Labelling Improves False Belief Understanding. A Training Study

Elisabet Serrat Sellabona, Carles Rostan Sánchez, Eduard Vallès Majoral, Moisès Esteban Guitart, Francesc Sidera Caballero and Jèssica Serrano Ortiz

Universidad de Girona (Spain)

Abstract. A total of 104 children aged between 41 and 47 months were selected to study the relationship between language and false belief understanding. Participants were assigned to four different training conditions: discourse, labelling, control (all with deceptive objects), and sentential complements (involving non-deceptive objects). Post-test results showed an improvement in children's false belief understanding in the discourse and the labelling conditions, but not in the sentential complements with non-deceptive objects or the control group. Furthermore, the most remarkable improvement in false belief understanding occurred in the labelling group. These results suggest that some types of linguistic experience promote the development of false belief understanding, provided that differing perspectives are confronted.

Theory of mind (ToM) refers to the ability to represent, reason about and conceptualize mental states (Malle, 2002). It is a fundamental factor in explaining human behaviour. However, human beings are not born with a fully-developed theory of mind system; it develops progressively over time.

Initial studies of the development of ToM supported the premise that there is a developmental milestone, at around four or five years of age, when false belief understanding is acquired. Nevertheless, the notion of ToM has expanded and it is now considered to be a series of processes leading to an increasingly sophisticated understanding of human behaviour. Against the backdrop of this broader concept of ToM development, the present study assumes that, despite the ongoing development of ToM abilities, the acquisition of false belief understanding is still a landmark. Once they have developed false belief understanding, children understand that different people represent the world in different ways, and may therefore have a variety of beliefs or thoughts about the same reality (Malle, 2002).

To solve false belief understanding tasks, it must be understood that what different people know or believe about the world might differ. Understanding false belief implies realizing how people construct knowledge, that people's beliefs are based on their knowledge, that those beliefs can differ from reality and also that people's behaviour can be predicted from beliefs or mental states (Malle, 2002). The development of false belief understanding is therefore crucial for the development of social cognition.

On the other hand, in this study we also assume that language can influence false belief understanding. Although several authors (e.g. Astington & Baird, 2005; de Villiers, 2005) maintain that language influences the development of false belief understanding, others defend a bidirectional relationship (Slade & Ruffman, 2005). It is generally agreed, however, that some aspects of such a complex question require further study (Milligan, Astington, & Dack, 2007). In the present work we adopt a specific approach to the influence of language on ToM, focusing on the development of false belief understanding and considering the fundamental role language plays in it.

Various proposals have been made regarding the different aspects of language that may influence ToM. The semantic-pragmatic or communicative approach emphasizes the role of the language employed to address children in the development of ToM. Another approach stresses the effects of learning certain aspects of grammar on ToM development. And additional data show that other aspects of language may also play a role in the development of ToM. Therefore, a summary of these proposals is provided.

The semantic-pragmatic approach focuses on the effects of the input directed at children. Within this approach, there are two principal lines of research: (a) Some authors consider that the effects of the input are due to the process of conversational interaction itself, in that it allows different viewpoints to be shared; (b) Other authors hold that ToM development is promoted by the effects of the significance of certain words. The first group maintains that children are capable of the competent use of psychological terms during social interactions mediated by language (Cutting & Dunn, 2006; Harris, 2005; Tomasello & Carpenter, 2007). Considering the large number of studies and the quantity of data collected along these lines, it seems clear that conversational interactions play a critical role in understanding people's mental worlds. In any case, it is still necessary to define what type of conversational input is most appropriate to stimulate the emergence of ToM, or what aspects of the conversational input favour this emergence. Most of the studies conducted have been correlational and it is thus difficult to determine whether the type of conversational input promotes the development of ToM, or whether the child's level of socio-cognitive understanding motivates others to express themselves through a more mentalistic discourse. It is important to consider, however, that the study by Lohmann and Tomasello (2003) shows that exposing children to interactive situations in which different points of view on the same subject are emphasized is sufficient to improve their performance on theory of mind tasks.

Other authors who share this communicative approach have focused specifically on the semantic aspects of conversational input. From this point of view, conversations which abound in references to mental content facilitate the abstraction of mental state concepts (Bartsch & Wellman, 1995; Olson, 1988). According to this perspective, the acquisition of terms that refer to mental states is important for developing an understanding of the mind, since it makes it possible to draw analogies between one's own experience and that of others. Language creates categories of expression to distinguish concepts such as beliefs and desires with a high degree of referential opacity. However, the above mentioned study by Lohmann and Tomasello (2003) shows that it is not necessary to use mental-state language to facilitate the understanding of false beliefs, and that a perspective-shifting discourse could be enough.

A second theoretical proposal assumes that the linguistic and grammatical abilities of children are critical for the development of false belief understanding. Syntactic abilities began to be thought of as closely associated with false belief understanding following the work of Astington and Jenkins (1999). Since then, a much more specific hypothesis has been proposed (de Villiers, 2005; de Villiers & de Villiers, 2000; de Villiers & Pyers, 2002), maintaining that the key factor when discussing false beliefs resides in the use by children of sentential complement syntax. Two important aspects of this hypothesis must be emphasized. First, it is not applied to all sentential complements, only to those that depend on mental and communication verbs that take complements about potentially real events, also called *realis* complements. Unlike *realis* complements, which can be true or false, *irrealis* complements, characteristic of other mental verbs such as *promise* or *want*, deal with future or hypothetical events, and a specific truth value cannot be attributed to them. Second, it has been postulated that children first master sentential complements that depend on verbs of communication, where there can be explicit evidence that the complement is false, and that only afterwards, by analogy, do they understand the sentential complements embedded in mental verbs (for a detailed explanation, see de Villiers, 2005).

Sentential complement sentences are complex sentences consisting of a main clause, into which is inserted another sentence (or complement clause), which acts as the grammatical object of the main clause. The type of sentential complement sentences on which de Villiers's hypothesis (2005) is based is illustrated here:

(1) *Mike says that the truck is yellow.*
(2) *Sean thought that Charles had broken his truck.*

Only this type of complement clause, found in sentences with verbs like *think* or *say* as a main verb, can be false even though the main clause which contains it is true. Sentence 1 may be true, regardless of the veracity of the complement clause. The same occurs in example 2. If the complement clause "*that Charles had broken his truck*" codifies the mind of Sean, the truthfulness of this proposition cannot be assessed by comparing it to the real world, but only to Sean's mental world. This property has led some authors to assume that sentential complementation provides the necessary format to represent false beliefs.

In this context, the study conducted by de Villiers and Pyers (2002) is fundamental. In it, a series of linguistic and false belief tasks was administered to children between 3 and 5 years of age on four different occasions. The authors found that children's performance in false belief tasks was not related to their general linguistic ability measured through mean length of utterance (MLU) or other syntactical measures, but was exclusively related to the mastery of the structures used to express communicative or thought complement verbs.

One line of research that has provided crucial empirical evidence for de Villiers's sentential complementation

hypothesis is based on data obtained from the performance of false belief tasks by deaf children. Research on false belief understanding in deaf children has generally confirmed the importance of children's linguistic abilities for the development of ToM, showing that while deaf children born into hearing families (non-signing) have difficulties with ToM tasks, deaf children with deaf parents (native signers) do not differ significantly from hearing children in their responses to ToM tasks (Peterson & Siegal, 1995, 2000). Along similar lines, Gale, de Villiers, de Villiers, and Pyers (1996) found that deaf children educated in hearing schools, who, as a result, are delayed in their linguistic development, tend to complete the false belief understanding tasks three years later than hearing children. When the individual differences in ToM development in these children were analysed, it was found that control over sentential complement structures continued to be the best predictor of false belief understanding. Schick, de Villiers, de Villiers, and Hoffmeister (2007) found similar results about sentential complement structures. In their research, ToM abilities and language skills were studied in deaf children who use either American Sign Language (ASL) or oral English, with hearing parents or deaf parents. The study included 176 children with a profound hearing loss divided in three groups: children being educated orally, deaf children with hearing families learning American Sign Language (ASL), and deaf children with deaf families exposed to ASL from birth. The study revealed that there was a significant delay on ToM tasks in deaf children of hearing parents, who typically demonstrate language delays, regardless of whether they used spoken English or ASL. In contrast, deaf children from deaf families performed identically to same-aged hearing controls. Both vocabulary and understanding syntactic complements were significant independent predictors of success on verbal and low-verbal ToM tasks. The study also revealed that the language skills in the deaf children were directly related to their Theory of Mind skills. However, it wasn't the children's general language skills but rather vocabulary skills and the specific ability to comprehend syntactic complements that predicted Theory of Mind skills.

In spite of the facts summarized here, some studies have questioned the role of sentential complements in false belief understanding. They have argued that semantics, not syntax, predict children's scores in ToM tasks (Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003), or that some more general linguistic abilities are better predictors of false belief understanding (Slade & Ruffman, 2005). A different line of evidence that has cast doubt on the centrality of linguistic complements for false belief understanding comes from the study of their relation in different languages (Perner, Sprung, Zauner, & Haider, 2003; Tardif & Wellman, 2000).

Among the attempts to contrast the sentential complement hypothesis with alternative explanations, two training studies stand out (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). Hale and Tager-Flusberg (2003) found that sentential complement training led to improvement in various ToM tasks. Furthermore, the influence was specific, given that the same effect did not occur with subjects trained in other types of sentences such as those including relative clauses. It is important to point out that, as a result of only using verbs of communication during sentential complement training, the ToM progress observed in the post-test could not be assumed to be an effect of learning the significance of specific mental content verbs.

Lohmann and Tomasello (2003) provide further support for the importance of complementation for the development of ToM. They tested the influence of two main variables related to false belief understanding: the confrontation of one's own perspective with a different perspective in discursive interaction, and the understanding of sentential complements with mental and communication verbs. Lohmann and Tomasello (2003) found that the most effective condition for promoting ToM progress was that which brought together all the variables mentioned. On the other hand, the condition in which only the deceptive nature of the experience was emphasized, without any verbal recourse, did not produce significant pre-post differences. Furthermore, false-belief understanding of both children trained in sentence complementation and children trained in the confrontation of perspectives improved significantly.

In addition to the approaches discussed above, some studies have shown that other aspects of language, apart from conversational interaction and syntactic abilities, are related to the development of ToM. For instance, Doherty (2000), Happé (1995) and Schick et al. (2007) found that receptive vocabulary is related to ToM. Other authors have discovered a relationship between having alternative names for objects and false belief understanding (Meins & Fernyhough, 2007; Perner, Stummer, Sprung, & Doherty, 2002). In fact, one theoretical proposal holds that the ability to consider multiple names for objects precedes false belief understanding (Jacques & Zelazo, 2005a). Therefore, receptive vocabulary skills, and also the understanding of synonymy (i.e., having alternative names for objects such category names, fruit or apple) have been related to the ToM, though no training studies have focused specifically on naming objects from more than one perspective.

The review of existent research established that language influences false belief understanding. There is evidence to suggest that the discursive input to which the child is exposed plays a crucial role in promoting false belief understanding. Some training studies show that it is not necessary to use communication or mental state verbs, or mental terms, to provoke progress in ToM tasks. In addition, there is also a specific type of sentence structure (i.e., sentential complements), which has been proven to be important in false belief understanding. In the present study we examine in detail the role of two aspects of language that have been strongly linked to false belief understanding: the syntax of complementation and the perspective-shifting discourse. We start from the basic premise that training studies are the most appropriate approach to this question, and will use a procedure adapted from Lohmann and Tomasello (2003).

On the one hand, we assume that training in sentential complement sentences improves false belief understanding, as demonstrated by Hale and Tager-Flusberg (2003) and Lohmann and Tomasello (2003). However, in this type of sentences, the semantic and syntactic characteristics overlap so that it is difficult to distinguish which one is responsible for improvements in ToM. In that sense, as Milligan et al. (2007) have suggested, "training studies could further separate the semantic and syntactic elements of sentential complement training in order to discover whether the effect is primarily due to falsity or to embedded syntax" (p.641). Taking into account that 'falsity' here refers to the semantic aspect, Milligan et al.'s proposal can be carried out using sentential complement sentences where the subordinate clauses are true and not false. Therefore, in the present study we attempt to separate the semantic from the syntactic aspects in sentential complement training using non-deceptive objects. Thus, the first objective of our study is to identify whether the effects of sentential complement training on children's false belief understanding are due to the syntactic structure of the syntax of complementation or to the semantic aspects derived from using non-deceptive objects. In line with de Villiers's (2005) proposal that it is children's mastery of false complements that promotes false belief understanding, our first prediction is that sentential complement training with non-deceptive objects will not improve children's false belief understanding, because in this training true complement clauses will be used. If this prediction is confirmed, our results compared with those obtained in prior studies will show that sentential complement training is only effective when false complement clauses are used, and therefore, that it is mainly the falsity and not the embedded syntax what fosters children's understanding of false beliefs.

On the other hand, the perspective-shifting discourse has been proven to be effective in the understanding of false belief (Lohmann & Tomasello, 2003). In the present study we assume that perspective-shifting discourse contributes to improve false belief understanding. However, it must be remembered that in conversational training the deceptive objects are labelled. So, bearing in mind that different studies have related having alternative names for the same reality to false belief understanding, it is possible that the aspect of the conversation which fosters false belief understanding is the giving of alternative names to deceptive objects. In this sense, the second objective of our study is to investigate the possible influence of labelling on the development of false belief understanding. In order to address this issue, in our study we will compare discourse-based training with object-labelling training. Our second prediction is that training based on labelling objects according to different perspectives will improve children's understanding of false beliefs, as prior research has demonstrated that discourse-based training has this effect.

Method

Participants

Participants included 104 normally developing preschool children (M_{age} = 3.7. range = 3.5 – 3.11) from different socio-cultural backgrounds in the Girona area of Spain. Four criteria were used to include subjects in the sample: a) they were between 3 years and 5 months old and 3 years and 11 months old; b) they were Catalan speakers; c) they had normal language development according to the Peabody Vocabulary Test (Dunn, Padilla, Lugo, & Dunn, 1986); and d) they did not respond correctly to a representational change task (false belief pre-test). Apart from the 104 participants, a total of 43 children were interviewed but were excluded from the study for the following reasons: 18 of them did not reach a score within their normative age range (according to the PPVT); 15 responded correctly to both test questions in the pre-test representational change task; and 10 of them missed training sessions.

The children in the study participated in a total of five sessions: one pre-test, three training sessions, and one post-test session. The pre-test tasks were administered in the first session. The training sessions started three or four days after the pre-test, and each session took place three or four days after the previous session. Finally, three or four days after the last training session, the post-test was carried out. The training sessions were different in each experimental condition, as described in the procedure section.

Materials and scoring

Pre-test material

Children were tested individually in three pre-test tasks.

Peabody Vocabulary Test. Children were administered the TVIP, which is based on the PPVT (Dunn et al., 1986). This test was used to determine if they had a normal linguistic development.

False Belief Pre-test. A version of Gopnik and Astington's (1988) representational change (RC) task was used to test children's false belief knowledge. In our study, an egg box was shown to the children and they were asked: "What is inside the box?" Then, the box was opened and the children could see there were coins instead of eggs inside. Afterwards, they were asked the representational change question: "What did you think was inside the box when you first saw it?" After closing the box with the coins inside, the children were asked the false belief prediction question: "If we show this box all closed up to a friend of yours, what will he/she think is in the box?" One point was given for correct answers to each of the last two questions. Children with two points were eliminated from the sample.

Sentential Complement Pre-test. Considering that one training condition included sentential complement sentences, it was important to ensure the children had a similar understanding of this kind of sentence. If the group trained in sentential complement sentences had, at the outset, a higher or lower level of understanding of the syntax of complementation than the other group, this might have affected the results.

The children were thus given two sentential complement pre-tests:

1) A version of the Hale and Tager-Flusberg (2003) task which consisted of telling the children stories accompanied by two drawings. In the first picture the protagonist performed one activity but in the second picture claimed to have performed a different activity. The children were then asked two test questions, one referring to what the protagonist said and the other to what the protagonist really did. Three stories were used in our study. Children were given 1 point for each correct answer, so the maximum score in this task was 6 points.

2) An adaptation of the Lohmann and Tomasello (2003) task intended to assess children's understanding of false belief. In this case, three illustrated stories were shown to the children. In each story the protagonist hold a false belief, and children were asked two questions, one referring to the protagonist's behaviour and the other concerning the protagonist's thoughts. One point was

awarded for each correct answer. A total of 6 points was possible for this task. Thus, the maximum score for the two sentential complement tasks together was 12 points. Some examples of the sentential complement tasks have been included in the Appendix A.

Training material

Children meeting the pre-test criteria were assigned to one of three experimental conditions or to the control condition. In each of three sessions, the experimenter showed four objects to the children. In two of the experimental conditions and in the control condition, three deceptive objects and one non-deceptive object were used in each session. At first sight, the deceptive objects seemed to be one thing (i.e. a tomato), but when observed closely a different function was evident (i.e. the tomato is really a candle). Conversely, the non-deceptive objects had the same function as their appearance suggested. A total of nine deceptive and three non-deceptive objects were used during training, except in the Sentential complements with non-deceptive objects condition, where all the objects were non-deceptive. A Little Red Riding Hood puppet (from now on *puppet*) was used by the experimenter to interact with the children in all conditions. The four conditions are explained in greater detail in the procedure section.

Post-test material

Four different tasks were used in the post-test.

An adaptation of the *Change of Location Task* (Wimmer & Perner, 1983). This task consisted of telling the children a story using toys. The instructions were as follows: "Jordi and Mariona are playing with a ball. They decide to hide the ball in a yellow box. Then, Jordi leaves the room. While he is away, Mariona takes the ball from the yellow box and puts it in a blue box. After that, Jordi comes back." Then the children were asked a prediction question: "Where will Jordi first look for the ball?" When the children said or pointed to the yellow box, they were given 1 point.

A version of the *Appearance-Reality Task* (Flavell, 1986). Deceptive objects similar to the ones in the training were used. These objects involved a discrepancy between the real and the apparent identity. First, the experimenter showed each object by its apparent side, and asked: "What does it look like?" If the child did not answer correctly, the experimenter told the child what the object looked like (i.e. "It looks like a tomato."). After that, the experimenter let the child see the real function of the object and asked: "And now, what is it really?" In case the child gave a wrong answer, the experimenter told the child what the object

really was (i.e. "It is really a candle."). Afterwards, the puppet appeared on the scene and the experimenter introduced her to the child. Then, the experimenter took the object by its apparent side (i.e. hiding that it was a candle) and asked the child a test prediction question: "If we ask the puppet what it looks like, what will she say?" Finally, the real identity of the object was disclosed and the experimenter asked: "And now, if we ask the puppet what this really is, what will she say?" One point was given when the children correctly answered each prediction question. Thus, a score of 2 was possible for each object.

A *Representational Change Task* similar to the one used in the pre-test but with a different object. Specifically, we used a Smarties tube containing tiny stones instead of sweets. The same two test questions used in the pre-test were asked. A score of 2 was possible in this case.

The scores of all the false belief post-test tasks were added to obtain a global false belief post-test score: 1 point was given for a correct answer in the change of location task, 2 points were given for answering correctly the two questions in the representational change task, and since the appearance-reality task included three objects, the total score for this task was divided by three, giving a maximum score of 2 points. So the three false belief post-tests tasks had a maximum score of 5 points.

Procedure

Three adult experimenters conducted the pre-test, the training sessions and the post-test. Children were tested individually in a quiet room in their nursery schools. In the first session the pre-test tasks were administrated. Pre-test sessions lasted about 20 minutes. Children passing the pre-test were randomly assigned to one of the four groups. After three or four days, the training sessions started. Training sessions lasted about 10 minutes and were separated by three or four days. A total of three training sessions were conducted. Finally, three or four days after the last session, children were given a post-test. The post-test session lasted about 25 minutes. The design of the training sessions varied according to experimental conditions. The four conditions are described in more detail below and are summarized in Table 1. (See the Appendix A for an example of each training condition.)

Discourse training group (DIS group). In this condition, the deceptive aspect of the objects was mentioned in a communicative exchange without using mental state verbs or sentential complement constructions. First, we showed the deceptive part of the object to the children (i.e. a tomato) and asked them what it was. The children always answered incorrectly (i.e. *a tomato*), because the real identity of the object was hidden to

them. After that, the experimenter told the children that the object looked like one thing (i.e. *"Ok, it looks like a tomato. But look carefully")* and showed them the hidden part of the object (i.e. a candle). Then, the experimenter asked the children what the object really was. The children were corrected when they answered wrongly.

Afterwards, the puppet came to visit the children. We showed the apparent aspect of the object to the puppet (i.e. a tomato) and asked the children what the puppet would say when we asked her what the object was, using communicative verbs. The experimenter repeated the children's answer. After that, the experimenter said to the puppet what the object really was (i.e. "*Look puppet, it's a candle!*"). The puppet did not believe the experimenter (i.e. "*A candle? No, it's a tomato, I've seen it!")*. Then the experimenter showed the puppet the hidden part of the object and told her again what the object really was. The puppet showed surprise. Afterwards, the experimenter asked the children again what the puppet would say when asked what the object really was. If the children answered wrongly by saying the deceptive label of the object (i.e. *"a tomato")*, the experimenter corrected them (i.e. "*A tomato? No, now the puppet will say a candle, Ok?*"), and suggested that the children should listen to the puppet's answer. The puppet's answer corresponded to the hidden part of the object ("*It's a candle!*").

Labelling training group (LAB group). As in the previous condition, the deceptive aspect of the objects was mentioned without using sentential complement constructions. No communicative or mental state verbs were used either. As in the case of the discourse condition, the deceptive part of the object was made obvious. However, in the labelling condition the children were not asked about the real and apparent identity of the object. Instead, they were shown labels, and were mere observers of the experimenter. First, the children were shown the apparent identity of the object and the experimenter stated its label (i.e. "Look X, a tomato"). Afterwards, the experimenter showed and labelled the real identity of the object ("Now look, a candle"). This part was done three times. Later, the puppet appeared on the scene. The experimenter showed the apparent identity of the object to the puppet, and the puppet labelled the object's appearance (i.e. "*A tomato!*"). Then, the experimenter showed the real identity of the object and the puppet stated its label showing surprise, when the real identity of the object was made obvious ("*Oh! A candle!*"). The part where the puppet arrived was not repeated.

Sentential complements with non-deceptive objects group (SND group). In this condition the experimenter used mental and communicative verbs, and also sentential complements. However, in this case the objects were

Table 1. *Characteristics of the training groups*

| | Elements in the training | | |
Training condition	Objects	Syntax	Verbs
Discourse	Deceptive experience	Simple clauses	Communication verbs. No mental state verbs
Labelling	Deceptive experience	Only vocatives and attention-getting expressions	No communication or mental state verbs
Sentential complements	No deception	Complement sentences	Mental state and communication verbs
Control	Deceptive experience	Only vocatives and attention-getting expressions	Only imperative 'look'

Note: Adapted from Lohmann and Tomasello (2003)

not deceptive (i.e. a candle with the typical round shape of a candle). First, the experimenter showed the object and asked the children what they thought it was. If the children gave a wrong answer, the question was repeated.

When the children gave a correct answer (i.e. *"a candle"*), the experimenter asked two questions about the characteristics of the objects (i.e. *"Do you think that this candle is soft or hard?"*). Children's responses could be right or wrong, but the experimenter always congratulated them (i.e. *"Very good, you think that the candle is hard / soft"*). Afterwards, the puppet appeared and performed an action with the object (i.e. *"Look at me, I can hold up the candle!"*). Next, the experimenter asked the children what the puppet had shown she could do. If the children responded incorrectly, the experimenter corrected them (i.e. *"She showed us that she can hold up the candle"*). If the children answered correctly, the experimenter repeated their words. After that, the puppet made a comment about the object (i.e. *"Oh, the candle is burning! Oh, I've burned myself!"*) and the experimenter asked what the puppet had said. As before, children were corrected after a wrong answer (i.e. *"She said that she had burned herself."*) and their words were repeated after a correct answer. Finally, the experimenter asked a question about the puppet's knowledge of the object (i.e. *"Does the puppet know that the candle is dangerous?"*) and asked how she knew this. Wrong answers were corrected and right answers were repeated.

Control group (CON group). In this condition, the deceptive parts of the objects were shown, but no language was used, except for certain expressions to direct the children's attention. The children were just observers. First the experimenter showed the children the deceptive part of the object and said: *"Look!"* Then, the experimenter showed the hidden part of the object, while showing surprise, and said: *"But now*

look!" This part was repeated twice more. Later, the puppet arrived and the experimenter showed her the object's deceptive side: *"Puppet, look!"* The puppet answered: *"Oh!"* After that, the experimenter showed the puppet the hidden part of the object: *"And now, puppet, look again."* The puppet reacted in surprise: *"Oh!"* The part of the training involving the puppet was not repeated.

Results

We divided the results section into three parts. In the first part, we analyzed differences among training groups at the pre-test using the Kruskal-Wallis and Mann-Whitney non-parametric tests and the Chi-square test. In the second part, we compared the performance on the different theory of mind post-test tasks. The Kruskal-Wallis and the Mann-Whitney tests were conducted to compare performance on the representational change and appearance-reality tasks. A Chi-square test was used to analyze performance on the change of location task as score on this task was a dichotomous variable. Analysis of Variance (ANOVA) was used to examine between-group differences for the aggregate false belief task scores. In the third part, we used the Wilcoxon non-parametric test to investigate whether the training led to significant changes in the scores obtained in the representational change task. Furthermore, the improvement in the representational change task (the difference between the post-test and the pre-test scores) was compared among the different groups using the Mann-Whitney test.

Analyses of Group Differences at Pre-test

Chi-square (for the false belief pre-test score and the sex variable) and Kruskal-Wallis analyses were conducted to check for between-group differences in the pre-test scores. The results are shown in Table 2. There

were no significant group differences for age, χ^2_3 = 4.308, p = .230; sex, χ^2_3 = .427, p = .935; false belief pre-test score, χ^2_3 = 2.140, p = .544; sentential complement pre-test score, χ^2_3 = 4.199, p = .241, or Peabody test score, χ^2_3 = 3.601, p = .308.

Results of the post-test tasks

The scores in the different post-test tasks were compared for the different training groups (see Table 3 and Figure 1).

For the representational change task, the Mann-Whitney non-parametric test showed significant differences between the labelling group and the other three groups, DIS: Z = −2.76, p = .006; SND: Z = −3.55, p < .001; CON: Z = −3.404, p = .001. However, there were no significant differences between the discourse and the SND group, Z = −1.038, p = .299, between the discourse group and the control group, Z = −.814, p = .416, or between the SND group and the control group, Z = −.237, p = .813.

For the appearance-reality task we found a different pattern. The Kruskall-Wallis test revealed significant differences among training groups, χ^2_3 = 13.478, p = .004. The Mann-Whitney test showed that differences were significant between the labelling group

and the SND and the discourse groups, SND: Z = −3.21, p = .001; DIS: Z = −3.05, p = .002, but no differences were found between the control group and the other groups (p > .05). Also, we found no significant differences between the discourse and the SND group, Z = −.497, p = .619. Thus, subjects in the labelling group outperformed the other three groups in the representational change task, but in the appearance-reality tasks they outperformed the discourse group and the complement sentences group only.

Concerning the change of location task, we found significant differences as a function of condition, χ^2_3 = 14.083, p = .003. Pair-wise comparisons showed that the LAB group was significantly better than the other three groups, SND: χ^2_1 = 6.470, p = .011; DIS: χ^2_1 = 3.820, p = .051; CON: χ^2_1 = 12.235, p = .001, whereas the difference between the DIS and the SND groups was not significant, χ^2_1 = .391, p = .532. Similarly, the differences between the control group with the DIS and the SND groups were not significant (p > .05).

An ANOVA was conducted to compare the aggregate false belief task scores. Results showed differences among the different conditions, F = 9, 287, p < .001. A Tukey post hoc test revealed that the differences were significant between the labelling and the other three groups, DIS: p = .001; SND: p < .001; CON: p < .001.

Table 2. *Means (and SD) of age, sex, and pre-test scores of participants in each training group.*

	Training groups			
	DIS *n = 26*	LAB *n = 26*	SND *n = 26*	CON *n = 26*
Age in months	43.62 (1.86)	43.19 (1.96)	42.92 (1.62)	43.81 (1.67)
Sex*	0.54 (.51)	0.58 (.50)	0.58 (.50)	0.50 (0.51)
FB pre-test**	0.31 (.47)	0.27 (.45)	0.42 (.50)	0.42 (.50)
SC pre-test***	5.50 (2.42)	6.04 (1.99)	6.15 (1.83)	5.62 (1.58)
Peabody	25.25 (4.37)	28.93 (9.36)	30.50 (3.54)	24.68 (3.04)

Note: *Proportion of girls
**Maximum score = 1
***Maximum score = 12

Table 3. *Means (and SD) of post-test scores for all training groups*

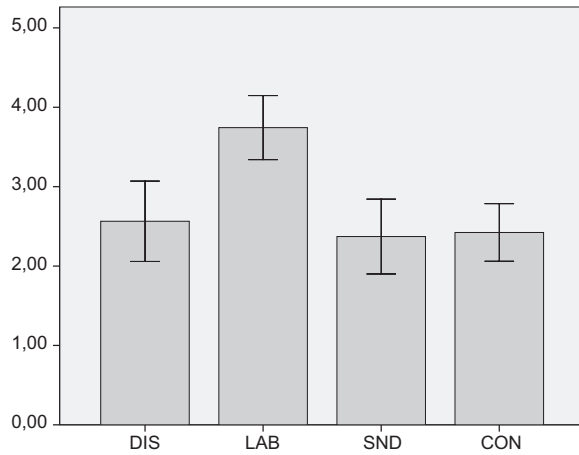|	Training groups			
	DIS *n = 26*	LAB *n = 26*	SND *n = 26*	CON *n = 26*
Change of location (0–1)	.31 (.47)	.58 (.50)	.23 (.43)	.12 (.33)
Representational change (0–2)	.73 (.72)	1.31 (.68)	.54 (.71)	.58 (.70)
Appearance-Reality (0–6)	4.58 (1.30)	5.58 (.81)	4.81 (.89)	5.19 (1.02)
Sum of False Belief Tasks (0–5)	2.56 (1.25)	3.74 (1)	2.37 (1.17)	2.42 (.90)

Figure 1. Sum of False Belief Tasks as a function of condition.

Thus, subjects in the labelling condition did better than the subjects in the other conditions on this aggregate score of false belief understanding (see Figure 1.).

Training effects: comparison of pre-test and post-test scores

The effects of each kind of training on false belief were assessed using the Wilcoxon non-parametric test and comparing the pre-test and post-test scores for the representational change task (RC) (see Table 4).

Both in the discourse and in the labelling training groups, subjects improved (i.e. did better on the post-test than on the pre-test) in the RC task, LAB: $Z = -2.67$, $p = .008$; $p = .803$; DIS: $Z = -3.95$, $p < .001$. Conversely, subjects in the SND and in the control group did not show any improvement, SND: $Z = -.90$, $p = .366$; CON: $Z = -1.000$, $p = .317$. In sum, in the representational change tasks, subjects increased their scores from pre-test to post-test in the discourse and labelling groups (see Figure 2).

When we compared the improvement in the representational change task (post-test minus pre-test scores) between the different conditions (using the Kruskal-Wallis test), we found that there was a significant difference, $\chi^2_3 = 20.837$, $p < .001$. When this comparison

was made group by group using the Mann-Whitney test, a significant difference was revealed between the improvement in the labelling group and the improvement in the other three training groups, DIS: $Z = -2.68$, $p = .007$; SND: $Z = -4.01$, $p < .001$; CON: $Z = -3.54$, $p < .001$. However, there were no significant differences between the discourse and the SND, $Z = -1.81$, $p = .070$, and the CON, $Z = -1.33$, $p = .182$, training groups. No significant difference existed between the SND and the CON training groups, $Z = -.249$, $p = .803$.

Discussion

Replicating findings from training studies (e.g., Lohmann & Tomasello, 2003), the present study found that the children in the discourse training group improved on the representational change task from the pre-test to the post-test. With regard to the two distinctive conditions of our study, the children from the labelling condition also improved their performance on the representational change task from the pre-test to the post-test, while the children trained in sentential complements with non-deceptive objects did not show any significant progress. Therefore, the training carried out in the discourse and labelling conditions improved children's performance on the representational change task. However, when we compare the post-test results between groups, the participants in the labelling condition improved significantly more than the other groups. The same pattern of findings emerged for the change of location task, which was not used in the pre-test or in the training. With respect to this task, children in the labelling condition performed superior to those in the other conditions. We can therefore assume that labelling training is more effective in improving false belief understanding than the other types of training.

In our study, the group trained in sentential complements with non-deceptive objects does not show any improvement in the representational change task, and obtains the lowest score of all the groups for the aggregate false belief score. These results, in conjunction with the studies showing the efficacy of the training with sentential complements and deceptive objects

Table 4. *Means (SD) of pre-test and post-test scores on representational change task.*

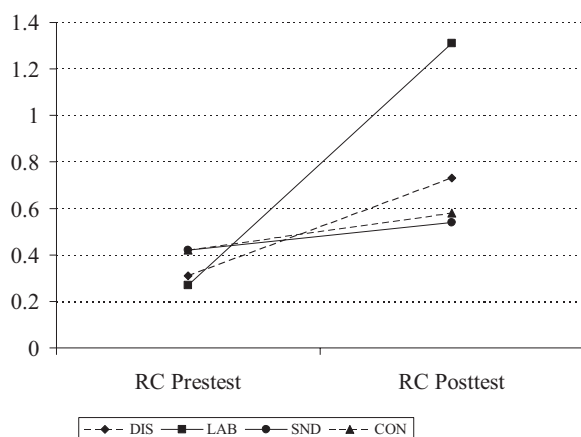|	Training groups			
	DIS *n = 26*	LAB *n = 26*	SND *n = 26*	CON *n = 26*
Representational change pre-test (0–2)	.31 (.47)	.27 (.45)	.42 (.50)	.42 (.50)
Representational change post-test (0–2)	.73 (.72)	1.31 (.68)	.54 (.71)	.58 (.70)

Figure 2. Pre-test vs. post-test scores on representational change task.

(Lohmann & Tomasello, 2003; Hale & Tager-Flugsberg, 2003), indicate that what causes improvement in false belief understanding is not the syntax of complementation alone, or the exposure to mental state or communication verbs, but rather the confrontation with a specific deceptive reality in which there is a double perspective: saying or thinking that an object is one thing while it is really something else. If this confrontation does not exist, the sentential complement syntax, despite the presence of mental state verbs or communication verbs, does not promote the development of false belief understanding.

As we have suggested above, a possible explanation of why the children trained in sentential complements with non-deceptive objects did not experience an improvement in false belief understanding is that they were not shown deceptive objects. However, it is important to consider here the results from the control group together with the results from the "no language" condition in Lohmann and Tomasello's (2003) study. In these conditions, the children were trained with deceptive objects, and no language was used, apart from some interjections and words to direct attention. Children's performance in false belief tasks did not improve following training, which strongly suggests that, apart from the exposure to deceptive objects, children need language to improve their false belief understanding. In other words, in order to develop their false belief understanding, both linguistic training and deceptive objects are required. In fact, these results are consistent with the hypotheses of de Villiers (2005), who argues that it is children's mastery of false complements –and not of true ones as in our study– what promotes false belief understanding.

On the other hand, in our labelling condition, under very similar circumstances to those in the control group or in the "no language condition" of Lohmann

and Tomasello (2003), and only adding the two possible labels for the reality and appearance of the object, there is a considerable improvement in false belief understanding. These results of the effect of labelling are rather surprising since previous studies did not indicate that labelling was the aspect of language that had the most notable impact on progress in false belief understanding. Some studies certainly produced data that were compatible with ours, but they were related to synonymy or to vocabulary skills. For Perner et al. (2002), the aspect of vocabulary that correlated most strongly with false belief performance was the ability to understand synonyms. Other studies, such as those carried out by Ruffman et al. (2003) and Schick et al. (2007) also demonstrate a significant relationship between semantics/vocabulary and false belief understanding. In any case, general vocabulary and synonymy are measures that, in some respects, are tapping into aspects of linguistic ability similar to the double labelling of a deceptive object. However, in other respects, they differ from double labelling. Although an object can have more than one name, being two different things, depending on whether its appearance or its function is considered, is a completely different matter. We do not know if training in the double labelling of a non-deceptive object (synonymy) would lead to an improvement in false belief understanding. This is an aspect that could be tested in future studies.

Even though the data from our study on the effects of labelling on false belief understanding lead us to consider that there is an aspect of labelling which is important for improving it, we must also remember that, in the study by Lohmann and Tomasello (2003), the condition in which the complementation and the discourse syntax were joined promoted the largest improvement in false belief understanding. It is therefore possible that what is most favourable to false belief understanding is the emphasis on the double perspective of objects, which can be transmitted in different ways, although always through the medium of language. The labelling condition, even though the communicative load is unidirectional, is sufficient to capture the essence of conversational input in relation to false belief understanding: conceiving of a single object from more than one perspective.

That said, and taking into consideration that deceptive objects are also labelled in the discourse condition, a critical question is why labelling has such a significant effect. If object labelling is found in both the labelling and the discourse conditions, why does training work better in the labelling condition? It should be remembered here that, in the discourse condition, the children participated in a conversational interaction about the objects during which the two possible

labels were conveyed, the double perspective was emphasized and the participants were included in a conversation. For its part, the labelling condition was a passive condition for the participants, who only observed and heard what the experimenter told them about the objects, but in which the two possible object labels were also emphasized. Thus, in both conditions the objects are labelled according to a double perspective and their deceptive character is emphasized. A possible explanation for the differences between them is that the labelling condition is more informative for the children since it isolates other aspects of the situation. In this way, no other linguistic elements have to be considered, which can make it difficult for the children to know what they have to pay attention to. It thus makes it easier for them to focus on what is critical, i.e. the deceptive character of the objects (in other words, their double perspective).

The theoretical perspective that best corresponds to the results of this study is one that assumes that an aspect of language (labelling) plays a decisive role in the development of cognitive flexibility, i.e. the ability to simultaneously consider multiple but conflictive representations of a single object or event (Jacques & Zelazo, 2005a). According to these authors, cognitive flexibility plays a central role in the development of ToM, and all the measures of false belief require cognitive flexibility. Jacques and Zelazo's (2005a) suggestion that labelling is effective, makes sense since the critical point is to conceive –signal, make obvious, note– different points of view, or representations of a single phenomenon, object or event. In some way, the proposals made by Tomasello (1999) and Harris (2005), which emphasize that a discourse that allows for different perspectives with respect to a single object or event is enough to produce improvements in ToM tasks, would also be compatible with the previous idea. The role of labelling in cognitive flexibility is therefore discussed in greater detail.

The labelling condition from our study includes the alternative naming investigated by Perner et al. (2002) and Meins and Fernyhough (2007). However, our condition also includes a conflict between the labels used to name the objects. In fact, whereas in the alternative naming system both labels are true (i.e. *rabbit and bunny*), in our condition the labels were opposite (if it is a candle it cannot be a tomato). This type of labelling is better adjusted to the executive function called cognitive flexibility. Labelling has been widely used to investigate the influence of language on executive functions, and more specifically, on cognitive flexibility (Jacques & Zelazo, 2005a; Müller, Jacques, Brocki, & Zelazo, 2009). In this sense, two kinds of measures have been used to evaluate cognitive flexibility: inductive and deductive tasks. These measures differ in the amount of information provided to children. In deductive measures all the information necessary to solve a task is provided to the children, and they simply need to use this information appropriately. On the other hand, in the case of inductive measures, the child does not receive all the information needed to solve the task, and has to infer it inductively. We can give an example of the distinction between a deductive and an inductive task in the form of two cognitive flexibility tasks. In the Dimensional Change Card Sort (DCCS), the children have to pair targets according to featural similarities, and the criterion of selection (e.g. colour) is given to the children. What they have to do is therefore to apply a given rule (deduce). In another task, the Flexible Item Selection Task (FIST), the children also have to pair targets according to featural similarities, but in this case the criterion is not explicit. Thus, the children here have to induce the criterion needed to solve the task (see Jacques & Zelazo, 2005b). In fact, despite that labelling training is not used to assess the children, we could say that it is a form of deductive training because the children receive all the information needed to understand the two perspectives of objects.

The effect of labelling has consistently been related to inductive measures (see Jacques & Zelazo, 2005b for a revision), while the results obtained using deductive measures are more inconsistent. For example, in a study with the DCCS, Kirkham, Cruess and Diamond (2003) found that labelling improves the performance of the task, while other research (Müller, Zelazo, Lurye, & Lieberman, 2008; Towse, Redbond, Houston-Price, & Cook, 2000), did not find such an effect. On the other hand, Müller, Zelazo, Hood, Leone, & Rohrer (2004), using another deductive paradigm, also found that three-year-olds improved their performance in the task when they were asked to use labelling. However, there are probably other factors, in addition to task format, which can affect the effectiveness of labelling.

In a discussion of the studies designed to investigate the effects of language on children's executive functions, Müller et al. (2009) examined different possibilities about the role of labelling in cognitive flexibility. They concluded that labelling could be differentially effective at different stages of attentional processing. Specifically, due to labelling, "children must be capable of quickly disengaging attention from one attribute and engaging attention to a different attribute (…). The increased attentional flexibility then would be a prerequisite for the coordination of lower order rules in a higher order system" (Müller et al., 2009, p. 65). In relation to our study, it is plausible that providing a label in the appearance-reality training could shift children's attention to the relevant stimulus dimension. This is indeed related to what we said before: labelling

makes it easier for the children to focus on what is critical, the double perspective of deceptive objects.

In short, according to our data, various aspects of language can influence or promote progress in false belief understanding, as long as they are related to a reality that allows more than one perspective. The reasons why some aspects of language are more strongly related to false belief understanding than others, and whether there are other aspects that might also have an influence, need to be studied in greater depth.

References

Astington J. W., & Jenkins J. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35, 1311–1320. http://dx.doi.org/10.1037/0012-1649.35.5.1311

Astington J. W., & Baird J. A. (2005). *Why language matters for the theory of mind*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195159912.003.0001

Bartsch K., & Wellman H. M. (1995). *Children talk about the mind*. New York; NY: Oxford University Press.

Cutting A., & Dunn J. (2006). Conversations with siblings and with friends: Links between relationship quality and social understanding. *British Journal of Developmental Psychology*, 24, 73–87. http://dx.doi.org/10.1348/026151005X70337

de Villiers J. (2005). Can language acquisition give children a point of view? In J. W. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp.186–219). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195159912.003.0001

de Villiers J., & de Villiers P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, UK: Psychology Press.

de Villiers J., & Pyers J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17, 1037–1060. http://dx.doi.org/10.1016/S0885-2014(02)00073-4

Doherty M. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, 27, 367–392. http://dx.doi.org/10.1017/S0305000900004153

Dunn L., Padilla E., Lugo D., & Dunn L. (1986). *Test de Vocabulario de Imágenes Peabody. Adaptación Hispanoamericana.* [Peabody Picture Vocabulary Test. Latin American adaptation]. Circle Pines, MN: American Guidance Service.

Flavell J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist*, 41, 418–425. http://dx.doi.org/10.1037/0003-066X.41.4.418

Gale E., de Villiers P., de Villiers J., & Pyers J. (1996) Language and theory of mind in oral deaf children. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds.), *Proceedings of the 20th annual Boston university conference on language development* (Vol. 1, pp. 213–224). Somerville, MA: Cascadilla Press.

Gopnik A., & Astington J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59, 26–37. http://dx.doi.org/10.1111/j.1467-8624.1988.tb03192.x

Hale C. M., & Tager-Flusberg H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, 6, 346–359. http://dx.doi.org/10.1111/1467-7687.00289

Happé F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843–855. http://dx.doi.org/10.2307/1131954

Harris P. (2005). Conversation, pretense, and theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 70–83). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195159912.003.0001

Jacques S., & Zelazo D. (2005a). Language and the development of cognitive flexibility: Implications for theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 144–162). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195159912.003.0008

Jacques S., & Zelazo P. D. (2005b). On the possible socio-communicative roots of cognitive flexibility. In B. D. Homer & C. S. Tamis-Lemonda (Eds.), *The development of social understanding and communication* (pp. 53–81). Mahwah, NJ: Lawrence Erlbaum.

Kirkham N. Z., Cruess L., & Diamond A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6, 449–467. http://dx.doi.org/10.1111/1467-7687.00300

Lohmann H., & Tomasello M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, 74, 1130–1144. http://dx.doi.org/10.1111/1467-8624.00597

Malle B. F. (2002). The relation between language and theory of mind in development and evolution. In T. Givón, & B. F. Malle (Eds.), *The evolution of language out of pre-language*. (pp. 265–284). Amsterdam, The Netherlands: Benjamins.

Meins E., & Fernyhough C. (2007). Preschoolers' understanding of multiple orientations to reality: The Adjectives task. *Cognitive Development*, 22, 289–297. http://dx.doi.org/10.1016/j.cogdev.2006.10.008

Milligan K., Astington J. W., & Dack L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78, 622–646. http://dx.doi.org/10.1111/j.1467-8624.2007.01018.x

Müller U., Zelazo P. D., Hood S., Leone T., & Rohrer L. (2004). Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development*, 75, 1594–1609. http://dx.doi.org/10.1111/j.1467-8624.2004.00759.x

Müller U., Zelazo P. D., Lurye L. E., & Liebermann D. P. (2008). The effect of labeling on preschool children's

performance in the dimensional change card sort task. *Cognitive Development*, 23, 395–408. http://dx.doi.org/10.1016/j.cogdev.2008.06.001

Müller U., Jacques S., Brocki K., & Zelazo P. (2009). The executive functions of language in preschool children. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 53–68). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511581533

Olson D. R. (1988). On the origins of beliefs and other intentional states in children. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 414–426). Cambridge, MA: Cambridge University Press.

Perner J., Stummer S., Sprung M., & Doherty M. (2002). Theory of mind finds its piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development*, 17, 1451–1472. http://dx.doi.org/10.1016/S0885-2014(02)00127-2

Peterson C. C., & Siegal M. (1995). Deafness, conversation and Theory of mind. *Journal of Child Psychology and Psychiatry*, 36, 459–474. http://dx.doi.org/10.1111/j.1469-7610.1995.tb01303.x

Peterson C. C., & Siegal M. (2000). Insights into theory of mind from deafness and autism. *Mind and Language*, 15, 123–145. http://dx.doi.org/10.1111/1468-0017.00126

Ruffman T., Slade L., Rowlandson K., Rumsey C., & Garnham A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18, 139–158. http://dx.doi.org/10.1016/S0885-2014(03)00002-9

Schick B., de Villiers P., de Villiers J., & Hoffmeister R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, 78, 376–396. http://dx.doi.org/10.1111/j.1467-8624.2007.01004.x

Slade L., & Ruffman T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of sintax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23, 117–141. http://dx.doi.org/10.1348/026151004X21332

Tardif T., & Wellman H. M. (2000). Acquisition of mental state language in Mandarin- and Cantonese-speaking children. *Developmental Psychology*, 36, 25–43. http://dx.doi.org/10.1037/0012-1649.36.1.25

Tomasello M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello M., & Carpenter M. (2007). Shared intentionality. *Developmental Science*, 10, 121–125. http://dx.doi.org/10.1111/j.1467-7687.2007.00573.x

Towse J. N., Redbone J., Houston-Price C., & Cook S. (2000). Understanding the dimensional change card sort. *Cognitive Development*, 15, 347–365. http://dx.doi.org/10.1016/S0885-2014(00)00021-6

Wimmer H., & Perner J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128. http://dx.doi.org/10.1016/0010-0277(83)90004-5

APPENDIX A

Sentential complements tasks

An example of each kind of sentential complements task is given.

Communicative Verbs task

Drawing 1: *"Marc is playing with cars."*

Drawing 2: *"Later, his mother arrives and asks him: "Marc, what were you doing?"* Marc answers: *"I was playing with a ball." Then, Marc goes away."*

Test question 1: *"What did Marc say he was playing with?"*

Test question 2: *"What was Marc really and truly playing with?"*

Belief Verbs task

Drawing 1: *"Raquel is at her home right now. She wants to play with a ball outside."*

Drawing 2: *"Raquel thinks that it is still daytime."*

Drawing 3: *"But it is really and truly dark outside."*

Drawing 4: **Test question 1:** *"Will Raquel go outside to play with the ball?"*

Test question 2:*"What did Raquel think first, that it was daytime or night-time?*

Training procedures

Example of a Discourse training session

The experimenter shows the deceptive part of the object and asks:

E: *"What is this?"*

Child: *"It's a tomato."*

E: *"Ok, it looks like a tomato. But look carefully."*

Then the experimenter shows the hidden part of the object:

E: *"What is it really?"*

- Child: *"It's a tomato!"* → E: *"Ok, it looks like a tomato. But look carefully, it's a candle! What is it really?* (If the child says "a tomato", the experimenter repeats this sentence another time)

When the child answers "a candle", the training continues.

A puppet appears and sees the deceptive part of the object:

Then, the experimenter asks the child:

"When I ask the puppet what is this, what will she say?"

- If the child answers: *"a candle"* → E: *"a candle?"*

- If the child answers: *"a tomato"* → E: *"a tomato?"*

E: *"Let's hear what the puppet says. Look puppet, I have a candle."*

Puppet: *"A candle? No, it's a tomato, I've seen it!"*

Then the experimenter shows the puppet the hidden part of the object:

E: *"What?! Look puppet, it's a candle!"*

Puppet: *"Oh!"* (surprised)

E: *"Now, when I ask the puppet what is it really, what will she say?"*

- If the child answers: *"a candle"* → E: *"a candle? Very good!"*

- If the child answers: *"a tomato"* → E: *"a tomato? No, now the puppet will say a candle, Ok?"*

E: *"Ok, let's hear what the puppet says. Puppet, what is it really?"*

Puppet: *"It's a candle!"*

Example of a Labelling training session

The experimenter shows the deceptive part of the object and says:

E: *"Look X, a tomato."*

Then the experimenter shows the hidden part of the object:

E: *"Now look, a candle."*

This part of the training is done three times.

Then, a Little Red Riding Hood puppet appears and sees the deceptive part of the object:

E: *"Look puppet!"*

Puppet: *"A tomato!"*

After that, the experimenter shows to the puppet the hidden part of the object:

E: *"Now look!"*

Puppet: *"Oh! A candle!"*

The part when the puppet arrives is not repeated.

Example of a Sentential complements with non-deceptive objects training session

The experimenter shows the object:

E: *"What do you think this is?"*

Child: *"A candle."* → E: *"Very good, you think that it is a candle."*

Child: Another answer → E: *"What do you think this is?"*

E: *"Do you think that this candle is soft or hard?"*

Child: *"Hard."* → E: *"Very good, you think that this candle is hard."*

Child: *"Soft."* → E: *"Very good, you think that this candle is soft."*

E: Do you think that this candle can be burned?

Child: *"Yes."* → E: *"Very good, you think that this candle can be burned."*

Child: *"No."* → E: *"Very good, you think that this candle cannot be burned."*

The puppet appears and sees the object:

Puppet: *"Look at me, I can hold up the candle!"*

E: *"What has the puppet shown that she can do?"*

Child: *"That she can hold up the candle."* → E: *"Very good, she showed us that she can hold up the candle."*

Child: Another answer → E: *"She showed us that she can hold up the candle."*

Puppet: "Oh, the candle is burning! Oh, I burned myself!"

E: *"What did the puppet say?"*

Child: *"That she burned herself."* → E: *"Very good, she said that she had burned herself."*

Child: Another answer → E: *"She said that she had burned herself."*

E: *"Does the puppet know that the candle is dangerous?"*

Child: *"Yes."* → E: *"Very good, she knows that the candle is dangerous."*

Child: *"No."* → E: *"She knows that the candle is dangerous."*

E: *"How does she know this?"*

Child: *"Yes."* → E: *"Very good, because she has seen that it burns."*

Child: Another answer → E: *"Because she has seen that it burns."*

Example of a Control training session

The experimenter shows the deceptive part of the object:

E: *"Look!"*

Then, the experimenter shows the hidden part of the object:

E: *"But now look!"*

This part of the training is done three times.

Then, the puppet arrives, and the experimenter shows her the deceptive part of the object:

E: *"Puppet, look!"*

Puppet: *"Oh!"*

Finally, the experimenter shows the puppet the hidden part of the object:

E: *"And now, puppet, look again."*

Puppet: *"Oh!"*

This part is not repeated.