33) gave long-range explanations that related back to earlier events in the story: Two children said that the fox felt disappointed and hungry because it did not get any food, and two stated that the fox felt bad because he was ashamed or full of remorse for having attacked the goat. These were very unusual answers, and their rarity suggests that the average 6-year-old is far more likely to verbalise internal states of protagonists that are depicted in individual pictures than internal states that only can be deduced by taking into account the entire plotline (Bohnacker 2016, p. 40).
Comprehension of *goals*, on any of the MAIN stories, should be high already at 4 years, presumably substantially above 50% (in our study, 79% at age 4 for the monolinguals). Comprehension of goals at age 5 and 6 can be expected to be very high, both at group level but also for individual children (in our study, both monolingual and bilingual 5- and 6-year-olds performed close to ceiling).

Response accuracies on D8/D9, which target the *internal state as reaction* of the protagonist in episode 3, cannot be expected to be high for Baby Birds/Baby Goats at age 4–6. In fact, the average 4- or 5-year-old cannot be expected to be able to answer D8/D9 correctly at all (in our study, only 8% of the 4-year-olds and 13% of the 5-year-olds were able to), and even at age 6, only around half of the children could do so. In contrast, response accuracies on D8/D9 for Cat/Dog can be expected to be relatively high already at age 4 and 5 (in our study, 67%–71% of the 4- and 5-year-olds answered correctly), and to go up to very high levels by age 6 (in our study, 96% of the monolinguals answered correctly.)

We are aware that the above proposal for what can be expected of typically developing children’s comprehension performance on the MAIN must be considered provisional. Even though the number of children studied (N=124) is not small, only two languages (Swedish and English) were included, so further studies should look at different languages and language combinations. For Cat/Dog, we have only reported data from monolinguals; here bilingual children would need to be studied (see Lindgren 2018 for a comparison of Swedish monolinguals with Swedish–German and Swedish–Turkish 4–6-year-olds, and Öztekin (in progress) for Swedish–Turkish 4–7-year-olds, which go in the same direction).

Another limitation of the present study is the fact that all children came from mid- to high-SES backgrounds, as measured in parental levels of education. It is likely that their SES background has been beneficial to the development of the children’s language and narrative abilities, as all families reported regular book-reading and storytelling. Moreover, in Sweden, where preschool is widely available and affordable for everyone irrespective of family income, most children attend preschool full-time from an early age (1 or 2 years), and are likely to be exposed to frequent book-reading and storytelling activities there. In the Swedish context, preschool may act as a leveller of SES-differences that in other contexts might impact strongly on children’s language and narrative development. It therefore remains to be seen whether children growing up in other contexts show the same response patterns as the children in the present study.

To conclude, we hope to have shown that a more detailed investigation of children’s performance on the comprehension part of the MAIN can be highly informative. There is development with age for comprehension, but it may occur at different ages for different types of questions, and not all of it shows in the overall composite comprehension score. Since several studies, including the present one, have found that the overall comprehension scores are generally quite high already at age 4–5 (especially when compared to production scores on the MAIN), the composite score of an individual child may not always be the most informative for assessment. When the child is already at an age where group means are high and do not increase much with age, the child’s performance on certain selected comprehension questions may tell us more with regard to age-adequate development.

References


