The Link between Language and Theory of Mind: Evidence from Internationally-Adopted Children

Joy Geren[1], Jesse Snedeker[1] & Carissa L. Shafto[2]

[1]Harvard University

[2]University of Louisville

Draft 12/10/2009
Comments welcome
Please request permission before citing

Address correspondence to:
Joy Geren
Lab for Developmental Studies
202 Shannon Hall
25 Francis Avenue
Cambridge, MA 02138
Phone: 617-384-8357; fax: 617-496-0975
geren@fas.harvard.edu

Abstract

There is ample evidence for a connection between linguistic abilities and performance in theory of mind tasks (ToM), however there is considerable dispute about precisely how these domains are linked. Disentangling the causal relations between domains in typically developing children is tricky because many skills are developing in synchrony. If ToM performance depends upon children's current linguistic abilities, then the relation should be present even when language acquisition is delayed. We explored these issues in a group of late English learners, 45 internationally adopted children who came to the US at 2.5 years or older, and a control group of language matched preschoolers who were learning English as a first language. The children were tested on general English language abilities, sentence complement comprehension, English vocabulary and ToM. The two groups performed similarly on standard verbal ToM tasks, even though the adopted group was on average nearly 3 years older. However, the adopted children outperformed the controls in ToM tasks with reduced linguistic demands. General language skills predicted ToM ability in both groups and complement comprehension did not account for any additional variance. The data suggests that executive functioning along with general language skills may be the critical components in ToM success.

INTRODUCTION

Classic Theory of Mind

For twenty years developmental psychology has been fascinated by an apparent leap that occurs in children's ability to represent mental states at around four years of age. In a variety of tasks three-year-olds act as if their current beliefs were shared by everyone, while four-year-olds correctly recognize that people's beliefs can differ and depend upon their perceptual experience (Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987; Surian & Leslie, 1999). For example in Perner and colleagues classic unseen displacement task children hear a story in which a character, Maxi, places his chocolate in the cupboard and leaves (Perner et al., 1987). Then his mother, unbeknownst to Maxi, moves the chocolate to the refrigerator. The child is simply asked where Maxi will look for the chocolate. Most three-year-olds will cheerfully respond that Maxi will look in the refrigerator, because that is where the chocolate is. In contrast many four-year-olds will explain that Maxi will look in the cupboard, because that is where he believes the chocolate is.

This robust and highly replicable shift cried out for a causal explanation, and variety of proposals rapidly followed. While some researchers argued that the shift reflected the maturation or development of a domain-specfic system for representing mental states, others noted that many related abilities emerged during the second year of life (Leslie, 1994; see Scholl & Leslie, 1999 for review). Success on false belief tasks was predicted by a variety of abilities spawning proposals that linked ToM with peer experience, the development of memory, and executive functions. Accounts of this kind are difficulty to test in typically developing children. Children are simultaneously developing skills in many domains, and development across domains is often correlated. These correlations can reflect the underlying causal structure of the problem, but they can also be driven broad factors (SES, neural efficiency, little g) that indirectly influence development in all domains.

The present study explores the proposal that success in the false belief is a result of the child's emerging linguistic abilities (citations for several such proposals). To disentangle the contributions of linguistic development from the growth in other cognitive skills, we focused on a population in which language abilities systematically lag behind other aspects of cognition: children who have been internationally adopted after that age of 2 and are learning English as a second language. If performance on explicit theory of mind tasks depends largely on a child's current linguistic abilities then these children should perform like much younger children with the same level of linguistic skills. However, if the relation between language and ToM performance, is driven in whole, or in part, by concurrent development in other domains, then we would expect the adopted children to out perform language matched controls.

In the remainder of this introduction, we discuss three issues. First, a number of recent studies demonstrate that infants and toddlers can represent the mental states of others. We briefly discuss this work and the impact it has on our understanding of the developmental transition that is measured by classic false belief tasks. Second, we describe the range of proposals that have been put forward linking language and performance on the false belief task and the evidence supporting this proposal. Finally, we describe the population that we are testing, internationally adopted children, and how their situation differs from other populations that have been used to explore these questions.

Early Evidence of Understanding Others

Failure at these classic tasks does not, however, mean that younger children are completely unable to reason about the desires, perceptions and knowledge of others prior to age

4. In fact, a number of studies have demonstrated these abilities in the second year of life. Several studies have found that 12- to 18-month-old children are able to distinguish between objects that are familiar to an actor from those that are novel based on what the actor has seen (e.g. Liszkowski, Carpenter, & Tomasello, 2008; Luo and Baillargeon, 2007; Tomasello & Haberl, 2003).

Given children's early competence in understanding others, why do they continue to fail classic ToM tasks for several years? There are three possible answers to this question: 1) early abilities are real but extremely limited in scope, 2) failures on classic theory of mind tasks are due to task demands and not limitations in ToM understanding, or 3) there are differences in the type of understanding required by each of the two task types. Each of these explanations will be considered briefly.

1) *Early competence is limited in scope: Even one-year-olds may be able to identify that a person is familiar with an object that they have seen, but they may be limited in the type of beliefs that they can represent.* One possibility is that very young children can use a simple "seeing is knowing" theory of other's beliefs, but they may fail to follow the actor's representation when it is more complex, for instance when it is inconsistent with what they know is true, or based on other types of information. This does not appear to be the case. Under optimal conditions toddlers reveal that they are not limited to simple understanding of familiarity. Song and Bailargeon (2008) demonstrated with a looking time task that 14.5 month olds, aware of its inconsistency with reality, are still able to represent an actor's false belief. Another study found that 18-month-old children understand that an actor's representations can be influenced by linguistic information as well as visual information (Song, Onishi, Baillargeon, & Fisher, 2008). Understanding is demonstrated in unintentional (i.e. looking time, Song, Onishi,

Baillargeon, & Fisher, 2008) and intentional (i.e. pointing, Liszkowski, Carpenter, & Tomasello, 2008) actions. The variety of beliefs young children can interpret as well as the multiple ways in which this understanding is demonstrated constrain the plausibility of a limited scope hypothesis.

2) *Task demands limit ToM performance: Children under 4 continue to fail classic ToM tasks because, despite a thorough ToM understanding, they are unable to deal with the demands of these tasks.* While it is true that the classic ToM tasks require linguistic and general cognitive abilities not yet possessed by children in their second year, this fails to account for the failures of young preschool children. Numerous attempts have been made to reduce the task demands of the classic ToM tasks with very little effect on the age of passing. In a meta-analysis, Wellman, Cross and Watson (2001) reported minimal deviations in the reported age of passing ToM tasks from 178 studies representing nearly 600 ToM testing conditions. The consistency of findings despite varying task demands suggests that there is a meaningful transition in children's understanding around age four.

Additional support for the idea that there is a meaningful shift in children's understanding of others minds comes from children's spontaneous production. Bartsch and Wellman (1995) report three phases of development in how children talk about desires, thoughts and beliefs. Very young children talk only about desires, while children in the second stage, starting around their third birthday, also speak of thoughts and beliefs. However, it is not until children reach the final stage, around age 4, that they begin to you use language to describe the connection between people's beliefs to their behaviors (Bartsch & Wellman, 1995, 143-144). The stages observed in natural language production provide support for the timing and ecological validity of classic ToM tasks.

3) *Early competence and classic ToM success represent different levels of understanding: Children have some understanding of others at a very young age while the more robust understanding of beliefs required by classic ToM tasks is not achieved until around age 4.* Children in their second year of life are, under certain conditions, able to demonstrate an understanding that others can differ from themselves in their familiarity with objects (e.g. Liszkowski, Carpenter, & Tomasello, 2008; Luo and Baillargeon, 2007; Tomasello & Haberl, 2003) and desires (Repacholi and Gopnik, 1997). In some cases they can even predict an actor's false belief and what type of linguistic cues will lead to successful revision of that belief (Song & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fisher, 2008). However, it is not until around age 4 years that children attain the more robust understanding that allows them to pass the classic ToM tasks which require an explicit understanding of what informs other's beliefs and how those beliefs influence behavior.

Both the early competence in understanding of others and the shift in understanding during the preschool years appear to be valid phenomena. The next step then is to determine what leads to the predictable shift in understanding around age four that allows for a more robust understanding of others.

Predictors of Classic Theory of Mind Success

Some researchers have suggested that increases in cognitive abilities, such as changes in executive function, occur between age 3 and 4 years allowing children to pass classic, explicit ToM tasks (e.g. Frye, Zelazo, & Palfai, 1995). One necessary ability is the suppression of knowledge of reality in order to accurately judge the beliefs and associated behaviors of a naïve actor. On typical ToM tasks adults are unimpaired by this conflicting knowledge, but more sensitive measures reveal that even in adulthood there are processing costs associated with needing to suppress knowledge of reality (Apperly, Back, Samsom & France, 2008; Birch & Bloom, 2007). One possibility is that around age four suppression abilities may improve enough to allow children to override interference in standard ToM tasks (Wellman & Bartsch, 1988).

Other researchers believe that ToM success is critically related to linguistic competence (see Milligan, Astington & Dack, 2007 for meta-analysis). Some studies have linked ToM performance to broad measures of language development or multiple different linguistic skills (e.g. Cheung et al., 2004; Slade & Ruffman, 2005), while others have suggested that the acquisition of specific lingusitic skills are critcial. In particular, a number of researchers have argued that ToM competence is linked to either general (?) syntactic ability or to the mastery of syntax and semantics of sentential complements (de Villiers & Pyers, 2002; Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). The sentential complement structure allows a speaker to express the propositional content of a person's beliefs or communications, regardless of whether these propositions are consistent with reality. De Villiers and colleagues have proposed to this linguistic structure provides an internal representations of belief states that allows them to be clearly distinguished from reality. On this account comprehension of sentential complements is a prerequisite for ToM understanding.

A challenge for ToM research is that typical developing children are simultaneously developing many different cognitive and linguistic skills making it difficult to confirm causal relationships between specific abilities. Groups with delayed language development, such as deaf children, support the notion that ToM abilities are linked to language development since they demonstrate delays on typical ToM tasks. Several studies of deaf individuals concur that language exposure and ability are predictive of performance on ToM tasks (e.g. Courtin & Melot, 2005; Jackson, 2001; Pyers, 2005). These findings hold even when the linguistic demands of the ToM assessment are minimized (Pyers, 2005), demonstrating that the relationship is not due to the linguistic demands of the ToM tasks themselves. Children with SLI also demonstrate a lag in their language skills relative to their chronological age and general cognitive abilities. Miller (2004) examined the relationship between sentence complement performance and ToM in children with SLI. Unfortunately the results were unclear, possibly due to a small sample size.

By testing the relations between ToM and language ability in a variety of populations could provide critical information about whether particular features of language are central to performance on explicit ToM tasks. In typically developing children syntactic abilities, lexical knowledge, and the mastery of sentence complements tend to develop in synchrony and may fuel the development of a variety of cognitive skills that rely on linguistic instruction or experience. These associations may not be as strong in children who learning a language at a later stage in development. For example, children who have delayed access to spoken language, because they received a cochlear implant after infancy, rapidly catch up in vocabulary acquisition but have lingering syntactic delays (e.g. Bollard et al., 1999; Geers, Nicholas, & Sedey, 2003; Geren & Snedeker, in preparation; Spencer, 2004; Svirsky et al., 2000; Young & Killen, 2002).

Another population of potential interest is children who experience a temporary language delay because of a sudden change in their linguistic environment. Our recent research has focused on internationally adopted children who enter the US during the preschool or early school years. These children show rapid attrition of their native language, well before they have acquired age appropriate skills in English (see Snedeker, Geren, Shafto, in press). Consequently, their linguistic abilities are, temporarily, well behind their chronological age, yet because they previously had another language to communicate in they do not have the massive cognitive or social delays that might otherwise be expected of older children with very limited linguistic knowledge.

International Adoption as a Theory of Mind Test Case

Our population of internationally adopted children has come to the U.S. after having started to learn their native language. At the time of adoption these children spoke no English and all were adopted into monolingual English speaking families. Under these circumstances children lose their birth language and become monolingual English speakers (see Snedeker, Geren & Shafto, in press). A study of 4-8 year old children adopted from Russia found that the children's productive Russian was greatly diminished after only 3-6 months in the U.S. and one year after adoption the children retained no functional command of their birth language (Gindis, 1999). Studies of adult adoptees offer additional evidence for the loss of birth language. One study examined adults who had been adopted from Korea into French families as children. The adult adoptees showed no behavioral or neurological evidence of recalling Korean or even being able to discriminate Korean from Japanese (Pallier et al., 2003).

While they are losing their birth language, these children are rapidly acquiring English. Nevertheless there is a period of several years in which their linguistic abilities lag behind their chronological age (Geren, Shafto & Snedeker, in progress; Snedeker, Geren & Shafto, 2007). Therefore we are able to test a unique population of children who, despite normal perceptual abilities and continuous language exposure, currently have limited language abilities for their ages. Thus our working hypothesis was that the adopted children in our sample, having spent months or years away from their native language, would have lost the abilities in their native language that are typically associated with success on ToM tasks. If this is true, then there are three possible predictions that one might make about their performance depending on our hypothesis about the relation between language and ToM.

1. Language as a trigger: On this hypothesis specific linguistic skills are necessary for acquiring the representations that support overt ToM performance, but, once those representations are acquired, full access to their linguistic expression is no longer necessary. If this is the case then we would predict that children who came to the United States after the age at which they would be expected to pass first order false belief tasks, would retain this ability even if their current level of linguistic ability in both their birth language and their adoptive language is well below the level typically associated with ToM success. For example, the acquisition of the sentence complement structure could allow children to develop a conceptual representation that supports the representation of beliefs. However, once this conceptual representation is available children may not need to access its linguistic expression in order to use it for nonlinguistic purposes. On this hypothesis, we would predict that children who come to the United States at or after age 5, but who no longer comprehend complements in either their birth language or their adoptive language, will still retain the ability to pass first-order false belief tasks.
2. Language as an implementation: The second possibility is that the representation of beliefs is tightly linked to the current linguistic abilities of the child. Such a hypothesis would be consistent with recent work showing that adults' performance in a nonverbal false-belief task is disrupted by verbal shadowing, but not by rhythmic tapping (Newton &de Villiers, 2007). One possible interpretation of these findings is that reasoning about false beliefs requires the use of the linguistic forms used to

encode these beliefs (possibly even including their phonological representations). If this is the case, then children, regardless of their progress in their first language, should be unable to pass ToM tasks from the time they lose the required language skills in their birth language, until the time that they master them in their adoptive language.

3. Language as correlate: Finally, it is conceivable that the systematic associations between language abilities and ToM performance are not attributable to the kind straight-forward causal relation posited above. Correlational patterns are by their nature ambiguous. Logically, the association between language and ToM in typically developing children could reflect: effects of language on ToM performance, effects of skills tapped in ToM tasks on language tasks, or the effects of a third variable (or variables) on both processes. The challenge for any alternate account is to explain why this association persists in populations with linguistic delays but no cognitive delay. We return to this question in the discussion. For now we simply note that hypotheses of this kind make clear predictions about the performance of the adopted children. If linguistic skills are not causally implicated in ToM performance then we might expect that our adopted children would perform reliably better than typically developing children with a similar level of linguistic ability, since they are considerably older and considerably more cognitively advanced (see Snedeker et al., in press; Geren, Shafto & Snedeker, in preparation). The cognitive advantages of older children may also provide some insight into the cognitive mechanisms that underlie different linguistic tasks. For example it is possible that some language skills are more closely yoked to the development of executive function than they are

to linguistic experience and thus might develop in advance of general language ability in adopted children. For example, tasks probing the comprehension of sentence complements might draw on some of the same inhibitory abilities required in classic ToM tasks. If children's mastery of complements is typically limited by the development of nonlinguistic skills, then we might expect older adopted children to acquire these structures in advance of other linguistic skills.

To distinguish between these three possibilities it is critical that we have a measure of ToM performance that does not require the comprehension or production of linguistic expressions that the child may not have acquired. Thus in these experiments we employed two types of ToM tasks, the common verbal tasks and low verbal tasks. Low verbal tasks reduce the linguistic demands of the tasks by presenting the background information as well as the question in pictorial form. Like standard verbal ToM tasks, these tasks still require the child to make explicit decisions about another person's thoughts, but are more appropriate for testing children who are currently limited in their language abilities. Our first study examined English language abilities and ToM in adopted children and native English speaking controls. Experiment 2 tests our working hypothesis that the current birth language skills of these adopted children are well below the level that is typically associated with success in ToM tasks.

Experiment 1

Participants

<u>The Experimental Group</u>

Forty-five internationally adopted children (23 female) from Eastern Europe and Asia were included in this study (see table 1). These participants were selected from a larger sample of 90 (in Geren, Shafto & Snedeker, in progress) based on their chronological age and their

English language abilities. Our goal was to focus on children who had reached the age at which children typically pass explicit ToM tasks, but who might not have the linguistic abilities that are typically associated with ToM task performance. For this reason we excluded all children younger than 4.5 years and all children who had Diagnostic Evaluation of Language Variation (Seymour, Roeper, & de Villiers, 2005) language scores higher than what is expected for an average 6 year old.

Participants were adopted from Russia (n=17), Kazakhstan (n=12), China (n=8), India (n=6), South Korea (n=1) and Cambodia (n=1). All children were adopted by monolingual English speakers and continued access to their native language was limited, with 33 children having no exposure to their birth language at the time of participation in this study. Eight children had access to a speaker of their language roughly 2-10 hours each week. The remaining 4 children were adopted as part of sibling groups less than a year prior to testing; their access to their native language was unrestricted but their adoptive parents reported that children's usage of their native language had decreased substantially over time. Children were excluded from participation if they had been diagnosed with a sensory, motor, or developmental condition that could affect language development (*e.g*. hearing loss or Down syndrome).

Participants were primarily recruited through national adoption support groups (*e.g.* Families for Russian and Ukrainian Adoption). The groups posted information about the study in their newsletters and on their web sites. Additional families were recruited through a national adoption agency specializing in international adoptions and yahoo adoption support groups.

<u>The Control Group</u>

Internationally adopted children were compared to a control group matched on general English language ability as measured by DELV raw scores. The control group consisted of 45 monolingual English speaking children (21 female) who were born in the U. S. The participants were significantly younger than the adopted group (see table 1) and had normal language abilities for their age, defined as a vocabulary score between the 10^{th} and 90^{th} percentile compared to norms for their chronological age. Fifteen of the control participants were between the ages of 3.0 and 3.9 years old and thus below the age for which the DELV was developed and normed.

Language Measures

Two standardized English language measures were used. The Peabody Picture Vocabulary Test-III (PPVT-III; Dunn & Dunn, 1997) was administered as a test of receptive English vocabulary. General English language abilities were assessed using the Diagnostic Evaluation of Language Variation, Norm Referenced version (DELV; Seymour, Roeper, & de Villiers, 2005), a broad test of syntactic, semantic, pragmatic and phonological development. Analyses focused on the syntax, semantics, and pragmatics sections which contribute to the total language composite score. The phonology section was not analyzed.

A separate measure tested comprehension of false complement clauses. The task (fully described in Schick, deVilliers, de Villiers, & Hoffmeister, 2007) uses communication verbs and does not necessarily require understanding of false beliefs (de Villiers & Pyers, 2002). Previous studies have found that comprehension of the complement structure is closely related to the ability to represent mental states (de Villiers & Pyers, 2002; Schick, et al., 2007). Twelve complement clause items, all using communication verbs, were tested. Ten or more correct answers was considered a passing score on this task.

Cognitive Measures

Children's nonverbal cognitive abilities were assessed using the matrices section of the Kaufman Brief Intelligence Test-II (KBIT-II NV; Kaufman & Kaufman, 2004). We chose not to use the verbal section of the test, as we wanted a cognitive assessment that was independent of English language ability.

Theory of Mind Measures

Children's ToM abilities were assessed using variants of two widely administered false belief tasks– an unexpected contents task (Perner, Leekam & Wimmer, 1987) and an unseen displacement task (Wimmer & Perner, 1983) and a pictorial ToM task, designed to minimize linguistic task requirements.

In the unexpected contents task the child was shown a familiar container, a crayon box, and invited to open the box and look inside. If they did not spontaneously produce a label for the unexpected contents ("oh, necklaces") they were asked what was inside. After the contents were once again hidden out of view in the box the child was asked "What did you think was in the box before you opened it?" and "Mom (or other non-present older family member) has not seen inside the box. What will she think is in the box before she opens it?" A second unexpected contents item, an egg carton containing dinosaurs, was presented in a later session.

In the unseen displacement task the child was told a simple story accompanied by series of photographs depicting a toy being moved while the owner was away. Two memory questions were asked to make sure the child comprehended the story and remembered the facts ("Where did Sally put the toy? Where is it now?"). Then, a final photograph depicting the toy's owner standing between the hiding places was presented and the child was asked two test questions "Where will Sally first look for her toy?" and "Why will she look there?" They could respond either with a verbal label or by pointing. The first item portrayed a little brother moving his sister's toy star, and in the second item the wind displaced a boy's toy ball and another helpful child retrieved it but returned it to a novel location. Only items for which the child correctly answered the memory questions were scored.

Each test question in the unexpected contents and unseen displacement was scored 0 or 1 resulting in a verbal ToM composite score of 0-8 for the two tasks. A composite score of 6, 7 or 8 was considered passing verbal ToM.

The pictorial, or low verbal ToM task (based on Custer, 1996) began with three trial items designed to test children's understanding of thought bubbles.[1] Each item was two-alternative forced-choice and children were required to answer all three items correctly in order to continue[2]. Next, two sets of test items were administered. In the first set the child was shown a drawing of a common scene, for instance a boy fishing. Then they were shown a second drawing depicting the character retrieving something from out of view (e.g. the fishing line was hidden behind some reeds). The child was encouraged to lift a flap to see what the person was about to retrieve (e.g. a boot on the line). Finally the child was asked to choose one of three pictures to place in the character's thought bubble. On the three unexpected outcome trials the choices included the expected item (a fish), the real item (a boot) and a distracter item (a bird). On two additional trials the character was retrieving the expected object and two distracter items were used. Verbal prompts on all items were limited to identifying the initial situation (e.g. "He's fishing") and asking the child to respond with the correct picture ("what goes here?").

The second type of low-verbal item presented a story as a series of 4 pictures depicting a person placing an unexpected item in a common container. In the final scene a second character

---

[1] The low verbal ToM task was created by Peter DeVilliers

[2] Occasionally children made errors stemming from a lack of world knowledge or an alternate construal of the situation. These children demonstrated understanding of the thought bubbles in the other examples and were not excluded.

who had (1 test item) or had not (3 test items) seen the unexpected item placed was about to open the container. The child had to put the correct picture into the second character's thought bubble. All items were two-alternative forced-choice with the choices being the expected contents of the container or the true contents. Verbal information was limited to labeling the expected and unexpected contents as they were initially presented (e.g. "crayons") and directing the child's attention ("See"). The two types of low verbal items were combined to give the child a total low verbal ToM score ranging from 0-9. A composite score of 7, 8 or 9 was considered passing low verbal ToM.

Procedure

Each participant's testing lasted between 2 and 4 hours conducted over one to three sessions depending on the needs of the child. When possible, testing was done in a quiet room with only the experimenter, but otherwise in a quiet area with minimal distractions. Normed assessments were administered according to standard procedures. During these session 21 children, all 6 years old or over, also completed tests of reading and phonological processing the results of which are reported in Shafto, Geren & Mervis (in progress). Parents completed a background questionnaire to document early language experience, pre-adoptive living situations, physical health parameters, and information about the child's school placement in the U.S.

RESULTS

Performance on General Language and Cognitive Measures

Independent sample t-tests were used to compare the adopted and control groups. See table 2 for means and test values. As planned, the two groups did not differ in overall DELV performance and the adopted group was significantly older than the control group. Standard scores indicate that the adopted children were performing substantially below age level on both the DELV and the PPVT-III. This is expected given the children's limited exposure to English.

Vocabulary scores were significantly higher in the adopted group than in controls. This is consistent with our other findings that children who are delayed in exposure to English due to being internationally adopted (Geren, Shafto & Snedeker, in progress) show more rapid progress in vocabulary development than syntactic development. DELV rather than PPVT-III scores have been used for matching the adopted and control groups because past studies suggest general language ability or syntactic abilities, both measured by the DELV, are more likely predictors of ToM ability than vocabulary knowledge.

Table 3 shows the correlations among raw scores (?) for the language measures for each group. PPVT-III vocabulary scores were highly correlated with DELV scores in both groups. Sentence complement scores were significantly correlated with both PPVT-III and DELV scores in the control group, but only with DELV scores in the adopted group. Thus the correlations between broad language measures and sentence complement comprehension were weaker in the adopted group than in the control group. The reduction in the intercorrelation in the adopted group is potentially useful for disentangling the contributions of general syntactic abilities and complements comprehension on ToM performance.

KBIT-II NV raw scores confirm that the adopted children had greater non-verbal cognitive skills than the controls, as expected given the ages of the respective groups. On average KBIT-II NV standard scores for the adopted kids fell within the normal range for their age. However the mean standard score of 89.9 was below the population mean for the normal sample (100). This is not particularly surprising. The vast majority of the adopted children (43 of 45) had spent time in institutional care. Prior studies have documented that children living in orphanages have cognitive delays (The St. Petersburg-USA Orphanage Research Team, 2005). For the purposes of this study it is not necessary that the adopted children have nonverbal abilities commensurate with their age. However, it is critical that their nonverbal intelligence is substantially greater than the younger language matched control group. The reliable difference between the raw KBIT-II NV scores confirms that this is the case.

Two of the measures, the KBIT-II NV and the DELV, were normed for ages 4 years and above, so the standard scores on those tests were calculated only for the controls who were 4 or older (n=30). The control group, on average, scored in the normal range for their age. Notably performance in the 10th to 90th percentile range on the PPVT-III was a prerequisite for controls to be included in the study. Average scores on all three standardized measures are slightly higher than the norming samples of their respective tests. This is typical in populations self selected to participate in research studies.

As expected the standard scores on the language measures were reliably different between the two groups. The adoptees performed well below age level, confirming that in the first years after adoption, their skills in English are substantially delayed relative to their chronological age. The focus on the present study is not on how adopted children compare to peers of the same age. Instead we are interested in how the child's absolute level of linguistic knowledge influences their performance on ToM tasks. Thus to facilitate comparisons between the two groups which are comparable in general language abilities but substantially different in age, raw scores, not standard scores, will be used in all further analyses.

Understanding Sentence Complements

The complements task was passed by significantly more adopted participants than controls (76% vs. 53% , $t(88)=2.24$, $p<.05$). A stepwise backwards regression was run with raw complement score as the outcome variable and DELV score, group (1=adoptee, 0=control) and the interaction between these two variables as predictors. Both group and DELV score were reliable predictors of complement score with no interaction (see table 4). Specifically, complements scores are higher for children with higher general language scores, however the adopted group had an advantage on the complements task that was not predicted by their general language abilities.

Language predictors of Theory of Mind

In both groups, children who passed the complements task scored significantly higher on the verbal ToM tasks (see table 5). In the control group, complement passers also scored significantly higher on low verbal ToM tasks. In the adopted group, complement passers were only marginally better on low verbal ToM tasks ($p<.07$).

Separate backwards regressions were run to determine the relationship between the vocabulary and general language measures and scores on the two types of ToM measures. Both PPVT-III and DELV scores were highly predictive of verbal and low verbal ToM scores in both the adopted and control groups (see table 6). Due to the high correlation between PPVT-III and DELV scores only one of two will be used in the remaining analyses. We have chosen DELV because it accounts for a greater amount of the ToM variance in our test population and a meta-analysis of false belief literature found receptive vocabulary measures to have a significantly smaller relation to false belief performance than general language measures (Milligan, Astington & Dack, 2007).

Verbal ToM

There was no difference between the adopted group and the control group in their performance on the verbal ToM task. Passing scores were achieved by 35.6% of the adopted participants and 35.6% of controls. Both groups answered about half of the questions correctly ($M= 48.3\%$ and $M= 47.8\%$ for adopted and controls respectively, $t(88)=.07$, $p=.94$). Because the adoptees were substantially older and more cognitively advanced, this suggests that linguistic abilities play a central role in these tasks.

A backwards linear regression was run to further evaluate the verbal ToM performance pattern of the two groups. DELV score, complements score and group membership were entered as possible predictors (see table 7). Only DELV was significant, no group differences were found. Next a regression was conducted for each group separately to determine if DELV and complements performance had the same relationship to verbal ToM in the adopted and control participants. In both groups DELV was significant and complements score was not (see table 8). Thus across all analyses, the two populations show no differences in verbal ToM performance suggesting that performance on these tasks is strongly associated with children's current language abilities.

Low Verbal ToM

In contrast with the verbal ToM measures, the low verbal ToM task was passed by significantly more (48.9%) adopted, than control (20%) participants ($t(88)=2.99$, $p<.01$). If we treat performance on this tasks as a continuous measure, the adopted group performed significantly better than the controls ($t(88)=2.62$, $p<.05$) averaging 64% correct compared to 48% for the control group.

The same regressions run for verbal ToM were repeated using low verbal ToM performance as the dependent variable. In a backwards linear regression all of the possible predictors entered. DELV and group were both significant predictors, while the contribution of the complements score was only marginally significant (see table 7). Thus, the findings for low-verbal task were similar to the high verbal task in one respect: both were strongly predicted by general language skills as measured by the DELV. However, the findings for the low verbal ToM task diverged from the verbal ToM task in one critical respect: the groups were significantly different. Although the two groups differed in their complements score, this did not fully account for the difference between the groups (group continued to have a reliable effect even after complement score was included in the model). Thus whatever advantage the adopted children possessed in the low verbal ToM tasks is not entirely accounted for by their better understanding of sentence complements.

To understand the nature of the differences between the two group, separate analyses were conducted for the adoptees and controls with DELV and complements scores as possible predictors. In the adopted children, the DELV was a significant predictor of low verbal ToM but complement comprehension was not. In the control population the DELV was also predictive, albeit to a lesser degree, however complements score was marginally predictive as well (see table 8).

Thus when linguistic demands were minimized, adopted children performed better on explicit theory of mind tasks than would be expected on the basis of their linguistic abilities alone. The advantage of the adopted children on this task cannot be attributed solely to their better comprehension of sentence complements.

Summary of Experiment 1 Results

The adopted and control groups in experiment 1, matched on general language skills (DELV), performed identically on the verbal ToM tasks. Despite having an advantage in nonverbal cognitive skills (KBIT 2 NV), vocabulary (PPVT-III), and sentence complement comprehension, the verbal ToM performance of the adopted group did not differ from the younger controls. This finding suggests that verbal ToM performance is tightly related to general language abilities.

In contrast, the adopted children performed better than the controls on the low verbal ToM tasks. This difference is not solely attributable to the adopted children's better comprehension of the English sentence complements; the difference between the groups continues to be reliable when the sentence complement score is included in the analysis. In fact, within the adopted group, complement score is not an independent predictor of the low verbal ToM score (see table 8).

Notice that this pattern of findings is clearly compatible with two of the hypotheses suggested in the introduction. The language as trigger hypothesis predicts that many adopted children should have access to the cognitive structures needed to represent false beliefs because they have already acquired the necessary syntactic machinery to trigger these representations in their first language. These abilities may be masked in the verbal ToM tasks because the children lack the mastery of English that is necessary for following the stories or responding to the prompts. However, in the low-verbal ToM tasks the children's skill is unmasked. The language correlate hypothesis is equally compatible with the present data. On this hypothesis the correlation between linguistic skills and ToM performance in typically developing skills is not attributable to a one-way causal connection. Instead it could reflect the complex interrelation of developments in multiple domains (e.g., effects of development in other domains on both language and ToM). On this hypothesis, we would expect that ToM performance in adoptees would be reliably better than it is in language matched controls because the other cognitive skills that contribute to this ability would be more developed in this population. Again, the lack of any difference in the verbal task would be attributed to verbal demands which masked the adopted children's underlying skill in representing mental states.

The present results, however, appear to be inconsistent with the language as implementation hypothesis. If performance on ToM tasks depends primarily on the current linguistic skills that a child possesses then we should expect internationally adopted children to perform as poorly as language matched controls, even if the ToM task itself has minimal overt linguistic demands. Specifically, on the hypothesis that the comprehension of false beliefs is tightly linked to the acquisition of sentence complements, we should expect no group difference between adopted children and controls when complement comprehension is factored out. But the difference persists.

There is however, one possible explanation of this data pattern that would be consistent with the language as implementation hypothesis. Perhaps the adopted children are succeeding at the ToM tasks because they have current access to the complement structure in their birth language. To test this hypothesis, Experiment 2 probed children's comprehension of basic vocabulary and the complement structure in their birth language.

Experiment 2

Participants

A subset of 15 subjects from Experiment 1 were tested on a computer-based task designed to tap children's knowledge of basic vocabulary and their understanding of complements in their native language (see table 9). Not all subjects were tested because this measure was introduced after testing was underway and because we only created versions in Russian and Mandarin. Adoptive families often have little information about the exact dialect used by their child prior to adoption and we wanted to minimize the chance of testing children in a language or dialect they had never heard, thus we focused on the most common standardized languages of our sample. Twelve children were tested in Russian (9 from Kazakhstan, 3 from Russia) and three children were tested in Mandarin (all from China).

Measure

The native language task included 4 types of test items- native vocabulary recognition, native complement comprehension, English complement comprehension and simple English control questions. On each trial the child was shown colorful photographs on a computer monitor while recorded sound files presented the questions. Participants indicated their response by pointing to one of four pictures on the screen at the time of the question. After 3 simple training items acclimated the child to listening to recorded speech and pointing to pictures on the screen the child received the remaining items in pseudo-random order. For native vocabulary items the child was asked “Where’s the (car/house/tree/girl/dog/pants)?” in their native language. Complements items were the same as in the comprehension of false complements task, only the pictures and voice were presented by the computer and rather than responding verbally the child chose from one of four pictures on the screen. All sound files were recorded by native speakers of the language and verified as accurate by a second native speaker.

Each child was given 6 complement items in English, 3 false complements and 3 true complements to prevent them from developing an alternative strategy to pass the native language task, and 6 false complements presented completely in their native language. Finally 6 control items asked simple questions in English to encourage continued participation and verify that the child was attending. These included items such as “Point to the one you can eat.” Two presentation orders were used for each language to counter-balance order of presentation.

Correctly answering 5 or 6 out of 6 questions in a category constituted a passing score for that task.

Results

Only 5 of the 15 children tested were able to pass the basic vocabulary task in their native language. Only 1 of these children, the oldest in the study, was able to pass the native complements task. That child also passed the English complement task. All 5 of the children who passed the native vocabulary task had been in U.S. for less than one year and had continued exposure to their native language, four of the children reported to still use their native language with siblings adopted as a sibling group.

None of the 10 children who failed the vocabulary task had continued exposure to their native language[3]. Though 8 of them had been away from their native country for less than 1.5 years, they no longer retained the basic vocabulary tested.

Summary of Experiment 2 Results

Performance on the computerized native language task was consistent with previous research showing rapid attrition of birth language skills. One might expect the sample in experiment 2 to be more likely than the general sample to have retained complements understanding in their native language. They were, on average, adopted at an older age than the general sample meaning they had a longer duration of exposure to their native language, and they have spent less time in the US leaving less time for language attrition. They also have, on average, a greater continued exposure to their birth language than the general sample. Therefore, although only 1/3 of the children in the study were tested on native complements, it is reasonable to use their outcomes to estimate the native language abilities of the general sample. If the same proportion of the larger sample has retained native complements abilities we would estimate that

[3] One child also failed the English controls suggesting general inattention on the task.

3 children (6.6%) in the study could pass native complements. Assuming random distribution this would result in a less than 50% chance that any of the children who failed English sentence complements would retain the ability to comprehend native complements. If we presume that older age of adoption and shorter time in the US are also contributing factors to native complement retention the estimate would be even lower. Therefore, it seems unlikely that a retained knowledge of complement structure in the child's native language is contributing to their level of success on the low verbal ToM tasks.

DISCUSSION

Performance on verbal ToM showed a close and stable association with language skills. Adopted children only succeeded on the verbal task when they had English language skills that were at approximately the same level as the first language learners who succeeded at this task. In contrast the association between low verbal ToM tasks and current language skills was not stable across the two populations. Adopted children performed significantly better on low verbal ToM tasks than language matched controls. This critical finding demonstrates that the specific level of current language skill that is a associated with ToM performance in first-language learners is not a necessary pre-requisite for representing false beliefs. Second-language learners with English language skills below the level of typically developing 4 year old perform quite well on non verbal false belief tasks.

Interestingly, second-language learners also have an advantage in comprehending sentence complements in English. However this advantage in complement comprehension cannot fully explain their ToM success. Even when complement comprehension is included as a predictor, there is a reliable difference between the two groups. Despite the advantage of adopted group in ToM, current language skill continues to be a predictor in this population

In the remainder of this discussion section we: 1) return to the hypotheses presented in the introduction, evaluating them in light of our data; 2) explore the ambiguity of the observed correlations between language and ToM performance; and 3) integrate these findings with the existing data on the development of ToM across linguistically divergent populations.

Three Hypotheses about the correlation between language and ToM

The present data bear most strongly on the language as implementation hypothesis. If reasoning about false beliefs requires the use of the linguistic forms used to encode these beliefs, then we would expect that internationally-adopted children would fail at verbal and low-verbal ToM tasks, until they gained the relevant skills in their adoptive language. While success on the verbal task showed a stable association with current language skill, performance on low verbal ToM tasks did not. Thus we conclude that children do not have to have current syntactic or semantic abilities at the level of a 4 year old to represent mental states. Our data also suggest that current access to the complement structure is also not necessary for success in explicit FB tasks: even when the adoptees advantage on complement comprehension was factored out the group difference persisted.

In contrast, the language as a trigger hypothesis provides a straightforward explanation of these findings. On this hypothesis we would expect adopted children to perform better than the controls on the low Verbal ToM task because many of them would have acquired the relevant linguistic abilities in their first language prior to adoption. Acquiring these skills provided them with the representational wherewithal to stably represent false beliefs, and this ability persisted even after they had forgotten how these semantic representations were phonologically encoded in their birth language. In the verbal ToM tasks, this ability was presumably masked by the children's poor language skills which were taxed by the difficulty of comprehending the stories and providing verbal responses. If the critical linguistic trigger is full acquisition of the sentence complement structure, such an account might also provides an intuitively satisfying explanation for the adoptees advantage in complement comprehension. On this story acquiring the complement structure in one's first language involves marshalling (or possibly even creating) conceptual representations that have the property of semantic opacity. Once these representations are made available, acquisition of the complement structure in a second language may be a far less onerous task.

The present results, however, are equally compatible with the language as a correlate hypothesis. The internationally adopted children are older than the controls, consequently their nonlinguistic cognitive skills are more likely to have advanced to a point where they can reliably succeed at explicit ToM tasks, consequently they perform better on low-verbal ToM measures. Again on this hypothesis these abilities are masked in the verbal ToM tasks by the demands that such tasks make on the limited English skills of the adopted children. This account might also provide an explanation for why the adopted children fared better on the sentence complements task than language matched controls. The sentence complements task has many of the same representational and inhibitory demands as the ToM tasks. To succeed children must represent two states (reality and what was said) and must inhibit the tendency to treat reality as privileged. The inability of typically-developing children to master this task before roughly 4 could be attributable to limitations in general representational capacity or inhibitory skills, rather than linguistic abilities per se. In older children these limitations are removed and the mastery of complements is desynchronized from language development. As we noted earlier the challenge for this account is to explain why ToM skills would be delayed in linguistically divergent populations, we return to this issue below.

Interpreting the correlations between current language ability and ToM

If language is not necessary for the implementation of ToM it raises the question of why we observed any correlation between linguistic abilities and ToM skills in internationally-adopted preschoolers. Before we address this issue, it is worth noting that the correlations that we observed were strongest in our broad measure of language skill. DELV scores were more robustly correlated with ToM performance than the child's score on the sentence complements task. This was true in both verbal ToM tasks and low verbal tasks, and in both the adopted children and the controls. This is consistent with prior findings indicating that broad language skills are strong predictors of ToM development. The relationship between sentence complement comprehension and ToM was particularly weak in the adopted children. This suggests that the systematic relation that has been observed in typically-developing children may be an artifact of the complex intercorrelation of development across multiple domains that characterizes typical development. Alternately, the weaker association between these abilities in the adopted children may suggest that the acquisition of sentence complements in a first language has effects on TOM tasks independent of the child's *current* linguistic knowledge, reducing the predictive value of the complement comprehension in the second language.

Earlier we noted that the difference between the two groups on low verbal ToM tasks was inconsistent with the language as implementation hypothesis. When the verbal demands are decreased, success in explicit ToM tasks does not require that children's current language skills be equivalent to those of first language learners who can perform these tasks. But if the children are not using English to succeed in this task, why is their performance correlated with their English language DELV scores? On the face of it the data appear to be incompatible with the language as trigger hypothesis and the language as correlate hypothesis. If the structure has been triggered in the first language or the child has achieved the necessary nonlinguistic skills then they should succeed with the low verbal task regardless of their level of English mastery.

There are three possible explanations for this correlation

1) English language comprehension is required for performance of the task even if it is not implicated in representing false beliefs.

2) The child's current level of English language comprehension has some effect on the representation of the false beliefs, even though adopted children do not have to achieve the same level of mastery as first language learners in order to succeed. Perhaps it is helpful, but not necessary.

3) The child's score on the DELV is a merely a proxy for other variables which were not entered into these analyses (or not measured) but do play a causal role in representing false beliefs. For example, the raw DELV score is correlated with the child's current age, the age at which the child was adopted, and the child's nonverbal IQ.

The first possibility seems unlikely. The language used in the low verbal task was minimal, children merely needed to understand individual nouns and understand that they should select the matching picture. Our adopted children had English language skills that were equivalent to typical monolingual children of 3 to 6 years. Thus we would expect them to understand this kind of instructions, and all of the children appeared to do so.

The present study however cannot distinguish between the remaining two hypotheses. This study was designed to explore whether age and prior linguistic experience contributed to ToM success independent of effects of current language ability. It was not designed to determine whether language was a unique predictor of ToM skill. Our assessment of nonlinguistic cognitive ability was limited to a single coarse measure and our design lacked the power to disentangle the contribution of multiple variables.

A related concern is whether the adopted children in this study have a deficit in low verbal ToM performance relative to their chronological age, which might reflect their current low level of linguistic skill. This is difficult to assess given our design. We did not have a group of nonadopted controls who were matched with our adoptees in cognitive ability, but who had substantially greater English language skills. Given the large number of potential cognitive abilities that would have to be measured, such a design would be infeasible. The performance of the adopted children does appear to be somewhat lower than what we would predict on the basis of their age alone.

Integrating the existing research on the role of language in ToM tasks.

Earlier we concluded that the present data is consistent with language as trigger hypothesis and the language as correlate hypothesis, but it is not consistent with the hypothesis that linguistic abilities which emerge at around four years of age are necessary parts of the implementation of TOM skills. Here we revisit these three hypotheses in light of other studies exploring the relation between language and ToM.

As we noted in the introduction, studies of typically-developing children provide limited leverage on these questions because children's cognitive abilities tend to be tightly correlated across domains for a variety of reasons. Thus, the most persuasive evidence comes from studies that manipulate linguistic variables or explore the development of ToM in populations with unique linguistic experiences but fairly standard cognitive equipment. Four lines of evidence are often used to argue that language is causally implicated in performance on explicit ToM tasks.

First, as we noted earlier, deaf children of hearing parents have delays in language and also in verbal and low verbal ToM tasks (Courtin & Melot, 2005; Meristo et al., 2007; Peterson & Siegal, 1995; Schick, de Villiers,de Villiers, & Hoffmeister, 2007). Deaf children of deaf parents are not delayed in either domain (Courtin & Melot, 2005; Schick et al., 2007) and so researchers have quite reasonably concluded that ToM deficit in deaf children of hearing parents must stem from limitations in their linguistic experiences. However, these findings cannot tell us what role of language is playing in ToM development. The linked deficits could reflect direct effects of language acquisition on the emergence of ToM skills (as suggested by the language as trigger and language as implementation hypotheses) or they could reflect more complex causal pathways (as suggested by the language as correlate hypothesis). For example, language could facilitate conversations with family members which make mental states more salient to children (Peterson & Siegal, 1995; Cutting & Dunn, 1999). Or linguistic experiences at an earlier age could promote the development of working memory and other executive functions which go on to facilitate performance on the false belief task. This possibility receives support from research documenting that deaf children of hearing parents have deficits in tasks involving impulse control, inhibition, working memory and cognitive set shifting, which are correlated with their linguistic skills (Corina & Singleton, 2009; Figueras, Edwards & Langdon, 2008; but see P. de Villiers, 2005).

Second, training studies document that verbal experience can lead to improvements in ToM performance. For example, when children are trained to verbally report a false communication (conveyed with a sentence complement structure), they improve on false belief and appearance/reality tasks (Hale & Tager-Flusberg, 2003; see also Lohmann & Tomasello, 2003). This suggests that an experience, which is largely verbal, can impact ToM performance, but it does not resolve the question of the mechanism by which it does so. Reporting the contents of a false communication has similar inhibitory demands to acting on a false belief—one must override the tendency to rely on reality. Alternatively, the condensed exposure to false communication could lead the children to consider why a character is saying something false (Is he ignorant of the true state of affairs or deceptive?) thereby focusing their attention on the importance of mental states. In short, these training tasks could be influencing ToM performance through their effects on linguistic representations or executive function, or they could be directly providing input to domain-specific mentalizing processes.

Third, and most persuasively, recent work on an emerging language documents a close relation between mental state language and ToM performance. Pyers and Senghas (2009) examined the use of mental state verbs and performance on a low verbal false belief task in two groups of adults who used Nicaraguan Sign Language (NSL) as their sole means of communication. The older participants were young adults who had been initially exposed to NSL as children prior to 1984 when the language was just emerging. The younger participants were currently adolescents and had been introduced to a richer form of the language in the late 80's or early 90's. At their first assessment, the researchers found that members of the older cohort used few mental-state verbs and were unable to pass low verbal false belief tasks. In contrast the younger cohort passed the false belief task and frequently produced mental-state verbs. In a follow up study two years later, after the groups began socializing more, the older cohort showed improvement in both their use of mental state verbs and the false belief task. These findings are hard to reconcile with the language as correlate hypothesis. While it is plausible that six year olds with language delays could lack the inhibitory control of typically-developing four year olds, it seems inconceivable that these young adults (with jobs and domestic responsibilities) could be so profoundly impaired. Thus these findings suggest that language either provides the implementation for representing mental states or serves as a trigger.

Notice that these first three lines of research document a link between language acquisition and TOM performance. Even on their strongest interpretation, such findings are compatible with both the language as trigger and the language as implementation hypotheses, and thus are readily reconciled with the findings of the present study. To the best of our knowledge, there is only one piece of evidence that uniquely supports the hypothesis that language is necessary to implement reasoning about mental states. Newton and de Villiers (2007) used an interference paradigm to demonstrate that adult's performance on a nonverbal false belief task was impaired by a verbal shadowing task, but not by an equally-demanding tapping task. They concluded that false-belief reasoning relies on linguistic representations, consistent with de Villiers' proposal that the sentence complement structure provides a conceptual resource that is necessary for stable representation of mental states.

This interpretation depends on the assumption that verbal shadowing interferes only in linguistic processes. We know of no evidence that this is the case. The differential effect of verbal shadowing relative to tapping simply shows that there is some resource or representation which is shared by verbal shadowing and the false belief task, but not by tapping. Precisely what this resource is remains unclear.

Verbal shadowing involves perceiving and repeating spoken language (in this case meaningful sentences). The task does not require the comprehension of this verbal material. But decades of research in psycholinguistics has demonstrated that the language processor has little respect for task requirements. The use of representations at one level of linguistic analysis automatically activates linked representations at other levels. Thus when people hear, think about or produce a word (even a meaningless one) they activate not only the phonological form of this word but also all the words that share phonemes with that form, and the meanings of those phonologically-related words. Such effects can be observed even in young children (Huang &Snedeker, in press; Mani, 2009), suggesting that they are a fundamental feature of linguistic architecture. Thus verbal shadowing would be expected to produce interference not only at the phonological level but also at higher semantic or conceptual levels (even if the material being shadowed has no semantic content). Whether this higher-level interference is construed as linguistic is largely a theory-internal matter. If we assume that language-specific semantic representations are the only domain-general combinatorial conceptual system (Spelke, 2003), then we might declare such interference linguistic by default. Absent such an assumption, there is no reason to believe that the rapid cascade of activation spreading upward from phonology halts suddenly at a hypothetical boundary between language and thought (see e.g., Chambers, Tanenhaus & Magnuson, 2004; Chamber & Juan, 2008; Fischer & Zwaan, 2008; Altmann & Kamide, 2009 for evidence that any such boundary is quite permeable). Thus the list candidate loci for the interference effect in Newton and de Villiers should include not only phonological and semantic representations but also the conceptual processes involved in representing: events, the temporal sequencing of events, and the goals and knowledge states of social agents. It is disappointing, but no one promised that cognitive science was going to be easy.

Thus all of the data to date, including the present study, appears compatible with the language as trigger hypothesis.

Conclusions

The present study demonstrates that internationally-adopted children are better at low verbal theory of mind tasks than one would expect given their current linguistic abilities. This fact sheds light on the tight correlation that has been observed between language development and performance on explicit theory of mind tasks. If this reflected the online use of linguistic representations in solving theory of mind tasks, then we would expect that adopted children would fail these tasks from the time that they lose these representations in their birth language up until the point at which they acquired them in their second language. This does not appear to be the case: adopted children with current syntactic skills below the level of a four and half year old performed quite well on low verbal theory of mind tasks while language matched control children did not. Thus while language may serve as a trigger to the development of stable representations of false beliefs, the language skills developed at around four years of age do not appear to be necessary for the implementation of the relevant representations.

ACKNOWLEDGEMENTS

We thank Peter deVilliers and Jill deVilliers for generously sharing the testing materials they developed, Susan Carey for her encouragement and all of the families who generously shared their time. We are grateful to Alexa Scarlett, Shannon Shaper, and Nadia Chernyak for help in preparing the study materials and to Katie Felkins for her tireless assistance over several years. This research was supported by the National Science Foundation under Grant No. 0418423 and by the Roger Brown fund.

Tables

Table 1 Participant age and adoption information

	Adopted Group			Control Group		
	Mean	SD	range	Mean	SD	range
Age at Test	7.4 years	1.7	4.5-10.8	4.6 years	1	3.0-6.1
Age of Arrival in U.S.	5.5 years	1.8	2.4-10.2	n/a	n/a	n/a
Time in the U.S.	1.8 years	1.3	.3-6.1	n/a	n/a	n/a

Table 2 Standardized test scores and comparisons

	Adopted Group			Control Group			Comparison		
	M	SD	range	M	SD	range	*t*	df	*p*
Age at Test (yrs)	7.4	1.68	4.5-10.8	4.5	0.97	3.0-6.1	10.125	88	<.001
DELV raw	57.47	14.83	21-76	56.2	16.69	22-83	0.381	88	0.704
KBIT-II NV raw	18.42	6.09	7-34	12.91	4.83	3-26	43758	88	<.001
KBIT-II NV SS*	89.9	14.5	60-121	101.17	12.81	72-139	3.446	73	<.01
PPVT-III raw	71.51	17.5	37-125	62.09	18.34	26-101	2.494	88	0.015
PPVT-III SS	79.56	15.2	51-123	103.47	8.53	87-118	9.164	88	<.001

*Standard scores only available for participants age 4 and over

Table 3 Language measure correlations

	Adopted Group			Control Group		
	DELV	PPVT-III	Comp. Score	DELV	PPVT-III	Comp. Score
DELV raw	1			1		
PPVT-III raw	.667**	1		.732**	1	
Complements Score	.313*	0.159	1	.523**	.437**	1

Table 4 Regression of complements score with DELV and group as predictors

Measure (DV)	Total Variance	Predictors	β in final model
Complements Score	R^2=.234	DELV raw	0.418***
		Group	0.228*
		Group X DELV	*ns*

Table 5 The relationship between passing complements task and ToM scores

	ToM Measure	Sentence Complement Success Fail	Pass	*t*	df	*p*
Adopted Group		n=11	n=34			
	Low Verbal	46.5%	69.6%	-2.47	43	<.05
	Verbal	31.8%	53.7%	-1.87	43	0.07
Control Group		n=21	n=24			
	Low Verbal	30.7%	63.0%	-4.41	43	<.001
	Verbal	33.9%	60.0%	-2.47	43	<.05

Table 6 DELV and PPVT as predictors of ToM score

		Adopted Group		Control Group	
Measure (DV)	Predictor	Total Variance	β in final model	Total Variance	β in final model
Verbal ToM score	DELV raw	R^2=.348	.602***	R2=.467	.691***
	PPVT-III raw	R^2=.283	.547***	R2=.472	.696***
Low Verbal ToM score	DELV raw	R^2=.461	.688***	R2=.268	.533***
	PPVT-III raw	R^2=.186	.452**	R2=.308	.569***

Table 7 Language ability and group as predictors of ToM score

Measure (DV)	Total Variance	Predictors	β in final model
Verbal ToM score	R^2=.417	DELV raw	.650***
		Complements score	*ns*
		Group	*ns*
Low Verbal ToM score	R^2=.414	DELV raw	.512***
		Complements score	.169 (*p*=.07)
		Group	.209*

Table 8 DELV and Complements score as predictors of ToM in each group

	Adopted Group			Control Group		
Measure (DV)	Total Variance	Predictors	β in final model	Total Variance	Predictors	β in final model
Verbal ToM Score	R^2=.348	DELV raw	.602***	R^2=0.466	DELV raw	.691***
		Complements Score	*ns*		Complements Score	*ns*
Low Verbal ToM score	R^2=.461	DELV raw	0.688***	R^2=.299	DELV raw	.401*
		Complements Score	*ns*		Complements Score	.253 (*p*=.09)

Table 9 Experiment 2 Participant age and adoption information

	Mean	SD	range
Age at Test (yrs)	7.6	2.0	4.8-10.8
Age of Arrival in U.S.	6.5	2.2	2.8-10.2
Time in the U.S.	1.1	1.1	.33-2.6

References

Apperly, I.A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adult performance on a non-inferential theory of mind task. *Cognition, 106,* 1093-1108.

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

Birch, S. A. J. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18,* 382-386.

Bollard, P. M., Chute, P. M., Popp, A., & Parisier, S. C. (1999). Specific language growth in young children using the CLARION cochlear implant. *Annals of Otology, Rhinology, & Otolaryngology Supplement, 177,* 113-118.

Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development, 75*(4), 1155-1170.

Courtin, C., & Melot, A. (2005). Metacognitive development of deaf children: Lessons from the appearance-reality and false belief tasks. *Developmental Science, 8*(1), 16-25.

Custer, W. L. (1996). A comparison of young children's understanding of contradictory representations in pretense, memory, and belief. *Child Development, 67*(2), 678-688.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1), 1037-1060.

Dunn, L. M., & Dunn, E. S. (1997). *Peabody picture vocabulary test (3rd ed.)* . Circle Pines, MN: American Guidance Service.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483-527.

Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24(1 Suppl),* 46S-58S.

Geren, J., Shafto, C. L., & Snedeker, J. (in progress). Internationally adopted school-age children: When is their English proficient?

Geren, J. & Snedeker, J. (in progress). Syntactic development in deaf children with cochlear implants.

Gindis, B. (1999). Language-related issues for international adoptees and adoptive families. In T. Tepper, L. Hannon & D. Sandstrom (Eds.), *International adoption: Challenges and opportunities.* (pp. 98-107). Meadowlands, PA: First Edition.

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*(1), 26-37.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*(3), 346-359.

Jackson, A. L. (2001). Language facility and theory of mind development in deaf children. *Journal of Deaf Studies and Deaf Education, 6*(3), 161-176.

Kaufman, A. S., & Kaufman, N. L. (2004). *KBIT-2: Kaufman brief intelligence test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*(3), 732-739.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*(4), 1130-1144.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition, 105*(3), 489-512.

Miller, C. (2004). False belief and sentence complement performance in children with specific language impairment. *International Journal of Language and Communication Disorders, 39,* 191-213..

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, 77,* 622-646.

Newton, A. M. & de Villiers, J. G. (2007). Thinking while talking: Adults fail nonverbal false-belief reasoning. *Psychological Science, 18,* 574-579.

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex, 13*, 155-161.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125-137.

Pyers, J. E. (2005). *The relationship between language and false-belief understanding: Evidence from learners of an emerging sign language in Nicaragua.* Unpublished PhD, U California, Berkeley.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*(1), 12.

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.

Seymour, H., Roeper, T., & de Villiers, J. (2005) *The DELV-NR (Norm-referenced version). The diagnostic evaluation of language variation*. San Antonio: The Psychological Corporation.

Shafto, C. L., Geren, J., & Mervis, C. B. (submitted). Correlates of reading ability in children adopted from Eastern Europe: A preliminary investigation of effects of age of arrival in the U.S.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*(1), 117-141.

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*(6), 1789-1795.

Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition, 109*(3), 295-315.

Spencer, P. E. (2004). Individual differences in language performance after cochlear implantation at one to three years of age: Child, family, and linguistic factors. *Journal of Deaf Studies and Deaf Education, 9,* 395-412.

Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-yr-old children. *British Journal of Developmental Psychology, 17*(1), 141-155.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11,* 153-158.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology, 39*(5), 906-912.

Wellman, H. M., & Bartsch, K. (1988). Young children’s reasoning about beliefs. *Cognition, 30*, 239–277

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72,* 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Young, G. A. & Killen, D. H. (2002). Receptive and expressive language skills of children with five years of experience using a cochlear implant. *Annals of Otology, Rhinology, & Laryngology, 111,* 802-810.