On the relation between implicit and explicit Theory of Mind and linguistic competence - An empirical approach

Dissertation
zur Erlangung der Doktorwürde
durch den Promotionsausschuss Dr. phil.
der Universität Bremen

vorgelegt von
CHARLOTTE SONJA HERZMANN

08.11.2011
Erstgutachter: Prof. Dr. Dr. Manfred Herrmann
Zweitgutachter: Prof. Dr. Kerstin Konrad

Tag des Promotionskolloquiums: 14.05.2012
Le cœur a ses raisons que la raison ne connaît point.
The heart has its reasons, which Reason does not know.

Blaise Pascal, Pensées IV
On the relation between implicit and explicit Theory of Mind and linguistic competence  
- An empirical approach

**Table of contents**

**Acknowledgements**  
7

**Zusammenfassung**  
8

**Abstract**  
10

**Abbreviations**  
12

**List of Figures**  
13

**List of Tables**  
15

1. **Introduction**  
16

2. **Theoretical background**  
18

2.1 *Where does language come from?*  
18

2.1.1 Verbal behavior according to Skinner  
18

2.1.2 Nativism by Chomsky  
18

2.1.3 Tomasello’s usage-based account  
19

2.2 *What underlies knowledge of mental states?*  
19

2.2.1 A Theory for Theory of Mind  
20

2.2.2 Simulating others  
20

2.2.3 Theory of Mind Mechanism  
21

2.2.4 Mentalizing concept  
22

2.3 *How to tap Theory of Mind - the case of false belief*  
22

2.4 *Language and Theory of Mind - a complex affair*  
28

2.4.1 Mental state talk - the communicative hypothesis  
28

2.4.2 Complement on Theory of Mind - the parasitic hypothesis  
29

2.5 *Social and linguistic knowledge - a developmental timetable*  
31

2.5.1 Period 1: from birth to first year (0-12 months)  
32

2.5.2 Period 2: the second year (12-24 months)  
32

2.5.3 Period 3: the third year (24-36 months)  
33

2.5.4 Period 4: the fourth year (36-48 months)  
33

2.5.5 Period 5: the fifth year and beyond (from 48 months)  
34

2.6 *An integrative approach - the implicit-explicit distinction*  
35

2.6.1 The representational redescription model  
35

2.6.2 Theory of implicit knowledge  
37

2.6.3 Support from research: Early Theory of Mind  
39

2.6.4 Limiting factors of False Belief tasks measuring implicit ToM  
42
On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach

2.7 Aims and scope of the thesis
  2.7.1 Central questions and general considerations 45
  2.7.2 Hypotheses and goals 47

3. Methods
  3.1 Study participants 50
  3.2 Data protection, data security, and legal framework 51
  3.3 Experimental design and study procedure 52
    3.3.1 The False Belief task 52
    3.3.2 The Theory of Mind Inventory 55
    3.3.3 The Complement Comprehension Task 57
    3.3.4 The Reynell Developmental Language Scales 57
    3.3.5 The demographic questionnaire 58
  3.4 Strategies of data analysis 58
    3.4.1 Coding of demographic data 59
    3.4.2 False Belief task performance - analysis techniques 59
    3.4.3 Language measures - analysis techniques 61
    3.4.4 Theory of Mind Inventory - analysis techniques 62
    3.4.5 Associations between measures - analysis techniques 63

4. Results
  4.1 Demographic data and study participation 65
  4.2 Performance in the False Belief task 67
    4.2.1 Interrater-reliability analysis 67
    4.2.1 Action prediction question 67
    4.2.2 Looking behavior 70
    4.2.2.1 First look and summed longest look 70
    4.2.2.2 Looking time and shifts to each location 72
    4.2.3 Association of verbal and nonverbal False Belief task performance 75
  4.3 Language measures 76
    4.3.1 Knowledge of complements 76
    4.3.2 Developmental language level 77
    4.3.3 Association of CCT and RDLs scores 77
  4.4 Theory of Mind Inventory 78
    4.4.1 Demographic variables and Theory of Mind Inventory mean scores 78
    4.4.2 Comparison of ToMI scores with data from evaluation study by Hutchins et al. (in press) 79
    4.4.3 Factor analysis of Theory of Mind Inventory items 80
    4.4.4 Theory of Mind Inventory and the False Belief task 81
On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach

4.5 Comparison of Theory of Mind and language measures  
4.5.1 False Belief task performance and knowledge of complements  83  
4.5.2 General language ability and False Belief understanding  84  
4.5.3 Theory of Mind Inventory and language competence  86  

5. Discussion  90  
5.1 Data interpretation and integration into a theoretical framework of ToM development  90  
5.1.1 The "retelling" tendency in initial orientation  90  
5.1.2 Coding effects on anticipatory looking analyses and the disregarded distractor  93  
5.1.3 Control function of the True Belief condition  94  
5.1.4 Processing of the False Belief action prediction question  95  
5.1.5 Dual process models on ToM development  98  
5.1.6 Basic ToM accounts: theory, simulation, mentalizing but not mechanism  101  
5.1.7 The Theory of Mind Inventory as a valid informant measure  102  
5.1.8 Interpretation of "situational knowledge" and "linguistic knowledge" ToMI factors  103  
5.1.9 False Belief performance and language level  106  
5.1.10 The complex affaire of ToM and language revisited  107  
5.1.11 An integrative dual process model of ToM development and processing  108  

5.2 Present study's limitations, shortcomings, and consequences for future directions  110  
5.2.1 Universal pitfalls of behavioral studies  110  
5.2.2 Consideration of methodological constraints  111  
5.2.3 Specific limitations of the present study and perspectives for future research  111  

5.3 Final conclusions  114  

6. References  115  

7. Appendix  125  

8. Declaration of oath  160
On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach

Acknowledgements

There are many people that supported me in conducting and finishing this thesis.

First, I would like to thank my advisors and reviewers:

Prof. Dr. med. Dr. phil. Manfred Herrmann provided facilities, equipment, and contacts for planning and conducting this study, supervised my work, and believed in my competences. He endured my endless questions and taught me important lessons about scientific work. Without his continuous support this project would not have been possible - thank you!

Prof. Kenneth Wexler, Ph.D., introduced me to the research on language acquisition and many fruitful and long discussions actually gave rise to the idea for this thesis. The opportunity to work in his team and to gain insight into developmental research sharpened my wits for my own project. His encouraging words made me believe in myself and my ideas - thank you!

Prof. rer. nat. Kerstin Konrad played an important role for me being accepted to receive a scholarship of the German National Academic Foundation. Though not knowing me personally she agreed to serve as a reviewer for this thesis and she provided me with all the things that were required in the formal act of submit this piece of work - thank you!

This whole project would not have been possible without the ideational and financial support of the German National Academic Foundation (Studienstiftung des deutschen Volkes). I enjoyed interesting talks and fruitful discussions at meetings of Ph.D. students and tutorial group.

The following people significantly influenced the present study, either by assistance during testing and analyses, providing technical, methodological, statistical, and mental support, long discussions, or an imperturbable believe in me:

Visire Beqa, David Berron, Wil Beunen, Angela Chu, Angela Clauß, Sascha Clamer, Daniel Dilks, Ph.D.,
PD Dr. Thorsten Fehr, Dr. Linda Fröhlich, Dr. Sascha Frühholz, Dr. Daniela Galashan, Dr. Orlando Galashan, Lena Hartwich, Matthias Jahn, Jan Janssen, Mirijam Köhler, Roman Konstantinov, Rita Korsch, Helena Kromm, Dr. Stephan Miedl, Nadya Modyanova, Barbara Muntz, Tim Reeß, Dr. Gabriele Reichel-Kaczenski, Hanna Seekamp, Julia Siemann, Jun. Prof. Dr. Svenja Taubner, Dr. Sina Trautmann-Lengsfeld, Juliana Wiechert, Dr. Daniel Wiswede.

A special thank goes to my family for their endless love and believe:

Nils, Claudia, Ulli, Tina, Till, Matthias, Oskar, Maria, Dagmar, Jens, Jürgen, Anne.

I am grateful to all the participating kindergarten, daycares, nurseries, the Landesverband evangelischer Tageseinrichtungen für Kinder in Bremen, and most of all, to all the children and parents who all were so highly motivated - It was great fun and a pleasure!

This thesis is dedicated to my beloved grandparents...
Zusammenfassung

Die sogenannte Theory of Mind (ToM) bildet eine grundlegende Fähigkeit des menschlichen Sozialverhaltens und der Kommunikation. Sie ermöglicht dem Individuum eigene und fremde mentale Zustände in verschiedenen Kontexten und Situationen zu berücksichtigen. Es wird angenommen, dass die ToM sich in zwei Verarbeitungsebenen, die sogenannte implizite und explizite ToM Komponente, unterteilen lässt. Während letztere vermutlich auf sprachlichen Kompetenzen beruht und dem Bewusstsein zugänglich ist, wird die implizite ToM wesentlich früher erworben und zeigt sich im Spontanverhalten. Die Entwicklung der ToM wird meist mit dem sogenannten Test zum "Verständnis falscher Überzeugungen" (False Belief (FB) Test) untersucht. Es zeigte sich, dass das spontane Blickverhalten von Kindern ein frühes Verständnis wiederspiegelt und somit die implizite Komponente der ToM reflektieren könnte während sich auf verbaler Ebene erst im Alter von vier Jahren ein explizites Verständnis solcher Überzeugungen finden ließ. Ein Zusammenhang zwischen der verbalen Antwort im FB Test und dem sprachlichen Entwicklungsniveau wurde bereits in mehreren Studien gefunden. Mit der vorliegenden Arbeit wurde das Ziel verfolgt den Zusammenhang zwischen expliziter und impliziter ToM und sprachlicher Kompetenz zu untersuchen. Hierzu wurden 224 Kinder im Alter von zwei bis sechs Jahren mit verschiedenen Testverfahren untersucht: Ein FB Test mit drei anstatt zwei Antwortoptionen wurde durchgeführt und das spontane Blickverhalten und die verbale Antwort der Kinder erfasst. Zusätzlich wurde das Sprachverständnis und die Sprachproduktion mit einem allgemeinen Sprachentwicklungstest überprüft sowie das Verständnis für bestimmte syntaktische Regeln. Als Maß für die alltäglichen ToM Fähigkeiten der Kinder wurde ein umfassender Elternfragebogen erhoben. Die zentralen Fragestellungen der vorliegenden Arbeit lassen sich wie folgt zusammenfassen: a) lassen sich in vorherigen Studien beschriebene Blickverhaltensmuster in einem modifizierten FB Test replizieren, b) führen unterschiedliche Kodierungsmethoden des Blickverhaltens zu den gleichen Ergebnissen und kann somit das spontane Blickverhalten als Indikator für eine implizite ToM dienen, c) findet sich ein Zusammenhang zwischen dem sprachlichen Entwicklungsniveau der Kinder und der verbalen Antwort im FB Test, aber nicht zwischen sprachlichen Kompetenzen und Blickverhalten im FB Test und d) werden die implizite und die explizite Komponente der ToM im Elternfragebogen erfasst? Das spontane Blickverhalten im FB Test wurde auf vier verschiedene Arten kodiert (erster Blick, aufsummiert längerer Blick, Dauer der Blicke zu jeder Antworthitung, Anzahl der Blickbewegungen hin zu jeder Antworthitung). Die Ergebnisse der vorliegenden Arbeit zeigen, dass Kinder scheinbar die Geschichte des FB Tests mit ihren Augen nacherzählen und dass die Kodierung des
aufsummierten längsten Blicke eine ungenauere Methode darstellt als die Erfassung der Blickdauer zur jeder Antwortoption. Obwohl die Ergebnisse vorheriger Studien zum Blickverhalten nicht repliziert werden konnten, so könnten die höheren Anforderungen des angewandten FB Tests zu einer Maskierung des eigentlichen antizipatorischen Blickverhaltens geführt haben. Bezüglich des verbalen Antwortverhaltens im FB Test zeigte sich ein signifikanter Altersseffekt. Allerdings ergab sich für die jüngste Altersgruppe ein zunächst kontrointuitiv erscheinendes Muster: 30 Prozent der Zweijährigen bestanden die verbale Vorhersage des FB Tests. Möglicherweise können jüngere Kinder die scheinbar deklarative Aufgabe auch implizit verarbeiten so lange noch keine explizite Verarbeitungskomponente entwickelt wurde. Für die älteren Kinder fand sich ein signifikanter Zusammenhang zwischen verbalem Antwortverhalten im FB Test und dem sprachlichen Entwicklungsniveau, die auf eine mögliche reziproke Beziehung dieser beiden Kompetenzen im Entwicklungskontext hindeutet. Zusätzlich wurden mittels einer explorativen Faktorenanalyse auf Basis der Daten des Elternfragebogens zwei abhängigen Faktoren identifiziert, die jeweils die implizite und explizite Komponente der ToM wiederspiegeln könnten. Die Ergebnisse der vorliegenden Arbeit führten zur Formulierung eines integrativen dualen Verarbeitungsmodells der ToM und ihrer Entwicklung.
Abstract

Theory of Mind (ToM) lies at the core of human social behavior and communication. It enables the individual to appreciate mental states in problem solving and interactive situations. Dual process theories claim for the existence of two distinct components of ToM, namely, implicit and explicit representations. The latter is believed to depend on linguistic competence and to allow for conscious access while the former is hypothesized to be acquired earlier and to become evident in spontaneous, non-declarative behavior. Studies on ToM development mostly apply so called False Belief (FB) tasks and in regard to the dualistic model, anticipatory looking behavior seems to express early developing implicit ToM while explicit verbal mastery is observed around the age of four years. In addition, forced choice responses in FB task have been claimed to rely on children's language competences (general and specific, respectively). The present thesis aimed at an evaluation of the dual process model of ToM and its relation to children's language level. A total of 224 children aged two to six years were tested on a three options FB task, capturing spontaneous looking behavior and verbal prediction responses. In addition, children's expressive and receptive language level and comprehension of specific syntactic structures were assessed. As a measure of children's everyday ToM competences an informant questionnaire was given to parents. The key question of the present thesis were a) whether it is possible to replicate anticipatory looking patterns in a three options FB task, b) whether different coding methods of children's looking behavior would lead to the same results and thus, can be interpreted to measure implicit ToM, c) whether verbal prediction in the FB is associated with children's language level and syntactic competence while anticipation is not, and d) whether the parental questionnaire measures implicit and explicit aspects of ToM on latent factors. Children's anticipatory looking was coded and analyzed in four different ways (first look, direction of summed longest look, time spend looking to each location, number of gaze shifts to each location).

Data indicated a distinct looking pattern across all age groups in that children appeared to "retell" the FB story course with their eyes. Direction of summed longest look proofed to be an insensitive measure compared to time spent looking to each location. Though present data did not replicate earlier findings on anticipatory looking behavior it is possible that correct gazing patterns were shadowed by high task demands of the three options design. There was a significant age effect on children's verbal response in the FB task, however, 30 percent of the youngest group also mastered the forced choice prediction. This finding was interpreted to reflect the application of implicit ToM to the declarative task which only later is shadowed by explicit reasoning strategies. For older children, verbal FB task performance and language level were significantly associated, supporting the view of a
reciprocal facilitation during development. An explorative factor analysis revealed two correlated factors in the parental questionnaire which possibly reflect implicit and explicit aspects of ToM. Findings of the present thesis led to the formulation of an integrative dual process model of ToM development and reasoning.
### Abbreviations

*in alphabetic order*

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>CCT</td>
<td>Complements Comprehension Test</td>
</tr>
<tr>
<td>FB</td>
<td>False Belief</td>
</tr>
<tr>
<td>E level</td>
<td>Explicit level <em>(in the Representational Redescription model)</em></td>
</tr>
<tr>
<td>E1</td>
<td>Explicit 1 <em>(in the Representational Redescription model)</em></td>
</tr>
<tr>
<td>E2</td>
<td>Explicit 2 <em>(in the Representational Redescription model)</em></td>
</tr>
<tr>
<td>E3</td>
<td>Explicit 3 <em>(in the Representational Redescription model)</em></td>
</tr>
<tr>
<td>I level</td>
<td>Implicit level <em>(in the Representational Redescription model)</em></td>
</tr>
<tr>
<td>KMO</td>
<td>Kaiser-Meyer-Olkin criterion</td>
</tr>
<tr>
<td>MSA</td>
<td>Measure of sampling adequacy</td>
</tr>
<tr>
<td>RDSL</td>
<td>Reynell Developmental Language Scales</td>
</tr>
<tr>
<td>RR</td>
<td>Representational Redescription</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error of mean</td>
</tr>
<tr>
<td>SETK-2</td>
<td>Sprachentwicklungstest für Zweijährige Kinder</td>
</tr>
<tr>
<td>TB</td>
<td>True Belief</td>
</tr>
<tr>
<td>ToM</td>
<td>Theory of Mind</td>
</tr>
<tr>
<td>ToMI</td>
<td>Theory of Mind Inventory</td>
</tr>
<tr>
<td>ToMM</td>
<td>Theory of Mind Mechanism</td>
</tr>
<tr>
<td>WET</td>
<td>Wiener Entwicklungstest</td>
</tr>
</tbody>
</table>
List of Figures

FIGURE 1: EXAMPLE FOR THE STANDARD FALSE BELIEF (FB) TASK (ACCORDING TO BARON-COHEN, LESLIE & FRITH, 1985) 24

FIGURE 2: EXAMPLE FOR THE "SMARTIES" TASK (ACCORDING TO PERNER, LEEKAM AND WIMMER, 1987) 26

FIGURE 3: THE REPRESENTATIONAL REDescription (RR) MODEL ACCORDING TO KARMILOFF-SMITH (1992) 37

FIGURE 4: THE IMPLICIT-EXPLICIT KNOWLEDGE THEORY ACCORDING TO DIENES AND PERNER (1999) 38

FIGURE 5: TOTAL NUMBER OF CHILDREN IN EACH AGE GROUP, SEPARATED INTO MALES AND FEMALES 51

FIGURE 6: EXPERIMENTAL SETUP OF THE FALSE BELIEF TASK 54

FIGURE 7: PERCENT OF CORRECT RESPONSES TO VERBAL ACTION PREDICTION QUESTION IN FB TASK SEPARATELY DISPLAYED FOR GENDER IN EACH AGE GROUP AND TASK CONDITION (FB AND TB) 69

FIGURE 8: PERCENT OF CHILDREN LOOKING INITIALLY AND IN SUM LONGEST TO CORRECT, INCORRECT, AND DISTRACTOR LOCATION, RESPECTIVELY, SEPARATELY DISPLAYED FOR GENDER AND TASK CONDITION, ONLY INCLUDING CHILDREN WHO PASSED THE TB VERBAL PREDICTION QUESTION 71

FIGURE 9: MEAN TIME SPENT LOOKING (ERROR BARS INDICATE SE) TO CORRECT, INCORRECT, AND DISTRACTOR LOCATION IN EACH AGE GROUP FOR FALSE AND TRUE BELIEF CONDITION, RESPECTIVELY 73

FIGURE 10: PERCENT OF SHIFTS (AND SE INDICATED BY ERROR BARS) TO CORRECT, INCORRECT, AND DISTRACTOR LOCATION IN EACH AGE GROUP IN FALSE AND TRUE BELIEF CONDITION, RESPECTIVELY 74

FIGURE 11: MEAN TIME SPENT LOOKING TO EACH LOCATION (AND SE INDICATED BY ERROR BARS) SEPARATELY DISPLAYED FOR PERFORMANCE ON THE FB ACTION PREDICTION QUESTION FOR EACH AGE GROUP 76

FIGURE 12: MEAN SCORES (AND SE INDICATED BY ERROR BARS) ON CCT SUM, EXPRESSIVE AND RECEPTIVE RDLS FOR EACH AGE GROUP IN PERCENT 78

FIGURE 13: MEAN SCORE (WITH ERROR BARS INDICATING SE) ON TOMI IN EVALUATION STUDY (HUTCHINS ET AL., IN PRESS) AND PRESENT STUDY IN PERCENT 80

FIGURE 14: MEAN SCORES (ERROR BARS INDICATE SE) ON TOMI FACTOR 1 "SITUATIONAL KNOWLEDGE" AND TOMI FACTOR 2 "LINGUISTIC KNOWLEDGE" FOR ACTION PREDICTION RESPONSE IN FALSE BELIEF TASK, SEPARATELY DISPLAYED FOR AGE GROUPS 82

FIGURE 15: MEAN TIME SPENT LOOKING (ERROR BARS INDICATE SE) TO CORRECT, INCORRECT, AND DISTRACTOR LOCATION, SEPARATELY DISPLAYED FOR FALSE AND TRUE BELIEF CONDITION, AND CCT MASTERY 84
On the relation between implicit and explicit Theory of Mind and linguistic competence  
- An empirical approach

\textbf{Figure 16: Mean scores (error bars indicate SE) on RDSL expressive and receptive subscales in percent for action prediction response in False Belief condition, separately displayed for age groups}  
85

\textbf{Figure 17: Mean estimated regression factor scores (error bars indicate SE) on ToMI factor 1 "Situational knowledge" and ToMI factor 2 "Linguistic knowledge" for performance in CCT, separately displayed for each age group}  
86

\textbf{Figure 18: The fuzzy-trace model of reasoning according to Reyna and Brainerd (2011)}  
100

\textbf{Figure 19: The integrative dual process model of ToM}  
109
List of Tables

**TABLE 1: DEMOGRAPHIC DATA OF STUDY SAMPLE, SEPARATELY DISPLAYED FOR GENDER, AGE GROUPS, AND PERCENT OF AGE GROUPS**  
66

**TABLE 2: RESULTS OF MULTIPLE LINEAR REGRESSIONS FOR TOMI VARIABLES AND LANGUAGE MEASURES (N = 153)**  
88
1. Introduction

Research on social cognition has a long tradition in psychological sciences attracting increased attention within the last decades. The individual's perception of and interaction with its environment and inner processes are complex in their nature and require a broad and interdisciplinary approach for investigation. Researchers from various disciplines such as philosophy, neurology, linguistics, sociology, communication sciences, and psychology explore the neural, behavioral, and conceptual underpinnings of this complex construct. After all, the main question is: What is social cognition and how does it develop?

Cognition, generally speaking, forms the bridge between the individual and any state of the world. It encompasses conscious and unconscious processes such as perception, attention, language, executive functions, memory, mental imagery, emotions, the ability to infer other people's mental states, and many more. There exist various and heterogeneous theories and approaches to single processes and the general topic of cognition, some of them contradictory and competing in their background tradition, investigation strategies, and (behavioral) predictions. Trying to solve some of these constraints by narrowing the focus to socially relevant cognitive processes only leads to further discrepancies between these approaches. Human beings are social subjects and, therefore, each cognitive process may be analyzed in the light of social relevance, adding even more complexity by including a counterpart to the individual and its behavior. However, this does not keep researchers from addressing complex cognitive processes, but encourages them to take an interdisciplinary perspective on cognition and its development. Studies on the developmental course of cognition may yield insight into the general functioning of cognitive processes and their interrelation and dependencies during formation and later functioning.

A core feature of social cognition is the so called Theory of Mind (ToM). It is commonly defined as the capacity to impute mental states (i.e., beliefs, desires, and thoughts) to oneself and to others and to use this knowledge to predict and explain behavior. ToM has been studied extensively since the introduction of this term by Premack and Woodruff in 1978. Though it is spoken of as one single competence ToM is composed of many different capabilities such as pretense, appearance-reality distinction, visual perspective taking, understanding and production of mental state terms, deception, an understanding of the concepts of desire, intention, emotion and belief, and many more. The scope of this construct depends on the various approaches, some of them take social cognitive processes such as joint attention and empathy as early components of ToM, others describe them as possible precursors or parallel developing competencies. For instance, empathy
and ToM can be described to encompass different aspects of mental state inferences, in that the former includes the emotional part while the latter concerns metacognitive factors (Fonagy, 2008).

Language is another important social component of cognition. It enables humans to interact with each other and to share inner feelings and thoughts. The acquisition of mental state terms such as belief, hope, and wish is crucial for children's conscious dealing with inner emotions, mental processes and outer states of the world. Through the verbal path, children are enabled to grasp the interdependencies between thought and behavior and learn about social effectiveness in communication. Furthermore, communication offers the direct experience and training of self-efficacy in manipulating and directing verbal and social interchange.

Regarding the interrelations of cognitive processes there is the strong claim for a connection of the language domain and ToM. Taking a closer look at the main features of language, namely pragmatics, semantics, and syntax, it becomes obvious that they share some core functions with ToM, for instance, the property of “recursion”. In terms of ToM, “recursion” describes the capacity to include thoughts into thoughts and this process may be infinitely repeated (Whiten & Perner, 1991). The analogous potential is found for language in regard to embedded sentences. Another shared aspect is the requirement to cope with representations of the world being either true or false. Additionally, verbal expressions always carry a literal and a social meaning that may differ in their message but need to be processed in parallel during conversation. Semantics require a shared understanding of words and categories while syntax provides the grammatical structure to embed false and true representations. Thus, it seems reasonable to presume a relation between these two cognitive faculties.

The present thesis aims at a comparison of these two social cognitive competences in developmental context. Taking into account hypotheses for an implicit-explicit distinction of knowledge advanced by Karmiloff-Smith (1992) and Dienes and Perner (1999) and strong arguments for an unidirectional dependence of early ToM development on language (de Villers, 2007; Astington & Baird, 2005), a behavioral study with children aged two to six years has been carried out. Before going into the project details, a general introduction to major theories on language acquisition, ToM development and concept formation will be given in chapter two. Aims and scope of the present study are outlined at the end of chapter two, followed by methods in chapter three and results in chapter four. Finally, chapter five aims at an integration of the present study findings in the contemporary empirical and theoretical framework as well as a general discussion of the concept of ToM development and future directions.
2. Theoretical background

2.1 Where does language come from?

Behavioral studies of children’s talk resulted in a clear developmental timetable of receptive and productive language. Being exposed to language from birth on, infants appear to be specifically interested in social stimuli, for instance, they tend to turn their heads towards human voice at a few months of life even if the related face is not directly visible (Alegria & Noirot, 1978). Around the age of 10 months, they produce the first word and during the next three to six years they gain a profound linguistic knowledge (Guasti, 2002). Interestingly, children do not require explicit instruction to acquire their native language. In regard to the underlying mechanisms of language acquisition three major directions have emerged: the behaviorist, the nativist, and the social-cognitive account.

2.1.1 Verbal behavior according to Skinner

The behaviorist approach was advanced by Skinner (1986) and describes language development to rely on behavioral rules of imitation and reinforcement. Language is claimed to be just another form of human behavior. However, the behaviorist account was challenged and abandoned for the most part as it cannot account for the rich linguistic knowledge children acquire throughout development.

2.1.2 Nativism by Chomsky

Noam Chomsky introduced the notion of a “Universal Grammar”, that is, a grammatical scaffold that underlies all natural languages (Chomsky, 1968). Universal principles define the shared properties of all languages. Parameters, on the other hand, designate those properties that are specific to a certain language. For instance, sentences in all languages include subjects, nouns, and verbs while the concrete position of those elements in a phrase differs between languages. Chomsky claims for an innate mechanism, the so called language faculty, which endow children with a general linguistic structure and the ability to selectively choose the appropriate parameters from its “Universal Grammar” for the language it is exposed to (Guasti, 2002). Based on the recursion property of syntax and the poverty of stimulus argument, namely that there is only limited linguistic input and a finite number of components and, nevertheless, the system being capable of producing an infinite number of semantic-syntactic constructions, language acquisition in the nativist approach is claimed to consist of the setting of innate parameters and application of universal principles.
2.1.3 Tomasello’s usage-based account

Language acquisition, according to Michael Tomasello’s usage-based account, is based on the capacity to engage in social and communicative acts (Tomasello, 2000). Through observation and imitation strategies, children are able to discriminate linguistic units and to gain an understanding of a basic function of language, namely to jointly generate a space of shared attention and to manipulate the counterpart’s attentional focus in regard to an object or topic. By reproducing words and phrases they picked up from their environment, children do not simply imitate the linguistic structure but also the related meaning and underlying intentions and, hence, learn about the communicative function of these utterances. An early slot structure develops that encompasses combined structure-meaning relations and further experience leads to the abstraction of linguistic elements and enables the child to combine constituents in new expressions (Tomasello, 2000).

2.2 What underlies knowledge of mental states?

The general function of ToM is to enable its holder to infer opaque, internal mental states of another person as well as of oneself and to put this knowledge to use in predicting and explaining behavior. This definition is broadly accepted, however, the term itself may be too narrow in the sense that it includes the notion of a theory. Assuming a theoretical basis of this social competence would imply that individuals model part of the world in their mind creating an opinion on reality and postulating a set of rules. But is a theory really needed to understand mental states? This question underlies a controversial interdisciplinary debate and led to the formation of four major opponent positions: The Theory theorists, the simulationists, advocates for an innate ToM Mechanism, and the psychoanalysts’ explanation of mentalizing processes. Certainly, there are subversions of these distinct domains, even integrated Theory-Simulation accounts and combinations with Modularity, as well as additional, different claims about the underlying mechanisms of knowledge formation on mental states. The psychoanalytic theory stresses interpersonal experience in the formation of a representational system. Here, only a rough introduction to the four explanations of ToM will be given, with a focus on the functional mechanisms during childhood. This section closes with a short review of ToM tasks in developmental research.
2.2.1 A Theory for Theory of Mind

The Theory theory states that having a ToM is having a body of knowledge about mental states. This knowledge consists of rules on causal relations and forms the foundation for the interpretation of everyday experience. Based on this conceptual framework, statements about another person holding a mental state, for instance, an attitude towards a proposition (i.e., he thinks proposition; Davies & Stone, 1995), are possible. The theory includes knowledge about the condition of having a mental state as well as what beliefs, wishes, thoughts are about including their underlying conceptual structure and meaning.

The development of a ToM in the Theory theory view is believed to be based on changes in the availability of knowledge concerning psychological states and affaires and relies on a gradual process of knowledge accumulation until a full blown concept of ToM is acquired (Davies & Stone, 1995). The question on the origin of the developing capacity separates the Theory theorists in two fields: the nativists argue for an innately given psychological concept that follows an ontogenetic development with triggering experiences (i.e., Carruthers, 1996) while the non-nativists claim for an acquisition process through theorizing in order to organize the huge amount of experiences to a coherent framework (i.e., Wellman, 1991). Hence, nativists believe in an inborn unspecific theoretical knowledge that facilitates social interchange and through growth and specific experience progress to more sophisticated theories on mental notions (Carruthers, 1996). Non-nativists favor to compare the acquisition process to scientific progress, children are described as little scientists that form a theory on basis of the available data (Wellman, 1991).

2.2.2 Simulating others

From the Simulationists' point of view, the ability to infer other's mental states rests upon the imaginative identification with another person. The thought format, in contrast to the Theory theory, consists only of the individual's own propositional attitude (i.e., I think that proposition; Davies & Stone, 1995). The concrete process itself can be described by two different approaches: the so called Cartesian account and the radical view of simulation. Advocates of the former approach claim for a combination of a tacit, possibly inherent theory and the simulation process in that introspection underlies the understanding of other people's mental states (Goldman, 1995; Harris, 1995). By simulating the other person's situation, using the appropriate and relevant inputs, the resulting mental state and behavior can be read off and interpreted from a first-person point of view. This process requires on side of the subject the ability to entertain first person mental states and the
competence to conduct a simulation process without the need of a full blown (theoretical) concept of ToM. A more radical explanation of ToM processes states that there is no need for the self to be included in the simulation process (Gordon, 1996). In this regard, the mental representation of the single proposition builds the thought format (Davies & Stone, 1995) and ToM relies on basic object-level understanding and the ability to reconceptualize. The recognition of a mental state requires in a first step an understanding of the object as such and by mechanisms of reconceptualizing this object is acknowledged as having a mental location - this process being called an ascent routine (Gordon, 1996).

After all, the simulation process is claimed to develop throughout childhood by experience and practice, becoming gradually more adept and refined.

2.2.3 Theory of Mind Mechanism

Another explanation for ToM is given in the Theory of Mind Mechanism (ToMM) account. The major claim is that every human being is endowed with an innate and specialized mechanism or module that processes metacognitive contents. Again, two variants of the ToMM functioning are proposed: the modular view of ToM and the interaction with an universal selection processor. Both accounts share the assumptions that a) ToM does not develop with age but only the access to application of knowledge differs between ages, b) pretense is an early and elementary expression of innate ToM, and c) simple ToM generalizes mental state properties, for instance, that desires call for satisfaction and representations such as beliefs are true (Fodor, 1992; Leslie, 1991). The modular view, however, claims for a heuristic approach to mental state understanding and the existence of a specialized structure of the brain (Fodor, 1992). Children apply the simplest heuristic explanation based on metacognitive generalizations and their limited computational capacity while adults are able to take into account additional information about cognitive and situational factors (Fodor, 1992). Expertise in ToM, hence, relies on the access of computational resources and maturation of the ToM module. An expansion of this view is given by Leslie, Friedman and German (2004) who complement the innate ToMM with a universal, non-modular selection process. The ToMM is present from birth on and directs attention to relevant information but it is not until a mature selection processor has developed that false beliefs can be handled. That is, the mature selection processor enables the individual to overcome the default setting of beliefs to be true. Both approaches agree in the major assumption that there is no fundamental difference in the ToM of children and adults.
2.2.4 Mentalizing concept

Despite the accounts from experimental psychology and philosophy, there is an additional stance advanced by psychoanalytic theory, called “mentalization”. Fonagy (2008) defines “mentalization” as the capacity to understand intentional mental states, and this process functioning on a preconscious and imaginative level underlies human action. A major developmental goal in psychoanalytic theory is the organization and compilation of the self which relies on the competences of “mentalization”, affect regulation, and attention mechanisms. The capacity to mentalize develops during early childhood and, in contrast to the former cognitive accounts, the importance of affect regulation processes and secure attachment regulations are highlighted. Interpersonal experience is crucial for developing mental representations and creating a symbolic representational system. First of all, an internal attachment model is required to generate anticipation processes and to organize inner mental structures resulting in an interpretive function of interpersonal experience (Fonagy, 2008). Through marked mirroring processes by an adult caregiver, the infant learns to regulate inner affects, for instance, by associating negative emotions with regulating behaviors or coping strategies. In addition, pretense play enables the infant to relate symbolic representations and mental acts to reality. This process is divided in the psychic equivalence mode, which has the child experience mental content as being physical and real entities, and the pretend mode, decoupling representations from reality and allowing for experiencing various affects and modified mental content in imagination (Taubner, Nolte, Luyten & Fonagy, 2010). Again, interpersonal experience with marked mirroring processes enables the child to integrate these modes to a coherent concept of "mentalization" and the self.

In contemporary research, these accounts still play a significant role in explaining the general mechanisms and functions of mental state reasoning. Extensions, adaptations and integrations of these theories have led to a complex and multifaceted field of explanations. After this short outline of the major waves in ToM research, a review of tasks tapping ToM in childhood will be given.

2.3 How to tap Theory of Mind - the case of false belief

The term ToM compromises many different aspects or competencies related to the handling, initiation, and interpretation of mental processes. Therefore, operationalization for empirical research usually only assess certain parts of the complex domain. In addition, it is challenging to
invent a test design that exclusively assesses ToM capacities and excludes other cognitive competencies such as language and executive functions. On one hand, language plays a central role in communication and most tests rely on linguistic processes (i.e., verbal or written instruction, story-based tests, target questions and required verbal answers, linguistic processing of information). On the other hand, executive functions seem to underlie any test of cognitive functions (i.e., working memory capacity, that is, holding in mind certain information and manipulating it for the correct answer, follow story-based tests, and inhibition of salient but incorrect responses and stimuli). These are only some of the arguments and objections that have been put forward in regard to ToM assessment (i.e., Bloom & German, 2000; Roth & Leslie, 1998). Hence, when developing a ToM task or testing ToM abilities a central aim should be the reduction of additional demands put on the child, always keeping in mind the general influences and complex interdependencies with other cognitive processes.

Assessment of ToM subcomponents can be found, for instance, in Piaget's early work on cognitive development, especially the concepts of object permanence in the sensorimotor stage and the overcoming of egocentrism and the role of imitation during the preoperational stage (Piaget & Inhelder, 1973). Today, contributions from Piaget's theory are broadly acknowledged though the basic stage model of cognitive development has been largely revised. Interestingly, some experiments by Piaget resemble currently applied ToM tasks in their design (i.e., experiments on object permanence resulting in the A-not-B error on the one hand and the false belief task on the other).

Following the target article by Premack and Woodruff (1978) and related commentaries, an influential study on the representation of false beliefs in human children was published by Wimmer and Perner in 1983. They followed the assumption that ToM requires the ability to explicitly represent a situation, to take an attitude towards this representation and to understand the relation and possible differences between one's own and another person's proposition (Wimmer & Perner, 1983). This is particularly important if one's own perspective differs from the counterpart's point of view and, thus, the individual has to cope with two conflicting representations. The method to assess this specific component of ToM is called the False Belief task (FB), sometimes referred to as change-in-location or unexpected-transfer procedures. Basically, the standard FB consists of a story including two characters and a target object (see Fig. 1; according to Baron-Cohen, Leslie & Frith, 1985). Both characters are present in the first scene and subject A puts the target object at a certain location (i.e., a barrel) and then leaves. In the absence of A, character B takes the target object and hides it at another location (i.e., a box). At subject A's return, children are asked where they believe that
character A will look for the target object. The whole story is either acted out, presented in pictures or as a movie, or simply told verbally, providing the child with full information about the target’s transfer and location. As a result, the child has a representation about the actual location of the target object, but in order to answer the test question, the child has to inhibit its own perspective and take into account character A’s false belief. The child cannot rely on its own, in Piaget’s words, egocentric representation. The FB has been extensively used to study children’s ToM development, interpreting the ability to correctly anticipate the story character’s behavior as a strong indicator for an important step in mental state reasoning.

![Image of False Belief Task](https://example.com/figure1.png)

**Figure 1:** Example for the standard False Belief (FB) task (according to Baron-Cohen, Leslie & Frith, 1985). 1) Two story characters are introduced and A stores the target object at a certain location (i.e., football in barrel). 2) After A leaves the scene, character B takes the target object and puts it in another container (i.e., a box; so called change-in-location). 3) When A returns, he wants to retrieve the object and children need to indicate where he will look for the football.

In the early years of ToM research, mastery of the FB task relied on the comparison with the 50 percent chance level. Carpenter, Call and Tomasello (2002) have claimed for this interpretation to be misleading and, therefore, suggest to adopt an additional true belief (TB) control condition. In the TB condition, the story character A witnesses the transfer and therefore searches for the target object
Theoretical background

at the correct location. As a result, children can rely on their own representation when predicting the story outcome, because both mental models are in accordance. If children choose the wrong location in TB and the correct location in FB they may rely on a different reasoning or response strategy that accidentally coincides with the correct response in FB. Controlling FB performance by testing children's TB response supports the assumption that the child a) understands the story and related questions and b) appreciates the actual location of the target to match a true representation.

Beside various adaptations of the traditional FB, other tests of ToM have been published and applied in developmental research. Closely related to the concept of FB are the appearance-reality distinction and the unexpected-contents task (i.e., Flavell, Flavell & Green, 1983; Perner, Leekam & Wimmer, 1987). Appearances of objects or situations may be misleading due to their ambiguous, deceptive nature. Representations build upon first glance may require reevaluation and correction when resting upon ambiguous stimuli, thus, a false representation is experienced from a first-person perspective. In the so called 'smarties' task (Perner, Leekam & Wimmer, 1987), children are shown a box that presumably contains candy, indicated by pictures on the outside of the container (see Fig. 2). The actual contents of the box are some pencils, leading to a revised mental representation. The test questions focus on children's ability to report what they first thought was in the container, what really is in it, and what another person would believe to be in this box. Answers may serve the evaluation whether children can hold in mind two controversial representations of the world including the ability to memorize the earlier impression from a first person perspective and to assign the same first sight error to another person.
Theoretical background

Figure 2: Example for the “smarties” task (according to Perner, Leekam and Wimmer, 1987). In this unexpected contents task, children are first shown a smarties box and are asked to report what they believe to be inside (presumably smarties). In a next step, the box is opened and they actual content is shown to the child (i.e., pencils). After closing the box again, children are asked what they first believed to be in the box and to predict what another person would expect.

The tasks outlined above strongly rely on verbal instruction and require either verbal or voluntary behavioral responses such as pointing. In addition to these so called elicited-response or forced choice tasks, researchers recently applied methods that rely on children’s spontaneous behavior during the observation of a scene. As an indicator for spontaneous, non-declarative behavior there are two candidate patterns: pointing and preferential looking. Overall, there are convincing findings for pointing to strongly rely on declarative and communicative impulses (i.e., Liszkowski & Tomasello, 2011). Preferential looking, on the other hand, primarily does not have a communicative function but may be caused by a preference in children’s exploratory behavior due to initial conceptions and learning strategies (Spelke, 1985). In reviewing reasons to apply preferential looking methods in research on cognitive development, Spelke (1985) points out that visual exploration is an early developing capacity compared to other behavioral systems and that behavioral preferences with specific patterns are surprisingly reliable and stable. Infants’ active and exploratory looking patterns appear to be guided by initial concepts, that is, they tend to direct themselves towards certain objects and events (i.e., faces, moving objects), and by addressing those aspects they are able to extend their early knowledge by learning through observation. For instance, the novelty preference
Theoretical background

shows in looking reliably longer to new, unknown objects or events compared to habituated ones, leading to the assumption that infants are capable to detect unfamiliar aspects in their environment and that they may seek to understand them (Spelke, 1985). In ToM research, two spontaneous-response tasks have been applied: violation-of-expectation and anticipatory-looking. The former is based on the assumption that children build expectations of the outcome of a story or the future behavior of a person in a certain situation while the latter method focuses on children’s ability to orient their gaze to the location where they expect something to happen (He, Bolz & Baillargeon, 2011). Both tasks lend on the conception of the FB test. For instance, a violation-of-expectation task may consist of an actor who hides a toy in one of two boxes in full view of the child. In a next step a transfer occurs and the actor either comes to hold a true or a false belief about the location of the toy. Finally, the actor reaches in one of the two boxes and the scene ends. If children build an expectation of the outcome based on the inferred belief of the actor then they would look longer at the final scene if this expectation is violated, in accordance with the novelty preference (Onishi & Baillargeon, 2005). That is, if the actor is holding a false belief but reaches for the correct box in the final scene children should look reliably longer compared to the actor behaving in accordance to the expectation by reaching for the wrong location. The anticipatory-looking tasks, on the other hand, focuses on an earlier time point of the story. Originally, the task relies on the acted out FB test including only an additional behavioral measure to the declarative, response-eliciting question. After the transfer of the target object and shortly before the reappearance of the story character, the children’s looking behavior is analyzed. If children form an expectation for the character’s reappearance at one of the locations based on the inferred belief state then they should clearly direct their eyes to the container related to their anticipation (Clements & Perner, 1994).

Further ToM tests focus on specific aspects or related concepts such as desire reasoning (i.e., Wellman & Woolley, 1990), pretense (Leslie, 1987), and faux-pas (i.e., Baron-Cohen, O’Riordan, Stone, Jones & Plaisted, 1999). Though these tasks and their application in various studies yield interesting and important insights into early precursor or later high-level components of ToM, presenting designs and discussing their numerous results and interpretations would go far beyond the scope of this thesis.

As mentioned earlier, it is difficult to discriminate a single competence from the complex field of social cognition. There are early developing competencies and initial conceptions that may show in visual attention and playing habits (i.e., as-if actions) that may be interpreted as precursors to or early aspects of ToM. On the other hand, there are strong developmental interactions between different cognitive capacities, such as executive functions and language. In particular, language has
been claimed to play a crucial role in ToM and its development. An overview of some strong claims for the relation between language and ToM in developmental context will be presented in the next section.

2.4 Language and Theory of Mind - a complex affair

As noted above, there are strong claims for a relation between the language domain and ToM reasoning. Both significantly contribute to social interaction and form core functions of social cognition. Though each competence has a specified role and various separable theoretical accounts about its particular functioning, they share core properties and may even rely on each other in their input-output and processing modalities. Social interaction is based on communication, either verbal or non-verbal. Socially relevant signals, therefore, need to be processed in different, complementary domains to integrate the qualitatively different information and to result in reasonable interpretations and appropriate responses. Furthermore, in order to acquire adult proficiency in the social domain these two competencies may even require additional complex interplays during development. Research on language and ToM in childhood yields supporting data for a developmental relation, in particular, many studies report that language level or mastery of certain linguistic aspects predicts children's performance on the FB task (i.e., Astington & Jenkins, 1999; de Villiers & Pyers, 2002). According to Saxe and Baron-Cohen (2006), hypotheses about the association and dependencies of ToM and language during childhood can be divided into a communicative and a parasitic account which are briefly outlined in this section.

2.4.1 Mental state talk - the communicative hypothesis

Language includes a variety of competences, such as basal prosodic and related motor knowledge, linguistic aspects including syntax, semantics, and lexicon, as well as situational knowledge, discourse and pragmatic communicative competence. ToM, on the other hand, encompasses various mental states such as beliefs and desires. While language is directly observable, mental states do not have a common and reliable behavioral correlate but can only be captured by either deduce possible internal states from a person's behavior in certain circumstances or by listening to verbal reports of inner thoughts or feelings. It seems reasonable to assume that language may serve as a scaffold for ToM knowledge acquisition, granting access and insight into another person's mind and allowing to
represent and communicate one's own inner states. For instance, parents’ mental state talk has been shown to be significantly related to children’s later FB performance and use of mental state terms (Ruffman, Slade & Crowe, 2002). Additionally, language may be a vehicle to express representations that are true, that is, that match reality, and those that are hypothetical or counterfactual. However, ToM may facilitate certain aspects of language acquisition as the meaning of words may be figured out in social interactions. For instance, the usage-based theory by Tomasello (2000) takes the ability to engage in joint action to be necessary for acquiring language by mapping words and sentences on a primarily non-linguistic scaffold. It has been shown that engagement in joint attention in young children serves as a significant predictor of later receptive and expressive language and performance in ToM tasks (Charman, Baron-Cohen, Swettenham, Baird, Cox & Drew, 2000). In addition, empirical findings point to a reciprocal facilitation and interdependency between language components and ToM measures (Slade & Ruffman, 2005). There appears to be a universal representational system with common precursor abilities in social cognitive development that underlies both, ToM and language.

Longitudinal and training studies are in accordance with this statement, showing a strong relation between specific conversational experience, linguistic knowledge, and FB performance, highlighting the importance of general pragmatic competences (Astington & Jenkins, 1999; Lohmann & Tomasello, 2003). Interestingly, these studies report a significant role for certain aspects of syntactic knowledge, in particular sentential complementation appears to contribute independently to ToM development. Hypotheses that specific linguistic aspects play an important role in the acquisition of ToM knowledge are subsumed under the parasitic accounts.

2.4.2 Complement on Theory of Mind - the parasitic hypothesis

The most prominent and strongest parasitic account highlights the importance of a specific grammatical structure for ToM development, that is, sentential complements of mental state verbs. Discourse about mental states employs certain verbs (i.e., to think, to believe) as linguistic vehicles for expressing mental content. However, the acquisition of mental state verbs poses high demands on part of the individual. Concepts underlying verbs are harder to grasp than nouns which are directly perceivable objects that ordinarily do not change a lot. Verbs describing physical acts may be referenced to perception, but mental states themselves are hard to map to the world and have complex and abstract properties. Mental state verbs cannot be referenced directly to any perception of the world and their meaning offers a wide range of interpretations. It seems reasonable to assume
a ToM component to underlie the process of associating a verb with its referee in reality. However, it has been shown that certain linguistic components contribute even more to this mapping process (Papafragou, Cassidy & Gleitman, 2007). Verbs depicting mental content take a specific syntactical structure that allows referencing the properties to the perspective of the respective subject. Complements taken by mental state verbs are whole phrases that can be true or false compared to the world and, nonetheless, they express a true proposition of an individual (de Villiers, 2007). In turn, these grammatical structures index mental content and are claimed to provide the representational scaffold for reasoning about false beliefs. Several studies report mastery of propositional complement syntax to be a good predictor for later FB task performance (i.e., Astington & Jenkins, 1999; de Villiers & Pyers, 2002; Lohmann & Tomasello, 2003). The central claim is that FB reasoning is inseparably linked to this syntactical scaffold, offering the linguistic representation of true and false mental content not meaning that this knowledge must be consciously accessible (de Villiers, 2007). This view is given further support by studies focusing on deaf children’s development of ToM. Early learned sign language, that shares the linguistic features of verbal languages, resulted in a normal developmental course of ToM reasoning while it was delayed in non-native signers with a significant role of false complement syntax understanding (de Villiers, 2005). It is important to note, though, that most of these studies employed elicited-response measures for ToM and therefore, these claims may only be considered for a specific time point in development (around the age of four years; see chapter 2.5).

Overall, the developmental relation between ToM and language remains controversial. Sentential complements seem to have a considerable impact on FB reasoning but taking a more comprehensive look at the relation of both competences leads to the assumption that there is more to ToM development than linguistic complementation. The view of a reciprocal facilitation during development as proposed by Slade and Ruffman (2005) with varying pronunciation of each competence throughout their acquisition courses appears to be the best candidate to capture the complex affair.

In the next chapter, a composition of a developmental timetable of certain social cognitive, linguistic and maturational aspects during the first six years of life is attempted.
2.5 Social and linguistic knowledge - a developmental timetable

Human development relies on a complex interplay between environmental factors, increasing cognitive and motor abilities, body and brain growth, as well as genetic predispositions. As noted above, cognition itself consists of interwoven processes and for each aspect there exist various hypotheses and general developmental accounts trying to explain how the child accomplishes to become an adult with grown-up capacities and knowledge. In the recent decade, technological and methodological progress led to the emergence of a new research field, the functional brain development approach, adding information to the long tradition of behavioral studies. There seems to be a strong relation between the structural development of brain regions with growing connectivity and children's cognitive progress. Structural maturation includes, inter alia, proliferation of neurons, synaptogenesis, pruning, and myelinization of axons. Neural correlates of cognitive functions have been studied extensively in adults, leading to hypothesis for the infant brain. However, a direct link between behavioral performance and maturational level of a certain brain structure based on adult imaging studies seems fallacious as infant’s cognitive functioning must not accord to the adult capacity. Cognitive development may just as well follow a mixed course of domain general and domain specific processes. There are several theories on human functional brain development, for instance, the maturational view, arguing for the maturation of functionally specified regions in the brain and resulting cognitive capacities, while the skill-learning stance claims for the existence of a general learning mechanism also involved in adult skill acquisition, and the interactive specialization approach suggests that brain areas are competing and interacting for acquiring a specific function and, therefore, response patterns change during ontogeny (Johnson, 2011). Still, it is reasonable to assume a relation between brain maturation and cognitive functions, after all, the adult brain appears to have a more or less universal functional structure, whether organized in modules or functional circuits remains unsettled. In order to integrate the present project and its results into the complex developmental course, the next sections give a short summary of cognitive milestones acquired during an approximate period of age and the related maturation of brain structures. The focus lies on the postnatal development until the age of five years, as during these periods the strongest improvements in ToM and language competence can be observed.
2.5.1 Period 1: from birth to first year (0-12 months)

Infants are confronted with language from early on. The auditory systems has been shown to develop prenatally and leads to early hearing of tones and voices (Querleu, Renard, Versyp, Paris-Delrue & Crèpin, 1988). At birth, infants perceive their own voice for the first time and within a few weeks, they start to intentionally modulate their voice and to practice the intentional use of their mouth and tongue. The production of syllable-like sounds and babbling occurs around the age of six to eight months of life followed by the first word from the tenth month (Guasti, 2002). Early signs of perception of objects and the appreciation of their permanence becomes evident from four months of age (Baillargeon & DeVos, 1991) and infants demonstrate a growing recognition memory and novelty preference (Hayne, 2004) as well as an understanding of the self as a physical entity (Taubner, Nolte, Luyten & Fonagy, 2010). Early social cognitive capacities become evident around nine to twelve months, the so called nine-month revolution, promoting joint attention mechanisms such as imitation and following gaze directions (Tomasello, 2000; Tomasello, Carpenter, Call, Behne & Moll, 2005). Turning to postnatal brain maturation, synaptogenesis has been shown to increase around birth with a burst of synapse formation in the visual cortex and frontal areas around three months of age and a maximum volume is reached by the former during the first year while the latter peaks later in development (Johnson, 2001). Maturational processes of the brain including myelination, synaptogenesis, pruning, and increase of metabolism, appear to follow specific time courses with a general trajectory to start and peak first in primary sensory and motor cortices and pathways, followed by association areas (parietal, temporal, and dorsolateral occipital) and then turning to the frontal lobe (Tau & Peterson, 2010). At birth, the neonate brain is approximately 36 percent of the adult size and within the first year reaches 70 percent of the adult volume with the highest increase of subcortical and cerebellar structures (Knickmeyer, Gouttard, Kang, Evans, Wilber, Smith, Hamer, Lin, Gerig & Gilmore, 2008).

2.5.2 Period 2: the second year (12-24 months)

In regard to the language domain, the infant needs to develop techniques to identify single words and meaning from the stream-like input in order to include new words in its lexicon. After a slow progression until the eighteenth month follows the so called ‘word explosion’, resulting in an exponential growth of the lexicon and the differentiation of functional and lexical roles (Guasti, 2002). In this period, children start to produce verbs referring to physical acts (Papafragou, Cassidy & Gleitman, 2007). Information encoding becomes faster, infants exhibit increasing long-term recall
abilities and sensibility to more retrieval cues (Hayne, 2004). Towards the second birthday, infants start to engage in pretense play (Leslie, 1987; Taubner, Nolte, Luyten & Fonagy, 2010) and show an early understanding of desires and goal-directed behavior - at the age of 15 months, infants exhibit early false belief knowledge in expectation-violation tasks (Onishi & Baillargeon, 2005; Träuble, Marinović & Pauen, 2010). On the neuronal level, grey and white matter volumes increase fast until the age of two years with a pronunciation of grey matter growth (Matsuzawa, Matsui, Konishi, Noguchi, Gur, Bilker & Miyawaki, 2001) resulting in an adult like brain structure (Johnson, 2001). Still, the total brain volume only encompasses 83 percent of the adult size (Knickmeyer et al., 2008).

2.5.3 Period 3: the third year (24-36 months)

At about two to three years of age, children start to produce three word sentences including verbs, requiring the application of a grammatical system. Interesting to note are the characteristic errors at this age, children demonstrate an overgeneralization of certain syntactical rules and primarily use infinite verbs (Guasti, 2002). Memory capacity and encoding speed become more advanced (Hayne, 2004). In regard to mental state reasoning, children start to pass the anticipatory-looking FB task in this time period (Clements & Perner, 1994; Garnham & Ruffman, 2001; Southgate, Senju & Csibra, 2007). In regard to brain maturational processes, grey matter volume reaches adult level in temporal cortices while frontal areas still display an increase going beyond the adult size (Matsuzawa et al., 2001). The main fibre tracts are established around the third birthday (Johnson, 2001).

2.5.4 Period 4: the fourth year (36-48 months)

Although production of mental state verbs becomes evident a little bit earlier, it is not until their fourth year of life that children demonstrate their correct application and understanding (de Villiers, 2007; Papafragou, Cassidy & Gleitman, 2007). Interestingly, children do not pass the elicited-response FB tasks at this age. Children demonstrate a preference shift in this period from playing with adults to playing with peers (Fonagy, 2008), probably due to advanced verbal abilities and a growing understanding of other minds. Brain volume growth appears to be linear for white matter while grey matter increase now displays a regional specific pattern with peaks in late childhood (Giedd, Blumenthal, Jeffries, Castellanos, Liu, Zijdenbos, Paus, Evans & Rapoport, 1999).
2.5.5 Period 5: the fifth year and beyond (from 48 months)

After the basic principles of language (semantics, syntax, vocabulary) have been acquired during the first four years of life, linguistic knowledge becomes more advanced and refined throughout childhood and adolescence. Due to the growing capacity to handle and manipulate information, children develop an temporally extended concept of the self by integrating memories into a causal-temporal structure (Fonagy, 2008; Taubner, Nolte, Luyten & Fonagy, 2010). In regard to the ToM domain, children start to pass the standard FB tasks from their fourth birthday on and mastery of even higher and more complex structures becomes evident in the following years (i.e., Baron-Cohen, O’Riordan, Stone, Jones & Plaisted, 1999; Clements & Perner, 1994; Milligan, Astington & Dack, 2007). By five years of age, children’s total brain volume has reached approximately 90 percent of the adult size (Tau & Peterson, 2010), but processes of synaptogenesis, myelination, and increase of metabolism in all neural lobes continue until late adolescence. Grey matter volumes show a specific time course of reaching maximal size in different cortices between ten to sixteen years, while white matter continues to increase during adolescence (Giedd et al., 1999; Gogtay et al., 2004). Overall, the plateau phase, which means that grey matter volumes have reached their maximum at around 150 percent of adult size, was reported to be reached earlier in females than in males for frontal and parietal cortices (approximately one year difference) though total brain volume is usually larger for males (approximately 10 percent; Giedd et al., 1999; Tau & Peterson, 2010). These gender differences in brain size and maturation may be caused by hormonal processes (Giedd et al., 1999). After all, individual brain maturation seems to follow two trajectories: functional complexity and phylogenetic developmental time course. Brain areas associated with lower-order or basic functions (i.e., sensory-motor regions, primary visual cortex) mature first, followed by functionally higher-order regions (i.e., involved in language processing and attention), while areas associated with more complex and integrative functions seem to mature last (i.e., executive functions, object-recognition) (Gogtay et al., 2004). In addition, phylogenetically older cortical areas (i.e. inferior aspects of the temporal and frontal lobe) have been shown to develop ontogenetically early and maturation seems to proceed laterally from these regions (Gogtay et al., 2004).

The next section will turn to an integrative approach to the before mentioned claims on the relation between ToM and language, before introducing the main hypotheses and aims of the present project.
2.6 An integrative approach - the implicit-explicit distinction

In ToM research, the question arises how to interpret findings pointing to a strong relation between ToM reasoning and language acquisition and how to integrate this complex interaction in a broader theoretical and developmental framework. A promising approach offers the adoption of an implicit-explicit distinction of knowledge. Lending from other areas of cognitive research, explicit knowledge is defined as a representation that can be manipulated and that is accessible for the individual while its implicit counterpart mostly has a functional use (Dienes & Perner, 1999). Two prominent models including implicit and explicit knowledge aspects applicable to ToM research have been put forward: the representational redescription (RR) model by Karmiloff-Smith (1992) and Dienes and Perner's (1999, 2002) theory of implicit and explicit knowledge. These two approaches differ in their specific focus: while the RR model describes different levels of explicitness and the redescription of representations as such, the implicit and explicit model explains the representational elements and effects of their explicitness.

2.6.1 The representational redescription model

Karmiloff-Smith (1992) claims the transformation of intrasystem information to explicit knowledge to be a major process in cognitive development. That is, by inborn preferences and predispositions, certain environmental events gain more attention leading to the formation of representations that organize in so called domains, specific knowledge areas. Through the process of redescription, early implicit representations subsequently become explicit knowledge to the individual, starting within a domain and a later possible translation to other knowledge areas. Representational redescription is endogenously driven though external influences may trigger its occurrence and it forms a domain general process in that it follows the same course in each domain depending on the state of explicitness of the representations at a given time. Karmiloff-Smith (1992) proposes a developmental model based on recurrent phases within subdomains in contrast to a general and simultaneous change of different domains. During the first phase, the infant's attentional focus is on external information, gathering input to form so called representational adjuncts to certain domains leading to mastery of behavioral tasks. A focus shift takes place in the second phase: internal dynamics organize the incoming information into knowledge chunks in subdomains. Finally, in the third phase external information and internal representations are adjusted and harmonized (Karmiloff-Smith, 1992). Knowledge domains consist of representations that change throughout development through a process of redescription. The reconstruction of representations may occur at different time points,
in different subdomains, and leads to different levels of explicitness that may coexist in the mind. Procedural representations allow for rapid and effective processing of external stimuli and behavioral output, but its components are not available to the general cognitive system. The very basal representations are called implicit, the so called I level (Karmiloff-Smith, 1992). Explicitness is divided into three levels: explicit 1 (E1) representations result from a redescription and compression of implicit knowledge, though the interdomain availability is reached at the price of losing some information in the abstraction process. E1 representations are manipulable and related to other knowledge domains, and allow for alternative interpretations of incoming data, for instance, pretended actions and false beliefs. Whereas E1 representations are not under conscious control, the so called explicit 2 (E2) level allows for conscious reflection and, finally, through redescription these representations become available across domains and verbally accessible in the explicit 3 (E3) level (Karmiloff-Smith, 1992). Regarding knowledge formation, not every subdomain contains all levels of explicitness and those levels do not have to follow one another. For instance, through verbal instruction children may learn certain facts or features that are directly stored on the E3 level. In regard to ToM development, the RR model states that infants must first gain implicit behavioral mastery of truthful representations with a subsequent redescription to the E1 level in order to allow for the procedure to become manipulable. In a further step, E2 and E3 representations enable the child to consciously reflect a belief situation and to give verbal report. It seems that children at the age of four gained an understanding of propositional attitudes of mental statements and false beliefs, internally represented in a E2 and E3 format. Language, in the RR model, develops in the same fashion as all other cognitive competencies, through redescription processes, explication and increasing accessibility (Karmiloff-Smith, 1992). Only at the E3 representational level, language takes a special role in that knowledge on this level becomes available to verbal report, justification, and evaluation. Figure 3 contains a schematic and simplified illustration of the RR model, bearing in mind objections to capture the complex model as a whole in flowcharts expressed by Karmiloff-Smith (1992).
2.6.2 Theory of implicit knowledge

Promoted by findings that children demonstrate an early capacity for FB understanding (Clements & Perner, 1994; see chapter 2.3), Dienes and Perner (1999) put forward another hypothesis on the distinction of implicit and explicit knowledge. Knowledge is described to consist of propositional attitudes, that is, the individual forms representations of inner and outer states and situations and puts them in functional use to react or interact with the world. Importantly, this term does not imply that knowledge is represented in a language like fashion (Dienes & Perner, 1999). Language itself consists of implicitly and explicitly represented components and linguistic structure only has a specific duty in the verbalization of certain explicit representations. Representations have a functional role by indicating a particular event, state, object, or content, and creating a reference between themselves, the individual and the referee. Importantly, for a representation to constitute knowledge the individual needs to be able to form and possess a representation through a causal process which must be accurate and treated as a truthful reflection of reality by the individual (Dienes & Perner, 1999, 2002). Knowledge itself consists of certain elements, such as representational properties of a referring subject in reality in a certain temporal context, as well as the propositional attitude of the individual. These elements are usually implicitly represented until made explicit in a hierarchical order of elements (from property descriptions to the self being explicitly represented; see Fig. 4). Explicit knowledge means that something has a clear meaning while any further content of a representation is conveyed in an implicit manner, reaching beyond the

Figure 3: The Representational Redescription (RR) model according to Karmiloff-Smith (1992). Boxes represent different levels of implicit (I level) and explicit (E levels) knowledge, and the handling of representations at each level. Horizontal arrows indicate the process of redescription of knowledge to become accessible at the next level, whereby the last step from E2 to E3 is defined by verbal access only. Gray vertical arrows represent triggers for the redescription process. White vertical arrows indicate ways of knowledge formation which may take place at each level and does not need to pass through each redescription process.
Theoretical background

literal meaning and providing supporting information on context or conceptual features. For instance, the individual may explicitly represent properties of being a human without any concrete reference, leaving predication and the precise subject implicit. Predication of properties, alternatively, may be attributed to a subject without the explicit notion whether this is really a fact or just an opinion of the individual. According to Dienes and Perner (1999, 2002), for knowledge to become conscious and to be subject to voluntary control, all these components including the self to be holding this attitude need to be represented explicitly.

![Diagram](image)

**Figure 4: The implicit-explicit knowledge theory according to Dienes and Perner (1999).** Knowledge itself consists of certain elements, that is, the content, the attitude towards the content, and the self. Content of representations include specific (i.e. subject) and general (i.e. category) properties of the represented, the referring subject in reality, the predication of the properties to the subject, the factuality (i.e. whether the predication is fact or fiction), and the temporal context. The attitude towards the representational content is defined through the possession and accuracy of the representation, as well as judging it to be the case and to have a causal origin. The individual itself is the holder of the propositional attitude. Explicit representation of the self always contains explicit representations of the attitude and the content, while in turn content can be represented explicitly with attitude and self remaining implicit.

In regard to ToM, Dienes & Perner (1999) claim for two knowledge bases during development. The more advanced implicit knowledge is based on the abstraction of situational regularities and shows in spontaneous, non-declarative actions. A different processing underlies the later knowledge base that comes with a causal understanding of the conception of beliefs and supports declarative responses. The degree of declaration in behavior is associated with the extend of explicitness.
Theoretical background

Procedural knowledge underlies non-declarative actions and includes implicit notions of predication and factuality, resulting in an efficient but inflexible and consciously not accessible application. Declarative knowledge, on the other hand, is less efficient but more flexible and explicitly depict predication and factuality, allowing for conscious access (Dienes & Perner, 1999).

2.6.3 Support from research: Early Theory of Mind

Within the recent decade, several studies have been published on early, implicit ToM, starting with the influential study by Clements and Perner in 1994. Here, an anticipatory looking method was added to the standard FB task to capture children's early anticipation of the protagonist's reappearance (triggered by an anticipation prompt; see chapter 1.3 for a description of task design and procedure). Clements and Perner (1994) aimed at experimental support for an early component of ToM, possibly becoming evident in behavioral mastery of FB tasks preceding the explicit test question performance by analyzing children's looking behavior. They tested a total of 44 children aged two to four years and reported 80 percent of the children older than two years and eleven months to have passed the anticipation test but only the oldest group (four year olds) passing the explicit question (Clements & Perner, 1994). It is important to note that they chose direction of initial look after the anticipation prompt as measure of implicit ToM. In fact, first look has been argued to be an insensitive measure and subject to early self-correcting saccades while, in contrast, the analysis of looking behavior during a short time period after the anticipation prompt has been proposed as a more reliable indicator of implicit ToM (Garnham & Ruffman, 2001; Ruffman, Garnham & Rideout, 2001). Children's anticipatory looking response is coded into the time they spent looking at each location during the specified time period. Subsequent studies were able to provide further support for an early ToM using anticipatory looking measures and expectation-violation methods (i.e., Garnham & Ruffman, 2001; Onishi & Baillargeon, 2005; Träuble, Marionović & Pauen, 2010). In a nonverbal anticipation task design, Southgate, Senju and Csibra (2007) reported 85 percent out of 20 children as young as 25 months of age to initially focus and in sum look longer at the correct location.

However, hypotheses and empirical findings on an early, implicit ToM component have received many critique and constraining remarks have been put forward providing alternative explanations of the study findings. Different interpretations are based on various hypotheses such as the application of a seeing equals knowing rule, associative learning strategies, extended explicit-knowledge explanations including low confidence knowledge, verbal processing due to task design, and the question whether these findings can be transferred and generalized over different task designs and

39
types of information (i.e., de Villiers, 2007; Penn & Povinelly, 2007; Perner & Ruffman, 2005; Povinelli & Vonk, 2004). Over the recent years, some of these alternative interpretations have been empirically approached by adding control conditions and modifying task designs. In regard to the seeing equals knowing rule, for instance, the strong claim for children to predict the protagonist’s future behavior based upon the question whether the character had visual access to the scene has been ruled out in two studies by Garnham and Ruffman (2001) and Träuble, Marionović and Pauen (2010). The former study introduced a three options design to the anticipation task, claiming that if children apply a seeing equals knowing rule they would expect the protagonist to look for the target object in either of the two incorrect boxes and therefore direct their gaze at both locations. In addition, the task design allowed to further address the association mechanisms claim by having the protagonist looking in the third container after having stored the target in another box. If children associate the protagonist with the location he has been last before leaving the scene, then they would direct their look to this third container. However, 72 percent of 32 children aged three to four years displayed a correct anticipatory look in the FB condition but only 19 percent gave a correct verbal prediction (correct versus third location: p < .001, Wilcoxon-signed rank test; Garnham & Ruffman, 2001). Träuble, Marionović and Pauen (2010) added a second true belief condition to an expectation-violation method. In this additional condition, the actor did not have visual but manual access to the object transfer and, hence, had a true belief about the object’s location. In violation-of-expectation designs, each condition is presented with one consistent and one inconsistent outcome in regard to the actor’s belief and the child’s expectation, measuring how long children focus on the scene (see chapter 1.3 for a description of task design and procedure). Infants looked significantly longer if the actor searched for the object in the empty container in the above described manual access TB condition (N = 36, within comparison, t(11) = 3.35, p < .01, one-tailed; Träuble, Marionović & Pauen, 2010).

Explicit-knowledge-only hypotheses yield another explanation of children’s task performance and have been addressed in several studies (Garnham & Perner, 2001; Ruffman, Garnham, Import & Connolly, 2001; Southgate, Senju & Csibra, 2007): the dissociation between looking response and verbal, explicit behavior may result from either a lack of confidence, misinterpretation process, or temporal stacking. The lack of confidence hypothesis states that there may be two contradicting, explicit explanations available but the child distrusts the seemingly complicated correct one and due to this lack of confidence, the weaker response only translates to looking behavior. A modified version of the anticipation FB task served the evaluation of this claim, by introducing an additional fast behavioral response (Garnham & Perner, 2001). In order to get to one of the locations, the
Theoretical background

protagonist needs to come down one of two slides and the child is required to catch him with a mat. If the lack of confidence hypothesis applies to ToM reasoning, then there should be no difference between children’s verbal prediction and mat moving response as both require a certain amount of commitment. However, Garnham and Perner (2001) found children’s mat moving behavior to coincide with their anticipatory looks (McNemar’s $\chi^2 (1,47) = .05, p > .90$) and not with their verbal predictions (McNemar’s $\chi^2 (1,47) = 6.25, p < .02$). Further support for the rejection of the lack of confidence hypothesis was given by Ruffman, Garnham, Import and Connolly (2001) who asked children to bet on the outcome as a measure of degree of confidence and commitment. Analyzing performance of the younger children (mostly three year olds, N = 18) who did not pass the explicit prediction revealed 94 percent to bet exclusively on the wrong location though indicating the correct container with their eyes (Ruffman, Garnham, Import & Connolly, 2001). Both studies strongly imply that early ToM competence relies on unconscious, implicit knowledge. The question whether ToM competence in the prediction response is shadowed by the verbal nature of the task or results from a misinterpretation on side of the child was addressed by Garnham and Perner (2001). In the mat-moving design, children should be better in reflecting on their decision where to move the mat before executing the action compared to their verbal prediction if they misinterpret the explicit question to mean where the protagonist should look rather than where he will look. Garnham and Perner (2001) were able to replicate their findings from earlier experiments that children looked at and moved the mat to the correct location (about 65 percent correct for both conditions, N = 74) but most of them failed the action prediction and reflect mat move question (40 % correct, respectively; N = 74). In addition, they included an analysis of children’s response timing (spontaneous versus prompt required) for the mat moving, reflect mat moving and the prediction response. They reported a significant main effect for timing ($\chi^2 (1,71) = 8.19, p < .004$), specifically pronounced in the mat moving condition (only 35 percent of the prompted group chose the correct location; Garnham & Perner, 2001). Still, the temporal stacking hypothesis could not account for the clear difference between children’s correct anticipatory look and their failure on the explicit question as there remained an additional main effect of condition (for looking versus verbal response: $\chi^2 (1,52) = 3.38, p < .05$, one-tailed; for mat moving versus verbal response: $\chi^2 (1,38) = 3.37, p < .05$, one-tailed; Garnham & Perner, 2001).

Turning to the hypothesized relation between complementation syntax and ToM, Low (2010) examined these two competencies in three to four year olds. In three studies, children were administered an anticipation FB task, several standard ToM tasks, a complementation measure, and tests for verbal and nonverbal ability (Peabody Picture Vocabulary Test-III and Raven’s Colored
Matrices, respectively). Overall, about 87 percent of the children in each study (total N = 102) showed correct first looks and they looked significantly longer to the correct compared to the incorrect location (study 1: N = 24, F(1,23) = 8.02, p < .01; study 2: N = 36, F(1,35) = 16.62, p < .001; study 3: N = 42, F(1,41) = 60.14, p < .001; Low, 2010). A hierarchical regression analysis in study 1 revealed independent, significant effects of implicit ToM and complement mastery: While age, general verbal and nonverbal ability accounted for 49 percent of explicit ToM performance variance in a first step, eye gaze and complement mastery in subsequent blocks accounted for further 14 and 11 percent of variance in explicit FB task performance, respectively (similar effects were reported for study 2 and 3, respectively; Low, 2010). However, there was no significant bivariate correlation between implicit ToM measures and performance on the complementation task (study 1: r = .13, p > .05; study 2: r = .31, p > .05; study 3: r = .13, p > .05; Low, 2010).

### 2.6.4 Limiting factors of False Belief tasks measuring implicit ToM

The above outlined studies only form a selection of research on early ToM development. There appears to be convincing support for the claim of an implicit-explicit distinction in ToM development and a possible functional dissociation in the ToM domain. However, there are a number of limiting factors to the currently available approaches.

The standard FB task requires children to choose between two response options leading to an equal probability distribution and insensitivity to distinguish between at chance performance and different response strategies. The addition of another distractor may yield more information on underlying strategies and includes a statistically superior probability distribution.

Garnham and Ruffman (2001) used a three options design focusing on different interpretations of children’s anticipatory looking in the implicit-explicit FB task by offering different target locations. Albeit the confirming findings for looking behavior being interpretable as implicit ToM understanding outlined above, the study procedure yielded an important limitation: the object transfer during the story course only included the right and the left container. Though the protagonist was linked to two locations (i.e. performed some kind of action at the middle and the left box), the target object was never placed at the middle location. In this regard, the middle container did not have any relevant role in the story and it seems likely that the statistical superiority of the three options design does not apply. After all, it seems questionable why children should ever choose to look at the middle location at all.
Theoretical background

Samson, Apperly and Humphreys (2007) introduced a three response options belief understanding test based on the change-in-location and unexpected-contents tasks in a patient study. The specific task design allowed for a discrimination of different response strategies, assuming that perspective taking relies on the ability to inhibit one's own knowledge and to infer the other person's belief. Samson et al. (2007) presented participants a video sequence, showing a person who puts an object (a passport) in an empty pizza box witnessed by another character. Later, the first person takes out the passport and puts another object in the pizza box (scissors) which is not observed by the second character. Finally, participants had to report what the second character believes to be in the box. Beside the correct answer (a passport), Samson et al. (2007) describe the dissociation of two incorrect response alternatives: A reality-based response strategy may be used when inhibition mechanisms are damaged (response: scissors) or a simplified mentalizing strategy can be applied which includes inhibition of one’s own perspective but relies on inaccurate cues to infer the other person’s belief (choosing the box which depicts its content on the outside; response: pizza). The latter response indicates that participants did not take into account the full sequence of events. Up to date, a deceptive distractor has not been included in the assessment of implicit ToM in children.

Earlier studies applied different coding methods to children's looking behavior in FB tasks as indicators of implicit ToM, including direction of first look, direction of summed longest look (that is, to which direction did children looked longest summing up all looks during the target period), and time spent looking to each location. Most studies included only one scoring method in their analysis (i.e., Clements & Perner, 1994; Garnham & Perner, 2001; Granham & Ruffman, 2001; Ruffman, Garnham, Import & Connolly, 2001), except for two studies: Southgate et al. (2007) and Low (2010) assessed direction of initial look and time spent looking to each location. As noted above, first look has been criticized to be an insensitive measure as children may display self-correcting behavior, for instance, they may tend to initially look at the location where the object really is before turning to the correct location for the rest of the time; therefore, direction of summed longest look or time spent looking to each location have been preferred as more sensitive measures (Garnham & Perner, 2001; Garnham & Ruffman, 2001). On the contrary, Low (2010) argued that initial look may be a better indicator for implicit knowledge because it may be faster than children's inhibitory control and, therefore, first look could be the spontaneous accurate application of an implicit strategy that is corrected as soon as inhibitory processes come in. Studies assessing both, initial look and time spent looking, found children to show correct eye gaze in both measures (Southgate et al., 2007; Low, 2010). Still, direction of first look as well as summed longest look are coded as absolute values, resulting in less informative categorical variables. Coding time spent looking to each location, on the
other hand, extends the information offered by these categorical scores, allowing for a relational analysis of children’s looking behavior. It seems reasonable and necessary, in this regard, to include different metrics of children’s looking behavior to assess implicit ToM.

Finally, the specific period being analyzed in looking time studies requires further discussion. The majority of earlier studies included a predefined four seconds time period after the anticipation prompt in behavioral analysis. In this regard, two remarks need to be mentioned: Firstly, Garnham and Perner (2001) found the prompt not to be necessary to trigger children's spontaneous, anticipation-guided looking behavior and, thus, they took the end of the narrative before the reappearance of the protagonist as onset for the analysis. This shift or extension of the analysis period may be necessary to capture children's early looking responses, particularly for direction of first look. Secondly, earlier studies are lacking a clear explanation why they decided to include a four seconds time period in analysis. Assumably, this period was chosen to assure comparability between participants based on a fixed, equal time period for analysis. In addition and adopting arguments by Low (2010) in favor of first look measures, the primary application of implicit knowledge may be influenced by a later emerging tendency to actively indicate (i.e., verbal or pointing) the story outcome, that is, adopting an explicit ToM explanation. This explicit strategy may override the implicitly guided looking behavior. However, it is unclear at exactly when this possible explicit strategy comes into play and it seems arbitrary to choose a four seconds period.

Taking into account these limiting factors, further investigation of looking behavior as a valid and reliable measure of implicit ToM is required. In the next chapter, the central aim and major hypotheses of the present thesis are outlined.
Theoretical background

2.7 Aims and scope of the thesis

2.7.1 Central questions and general considerations

This thesis aims at a further evaluation of the concept and development of ToM in relation to linguistic competence. Based upon the explicit-implicit distinction of mindreading abilities by Karmiloff-Smith (1992) and Dienes and Perner (1992, 2002), and in consideration of the before mentioned limitations of earlier studies, the following questions arise:

- Do children display an implicit understanding of false beliefs by looking to the correct location (initial look, summed longest look, time spent looking to each location, respectively) if an additional incorrect location with an deceptive feature is added to the FB task?
- Do different coding methods of children’s looking behavior produce diverging or homogenous results? If they differ does this discrepancy rely on the insensitivity of categorical metrics or rather the existence of different underlying processes?
- Does children’s linguistic competence (developmental language level and/or knowledge of complementation, respectively) relate to implicit and explicit ToM capacities?

Development of an implicit ToM is believed to precede the explicit competence. Therefore, the study sample should cover an age range that is known to include the transition phase from failing the FB task to passing it. There are convincing findings that implicit competence can be captured in an anticipation design at 25 months of age at the earliest (Southgate, Senju & Csibra, 2007) while the majority of children tend to pass the explicit prediction after their fourth birthday (i.e., Clements & Perner, 1994; Träuble, Marionović & Pauen, 2010). In addition, this time period covers important milestones in children’s language acquisition (see chapter 2.5). Taking into account the fast and various developmental changes within this age span, it seems reasonable to include children between two to five years of age and to compute groups on a six-month step basis.

In the present study ToM capacity was approached from two sides: A direct FB task, which includes implicit and explicit measures of ToM, was used to elicit observable reactions from the child, and an indirect measure based on a parental questionnaire. The behavioral task consisted of a modified version of the three options design introduced by Garnham and Ruffman (2001) in that the target object was placed once at each location throughout the story course. This adaption resulted in an equal relevance of each location and adjusted probability of correct answers, possibly resulting in a more sensitive measure of children’s ToM capacities. In addition, the third container included a distracting and deceptive feature by showing a picture of the target object on the outside, allowing
for the interpretation of possible response strategies proposed by Samson, Apperly and Humphreys (2007). Children's responses were analyzed for verbal prediction and looking behavior, applying different coding methods for eye gaze outlined above. In addition, results in a four seconds predefined time period were compared to findings in individual analysis periods between the end of the narrative and the explicit prediction question.

With respect to the parental questionnaire, the Theory of Mind Inventory (ToMI; Hutchins, Prelock & Bonazinga, in press) was administered to provide indirect evidence for children's capacities. The questionnaire was developed in consideration of the limitations of the standard FB task and aimed at a broad assessment of the multi-faceted concept of ToM. It relies on the assumption that caregivers spend a lot of time with their children and, therefore, become experts in regard to their children's abilities including weaknesses and strengths, offering a rich source of information (see chapter 3.3.2 for detailed test description and information on the questionnaire's validity and reliability). Furthermore, it is hypothesized that implicit and explicit ToM capacity is assessed in ToMI that may be distinguishable in two latent factors.

Assessment of linguistic ability consisted of a general language task, the Reynell Developmental Language Scales (RDLS; Reynell, 1985) and the Complement Comprehension task (CCT; de Villiers & Pyers, 2002). The specific role and syntactical features of mental state verbs have been outlined above. They enable its user to express mental representations which may be true or false. It has been argued that mental state verb use and ToM ability are strongly correlated during development, mastery of both competencies becoming evident around the fourth birthday (i.e., Astington & Jenkins, 1999; de Villiers & Pyers, 2002). This finding is not surprising taking into account that both capacities rely on an appreciation of mental content that is oblique to direct observation. Interestingly, certain verbs of communication (i.e., say) share the syntactical property of sentential complements but lack the obliqueness of mental representations (de Villiers, 2007). If a person expresses a thought verbally there is no need to infer a mental state but a direct comparison between the person's expressed view and reality can be drawn. The CCT serves as a measure of children's understanding of tensed complements of verbs of communication. On the other hand, there has been the strong claim for a possible bidirectional association of ToM abilities and general language capacities (i.e., Astington & Jenkins, 1999; Lohmann & Tomasello, 2003). From early on, children are confronted with linguistic input and social interactions with major developmental changes in language and ToM taking place in approximately the same age span. In consideration of shared properties outlined above it seems reasonable to assume a (developmental) association of these competences. Since receptive vocabulary and linguistic competences precede expressive
capacities in early language acquisition (Tomasello, 2000) a developmental language task should measure receptive and expressive language competence in two separate subscales. In regard to the present study, the task should be available in a German version and it needs to be applicable to the age range of the target sample, that is, two to five years of age. This combination of features is particularly rare in developmental language tasks and narrowed down options to the German version of the Reynell Developmental Language Scales (RDLS; Reynell, 1985). The RDLS are a popular and extensively used measure in English-speaking countries and it has a good predictive validity in clinical settings for both, the receptive and the expressive subscale (Hagtvet & Hagtvet, 1990). There exist two official translations of the RDLS in German language (Edwards et al., 1997; Reynell, 1985) and for reasons of availability, the German adaption of the RDLS by Sarimski (Reynell, 1985) was used in the present study.

Certain demographic variables have been shown to be related to children’s ToM development and language acquisition as well as to parent’s ratings on the ToMl (Hutchins et al., in press; Pears & Moses, 2003; Perner, Ruffman & Leekam, 1994): the child’s chronological age, number of siblings and position of the tested child, languages spoken at home, parent’s level of education, annual combined household income, and amount of time parent’s spent with their child. A demographic questionnaire was developed for the present study, taking into account the before mentioned factors and following suggestions by Hutchins et al. (in press).

2.7.2 Hypotheses and goals

Based on earlier studies on implicit and explicit ToM using FB tasks described above, the following hypotheses were formulated:

If the implicit-explicit distinction applies to ToM and if eye movements are interpreted to reflect spontaneous behavior, the implicit component of ToM is hypothesized to show in children’s looking behavior. In an anticipation task, children are believed to look reliably longer to the correct location throughout the target time period, irrespective of number of distractors. Coding style of direction of longest look, that is, categorical scores (correct versus incorrect) and percent values of time spent looking to each location, should not differ but indicate the same response tendency. In addition, direction of first look and a further metric, namely, children’s shifting behavior, were assessed to check for their validity as a measure of implicit ToM. As mentioned above, children may display shifting behavior due to a self-correcting mechanism or a possible tendency to "retell" the story with
their eyes. After all, it is hypothesized that if children build an expectation about the protagonist’s reappearance at a certain location they may redirect their eyes at this specific location more frequently, probably to check whether something happened while they looked in another direction. If children initially look to the incorrect location and only turn to the correct box afterward then this would also be reflected in their shift-to behavior. If, on the contrary, children initially look to the correct location and only later shift to another location then this would support the claim for the spontaneous application of implicit reasoning mechanisms being overridden by later inhibitory control processes.

Taking into account arguments outlined above, time period for looking behavior analysis started at the end of the narrative and ended with the onset of the explicit prediction question and, in addition, four seconds and individual time periods were analyzed, respectively. Length of analysis period was hypothesized not to affect results of looking preferences. Based on earlier studies reporting very young children to pass looking measures in FB tasks (Southgate et al., 2007), implicit measure of ToM in the present study was expected to be independent of children’s chronological age. Explicit ToM understanding was hypothesized to be highly correlated with children’s age replicating earlier findings (i.e., Garnham & Perner, 2001; Milligan, Astington & Dack, 2007). Implicit and explicit FB task performance allowed for an analysis of different response strategies (according to Samson et al., 2007) by having the target object always starting at the deceptive distractor location. If children chose the distractor (verbally or with their eyes, respectively) than this would speak for the application of a simple mentalizing strategy (simply relying on the visual cue while inhibiting or avoiding their own knowledge of the true location of the object).

The ToMI has been shown to be a good measure of children’s everyday ToM competences (Hutchins et al., in press) but the implicit-explicit distinction has not yet been applied to the questionnaire. Implicit ToM has been claimed to show only in non-declarative, spontaneous behavior and eye gaze seems to serve as a good measure of this competence. Still, direct testing can only capture children’s behavior in a particular situation at a certain time point. The application of implicit ToM reasoning, however, may not show in an experimental setting but in everyday social situations yielding more information for the child about circumstances and participating subjects. In this regard, the ToMI offers the opportunity to assess parents’ perception of their children’s behavior in everyday situations, covering early through late developing aspects of ToM understanding compared to mere belief reasoning captured by FB task. The present study aimed at an evaluation of the ToMI, claiming for the questionnaire to possibly cover implicit and explicit ToM competences in underlying latent
Theoretical background

factors. Therefore, ToMI score is hypothesized to be correlated with performance on the FB task (implicit and explicit, respectively).

Additionally, the relation between ToM competence and language level will be addressed taking into account the communicational and parasitic hypotheses, respectively. In regard to the former hypothesis, a strong two-way interaction of children's developmental language level and explicit FB task performance is expected, supporting the claim for reciprocal facilitation by Slade and Ruffman (2005). The parasitic approach, on the other hand, leads to the hypothesis that children's explicit ToM builds upon specific language capacities, that is, comprehension of tensed sentential complements (de Villiers & Pyers, 2002; de Villiers, 2007). Implicit ToM, however, is hypothesized to be independent of language acquisition and linguistic processing, therefore, language measures were expected not to be correlated with looking behavior in the present study.

The main hypotheses can be subsumed as follows:

1) An early, implicit ToM understanding becomes evident in children's looking behavior in the FB task under the age of four years.

2) Coding methods for children's looking behavior do not result in different response tendencies and, therefore, can be claimed to measure the same mental process.

3) Explicit understanding of ToM shows in children's performance on the action prediction question in the FB and is strongly related to chronological age and linguistic competences (general language and complement comprehension, respectively).

4) The ToMI serves as an informant measure of children's ToM capacities and includes implicit and explicit components of ToM.
3. Methods

3.1 Study participants

Participants were recruited from fourteen different kindergartens and nursery schools in the area of Bremen, Germany. Day care institutions were either contacted by phone or email, followed by a short presentation of the project at internal meetings of the institution members and parent-teacher conferences. Day care managers, teachers, and parents received an information form containing a brief outline of the project and contact information. In order to participate in the study, parents had to fill in a consent form and declare whether they want their child to be tested and whether the parents agreed to fill in a parental questionnaire. Participation amounted to a mean of 40 percent of all children in each day care (ranging from 15 to 100 percent). Based on parents’ reports, no child was ever diagnosed with any physical or cognitive disabilities and all children had German as their native language.

In order to define the optimal sample size, an a priori power analysis was carried out using G*Power 3.1 (Faul, Erdfelder, Buchner & Lang, 2009) for three different comparisons based on the hypotheses described in the previous chapter and expecting a moderate effect (following Low, 2010):

For the comparison of grammatical competence (ordinal scaling) and ToM performance (categorical/dichotomous and interval scaling), the optimal sample size for a moderate effect ($f = 0.25; \eta^2 = 0.06$) and for seven subgroups (defined by age in six-month steps from two to five years) resulted in a total of 210 children (30 children per age group; parameters of F-test: $\alpha = 0.05$, $\text{Power} = 0.95$, df = 1, critical $F = 3.89$).

For the analysis of associations between implicit ToM variables (categorical/dichotomous scaling) and age and gender effects, the optimal sample size for a moderate effect ($\omega = 0.3$) resulted in a total of 172 children (parameters of $\chi^2$-test - goodness of fit: $\alpha = 0.05$, $\text{Power} = 0.95$, df = 2, critical $\chi^2 = 5.99$).

For the analysis of implicit ToM variables (interval scaling), a priori power analysis was calculated for a moderate effect ($d_z = 0.3$), using a nonparametric test with $\alpha$-level adjustment for multiple comparisons, leading to an optimal sample size of 211 children (Wilcoxon signed-rank test, two tailed: $\alpha = 0.01$, $\text{Power} = 0.95$, critical $t = 2.6$).

The final study sample consisted of 224 children (111 male, 113 female) aged two years (2;00) to six years and eight months (6;08) with an average of 32 children per age group (range 24 to 42; see
Methods

Figure 5). Chronological age was calculated in years, months, and days between the day of birth and the day of testing. Partitioning the sample into age groups was based on chronological age and each group encompassed children within a six months age frame, that is, age group two involved children from their second birthday until they were two years, five months, and 30 days old, while age group two years and six month span children aged two years and six months to two years and eleven months and 30 days, and so forth. Gender distribution within each age group did not differ significantly but indicated a trend of more females in older age groups ($\chi^2 = 11.28, p = .08$). Posteriori power analyses of the three comparisons resulted in a power of at least .93 for each comparison.

Figure 5: Total number of children in each age group, separated into males and females.

3.2 Data protection, data security, and legal framework

The study design, experimental procedure including all tests and questionnaires, the consent and the information form, as well as the handling of the collected data were in line with the ethics guidelines of the University of Bremen and based on the standards of the declaration of Helsinki (2004). The project was approved by the ethics committee of the University of Bremen (see Appendix A). Children were only tested if parents gave their written informed consent prior to testing and if children verbally agreed to participate. Parents were informed about data protection and data security of all experimental and personal data belonging to their child (see Appendices B1 and B2). Parents and children were informed about their right to quit the experiment during the entire course of testing, and that their participation is completely voluntary, and that terminating the study would
not result in any negative effects. Collected data, including videotapes, were assigned pseudonyms and have been securely stored separate from any identifying information.

3.3 Experimental design and study procedure

In this explorative study, a cross sectional design was chosen to assess and compare ToM abilities and language competence in different age groups. The study sample included seven age groups (two to five years) of approximately the same size and with an equal gender distribution, resulting in a quasi-randomized sampling strategy. Assessment included a direct and an indirect measure of Theory of Mind (False Belief (FB) task and Theory of Mind Inventory (ToMI), respectively) and two direct language tasks (Reynell Developmental Language Scales (RDLS) and Complements Comprehension Test (CCT)). In addition, a parental questionnaire was administered to control for demographic covariates.

Testing sessions took place in a quiet room in the respective kindergarten or day care institution. After a short warm-up period, each child was tested individually in one single session on all tasks described below. Order of tasks during assessment was fixed, starting with the FB task, followed by the CCT, and the RDLS. This order was chosen for reasons of attentional resources, the FB task required children to pay closer attention in order to elicit possible looking responses, while the language measures allowed for the experimenter to insert short breaks or playing intervals. Conditions of FB task (FB and TB) were administered in randomized order and took approximately five minutes per condition. Assessment of CCT required approximately ten minutes, while the largest amount of time was necessary for the RDLS (approximately 30 minutes), resulting in a total session duration of approximately 50 minutes per child. Children received a sticker after testing as well as a research assistant diploma and parents were given a short written behavioral observation report.

Each method and its application in the present study will be briefly described.

3.3.1 The False Belief task

A model of a mouse house (40 cm high and 120 cm wide) was made out of cardboard with three cut out holes as doors (30 cm apart; see Figure 6). Each door had a hook-and-loop tape on the back to attach a little curtain in order to avoid direct view into the house. Three cardboard boxes (5 x 5 x 5 cm) were placed in front of each door. Boxes and curtains belonging together had the same color (green, yellow, and white, respectively) to assure a strong association for each location. Due to the
Methods

hook-and-loop tape, allocation of colored curtains and associated boxes could be randomly assigned to the three locations. Throughout story course, the target object was placed into each of the containers at the three locations in order to have the same association between container and object for each location. The target object was either a toy piece of cheese or a toy candy which were randomly assigned to the first and the second story. The white cardboard box depicted the respective target object for each story on its outside (visual distractor). The color was chosen to avoid side effects due to associations with the color of the target object and to make contours of the picture more visible. Story characters were two mouse puppets (approximately 12 cm high), named Sam and Katie. Behind the house model, a video camera was set up right above the middle door as an anchor position (Sony TRV 250E, digital video camera recorder). The camera was adjusted to capture participants head in full view in order to record eye movements to each location during task assessment. The experimenter was situated behind the model and narrated the story while acting out the story events in front of the child. Each child received two story versions with story order being counterbalanced across study sample, one to test children's false belief (FB) understanding and the other to control for their true belief (TB) performance. At the beginning of each story, the child was introduced to Sam (protagonist) and his house, explaining that the three doors led to a central area inside the house. It was made sure that children understood that Sam could enter the house through any door and leave again through any other door (comprehension 1 questions). This was important because otherwise children might have thought that there are three separate rooms behind the doors which would influence the expectation of Sam’s reappearance after entering the house. Furthermore, the experimenter explained that if Sam wants to have a look in a certain container he uses the nearby door with the corresponding colored curtain. To assure children’s understanding they had to indicate where Sam would reappear if he wants to have a look in the left, right, and the middle container, respectively, which was acted out by the experimenter showing the children the empty boxes. The white container depicting a picture of the target object on the outside was always introduced last. Before opening the box, the child was adverted to the picture and only after this procedure the container was opened and the object was retrieved. In a next step, Sam put the target object in one of the other two boxes and entered the house through the corresponding door. In the FB story, Sam stayed behind the house and slept, in the TB condition, Sam reappeared on top of the house on his balcony where he could oversee the whole scenery. As a memory control (memory 1) children were asked to indicate where Sam had put the target object. Only if children answered this control question correctly, the experimenter proceeded - otherwise the story was repeated. Afterwards, the second story character, Katie, entered the scene. In the TB condition, she shortly talked to Sam, in the FB story she directly found the target object, put it in the third container
and left the scene. At this point the experimenter posed the anticipation prompt ("Sam is hungry for his cheese/candy. I wonder where he is going to look for his cheese/candy?"). A period of approximately 10 seconds after the anticipation prompt was later analyzed for children's looking behavior based on video recordings. During this period, the experimenter looked down to the floor to avoid eye contact with the child and any possible gaze following behaviors. Following that delay, the experimenter posed the action prediction question ("Now, Sam wants to get his cheese/candy. Where will he look for his cheese/candy?") and three further control questions (memory 2, memory 3, comprehension 2). In addition, children were asked to explain their response (explanation question). For a verbatim narrative of the TB and the FB story see Appendix C.

![Experimental setup of the False Belief task](image)

**Figure 6: Experimental setup of the False Belief task.** The child was seated in front of the house model. The house had three doors with colored curtains attached by a hook-and-loop tape, and one box of the same color stood in front of each door. The white box depicted the target object on its outside (distractor). The experimenter stood behind the model and acted out the false and true belief stories using two toy mice, Sam and Katie. Behind the model, centered above the middle door, a video camera captured the child’s head in full view for later analysis of looking behavior.

All children had to pass the comprehension 1 and memory 1 questions, otherwise the story was repeated. This was necessary for five children who passed both questions after the replay. All other control questions were coded as correct, incorrect, or no answer and the explanations were
transcribed from the video recordings. For younger children (age groups 2 to 3) who may have lacked necessary language capacities to cope control questions scoring for no answer was necessary.

Children’s verbal or pointing response to the action prediction question were coded as pass or fail whereby the latter was further divided into indication of incorrect location (in the FB story the actual location of the target, in the TB story the container where Sam had put the object first, respectively) or distractor location (white box).

Looking behavior was scored from the video recordings on a frame-by-frame basis (frame rate: 25 frames per second) using Virtual Dub (Virtual Dub 1.9.8, Lee, A., www.virtualdub.org). Two different time periods were analyzed: For both analyses, onset was the first mentioning of the target object in the anticipation prompt ("Sam is hungry for his cheese/candy. I wonder where he is going to look for his cheese/candy?"). Offset for analysis one was the start of the action prediction question ("Now, Sam is going to get his cheese/candy") while analysis two included only four seconds after the onset.

In analysis one, looking behavior for each child was adjusted to the individual time period and converted into percent scores. Comparisons of both analysis methods were carried out to evaluate the effect of defining a fixed versus individual time period for analysis. Differences in results of these analysis strategies would require a further investigation and discussion of the before mentioned possible underlying mechanisms.

Based on earlier studies, direction of initial look and direction of summed longest look over the whole period were coded as correct, incorrect, or distractor response (i.e., Clements & Perner 1994; Low, 2010). In addition, order and number of looks to each location, summed time spent looking to each location (in percent), and total time period for adjustment in analysis one were assessed. Order and number of looks were used to calculate the number of shifts from one location to another.

Following the main hypotheses of the present study, children are expected to look significantly longer to the correct location if they have an implicit ToM that shows in spontaneous looking behavior.

Taken together, the following measures of children's looking behavior were applied (for each story condition, respectively, and separately coded for individual and 4 seconds period analysis except for first look): Direction of first look, direction of summed longest look, percent of time spent looking to each location, and percent of shifts to each location.

\subsection{3.3.2 The Theory of Mind Inventory}

The Theory of Mind Inventory (ToMI; Hutchins et al., in press) is an informant, indirect measure of children’s ToM competences based on parental report. In the present project, the ToMI was
translated to German, closely oriented at the English version only introducing minor adjustments to cultural or linguistic specifications of German language (see Appendices D1 and D2). The task consists of 48 statements, each with a response continuum of 20 metric units, ranging from definitely true to definitely not applicable to the child and a center point for “don’t know” (Hutchins et al., in press). Continuums exactly matched the original version, in that the 20 metric units equaled 6.75 inches on an American 30 feet per inch engineering scale, in order to avoid possible effects caused by variations in line widths due to differences in German and American units of lengths (inch versus centimeters). Starting and end point of the continuum corresponded to absolute disagreement or agreement (0 and 20 points, respectively) and the center for indecisiveness (“don’t know” anchored at 10 points). The first page included instructions and three examples for indicating different degrees of confidence. Parents were instructed to read each statement carefully and place a marking at the continuum according to the degree to which the statement was applicable to their child’s behavior. Parents received the ToMI on the same day their child was tested on the FB task and language tests and were asked to complete the questionnaire at home within two weeks. Single items scores for each child were assessed using the American engineering ruler (0 - 20 points) and mean scores were calculated for each child.

The measure’s psychometric properties have been evaluated in a study of 124 American children aged two to twelve years who were administered a ToM task battery and whose caregivers completed the ToMI (Hutchins et al., in press). ToMI scores correlated positively with children’s age (accounting for 52 percent of the variation of ToMI scores; $r = .72, p < .05$) and with ToM task battery performance (ToMI scores accounted for 67 percent of ToM task battery performance; $r = .82, p < .05$; Hutchins et al., in press). The ToMI showed a very high internal consistency (Cronbach’s $\alpha = .98$) and a test-retest evaluation after an interval of two and twelve weeks revealed high reliability ($r = .89, p < .01$; Hutchins et al., in press). Ceiling effects only became evident for children aged nine to eleven years. One might argue that the informant nature of the ToMI yields the risk that parents overestimate their children’s capacities. However, this effect may be smaller than expected. In a study by Deimann, Kastner-Koller, Benka, Kainz and Schmidt (2005) parents predictions of their child’s performance on the Wiener Entwicklungstest (WET) only deviated to a small degree from the children's actual scores (differences of 0.3 of standard deviation for typically developing children). Hutchins et al. (in press) report strong correlations between ToMI scores and ToM task scores which further support the measure’s predictive validity. After all, the ToMI investigates children’s mindreading abilities from a family-centered perspective with a focus on real-life social situations and displays high sensitivity to different stages in ToM development in contrast to the standard FB task.
3.3.3 The Complement Comprehension Task

The Complement Comprehension test (CCT; de Villiers & Pyers, 2002) was carefully translated into German, oriented at the wording of the German sentential complements test described by Lohmann and Tomasello (2003). The CCT consists of ten picture sets, each depicting a short story on two photographs (see Appendix E). Participants were shown the first picture accompanied by a phrase describing what one of the depicted characters was saying (i.e., "The teacher says the girl has a bug in her hair"). A second picture proved this statement to be false which was supported by another phrase of the experimenter (i.e., "But it really is a leaf"). Pointing back to the first picture the experimenter asked the child to report what the person on the first picture had said. The answer required the child to merely represent the content of the first statement by holding the complement phrase in mind. Replies were scored for each item as pass or fail, resulting in a maximum score of ten points. In addition, following the recommendation by de Villiers (personal communication, March 1st, 2009), only sum values of at least nine points counted as passing the CCT.

3.3.4 The Reynell Developmental Language Scales

The Reynell Developmental Language Scales (RDLS; Reynell, 1985) measures children's receptive language capacity in 66 items with increasing level of difficulty using child-appropriate test materials. Items include simple reactions (i.e., pointing, verbal answers, manual combination, manipulation or displacement of materials) to certain word patterns and objects, more complex syntactical sentence structures, negations, and expressions concerning non-observable processes or phenomena. There exist two versions of the receptive scale, one for typically developing children and one for children with developmental delay including minimal motor responses. In the present study, only the former version was applied. Expressive language capacity is measured in three subscales: structure, vocabulary, and content. The structure scale includes a rating of children's structural linguistic abilities from first vocalizations to complex structures including main and subordinate clauses and scores 0 to 21 points. Vocabulary forms a block of prompts and questions testing children's knowledge of object names and meanings (semantic concepts), increasing difficulty by presentation mode (real objects, picture of objects, names of objects) and object label (objects to properties), scoring 0 to 22 points. Finally, the content subscale tests children's capacity to describe complex situations in their own words using drawings of familiar scenes at home (i.e., setting up the table), starting with a familiarization picture and three test items. Children's responses are rated for general description in a main clause, content, and number of additional sentences, over all scoring 0 to 24 points. Expressive language scores based on the sum scores of the three subscales (0 to 67 points).
Order of presentation and questions was maintained as determined in task manual, but object arrangement for each block was counterbalanced across children. Raw scores were converted into percent values for analysis.

Until today, the RDLS lack a German standardization and scores can only be compared to the English norms. Still, the German RDLS have been shown to have a good validity and reliability compared to other German developmental language measures (i.e., correlation with "Sprachentwicklungstest für Zweijährige Kinder - SETK-2" of r ≥ .7; Sachse, Anke & von Suchodoletz, 2007).

### 3.3.5 The demographic questionnaire

General information was gathered about family status (married/partnership, single parent, other) and whom the child was living with (biological parent/s, adoptive caregiver/s, other) to control for study sample homogeneity (see Appendix F). Parent's educational level and annual household income was scored on a four-points scale from low to high. In addition, parent’s had to indicate working hours (full-time to at home) and the average time they spend with their child during a regular week, not counting the time the child was asleep (in hours). Languages that were spoken at home as well as information whether the child ever received language training were assessed to control for possible effects. Finally, parents were asked to rate the developmental status of their child in comparison to their peers in five categories (mental, motor, social, autonomy, and language), according to the demographic questionnaire used by Hutchins et al. (in press). Overall, the demographic questionnaire included fourteen items. Parent’s received the demographic questionnaire on the day their child was tested and were asked to return it within two weeks.

### 3.4 Strategies of data analysis

Statistical analyses were accomplished using the statistical analysis software PASW 18 (SPSS Inc., Chicago, Illinois, USA) and the spreadsheet program Excel (Microsoft, Redmont, Washington, USA). Results of five measures were analyzed: the demographic questionnaire yielded general information about the study sample, the FB task assessed children's verbal and looking behavior as measures of ToM knowledge, children's understanding of specific syntactical complements was evaluated using the CCT, the RDLS served as a measure of children’s expressive and receptive developmental language level, and the ToMI assessed parent’s perception of their children’s ToM ability in everyday social situations. Statistical analyses were carried out for each measure separately before addressing
possible associations between performances in different tasks or scores in questionnaires. All statistical analyses were based on a significance level of $\alpha = .05$ or a Bonferroni corrected level of $\alpha' = .007$ for whole group comparisons ($\alpha' = \alpha/7$ age groups). Uncorrected $\alpha$ levels were applied if not otherwise indicated.

3.4.1 Coding of demographic data

Of the fourteen items of the demographic questionnaire, item 3 (date of birth) was used to calculate the exact chronological age at time of testing. Gender (item 4), family status (item 5), whom the child lived with (item 6), and whether the child had received a language training before (item 13) were coded on a two- or three-point scale and entered as categorical variables for analysis. A categorical variable indicating whether the child had an older sibling or not was computed by merging information in item 11 (number of siblings and and the child’s position). In accordance, item 12 (languages spoken at home) was entered as a dichotomous variable indicating whether the child was raised mono- or bilingual. Items seven and eight (parent's educational level and occupation, respectively) were separately coded for mothers and fathers and entered as ordinal variables ranging from low to high (no educational status to graduation from university and being at home to full-time jobs, respectively). Combined annual household income (item 9) was accordingly entered as ordinal data ranging from low to high. Percent values of each item for each age group were calculated. Item 14 (relative developmental status) was coded on a three-point scale (ranging from delayed to advanced compared to peers) referring to five asked categories (mental development, motor skills, social behavior, autonomy, and linguistic level). A ratio scaled variable contained hours spent with the child during a week (item 10) as stated by parents and was tested for normal distribution. Central tendencies for items 10 and 14 were calculated.

Analysis of demographic variables and possible associations with other task performances or scores only included children for whom the demographic questionnaire was returned. Missing responses for single items were coded as missing values for analysis.

3.4.2 False Belief task performance - analysis techniques

In order to allow for an interpretation of FB task performance to possibly reflect the application of ToM knowledge, analysis only included children who passed the action prediction question in the TB control condition (for explanation see chapter 2.3 and Carpenter et al., 2002).
Performance in FB and TB action prediction question, as well as direction of first look and summed longest look were coded on a three-point scale: one point referred to the respective correct location, zero indicated the incorrect location, and minus one related to distractor location. These categorical variables were used to calculate percent of children choosing the correct, incorrect, or distractor location, respectively. For statistical analysis and comparisons of response options, these variables were dichotomized resulting in three variables for each measure for each condition (action prediction, first look, and summed longest look, respectively). Dichotomous variables contained correct and incorrect responses, correct and distractor responses, or incorrect and distractor responses, respectively, with the omitted response option being entered as a missing value.

Chi-squared tests of goodness of fit were used for dichotomous FB and TB action prediction, direction of first look, and direction of summed longest look variables, respectively, to evaluate possible preferences compared to uniform distribution on a Bonferroni corrected $\alpha' = .007$. Possible differences between age groups were analyzed using Pearson’s chi-square and Somer’s d statistic with task performance as dependent variable. In addition, possible effects due to condition order (FB before TB or vice versa), gender, and demographic differences were assessed with Pearson’s chi-square tests.

Time spent looking and number of shifts to each location in FB and TB condition, respectively, were converted into percent values and entered as ratio scaled variables. Kolmogorov-Smirnov tests for looking time variables became significant ($z \geq 1.6$, $p < .01$), therefore, nonparametric statistics were used for analysis. Mann-Whitney-U and Kruskal-Wallis tests were applied to assess possible gender, condition order, and age effects. Wilcoxon signed rank tests were applied for comparisons of choices of each location using a Bonferroni corrected significance level of $\alpha' = .007$ for whole sample analysis, and $\alpha = .05$ for single comparisons within one age group. However, analysis of variance (ANOVA) are known to be rather robust against violations of underlying assumptions, thus, ANOVAs were applied to complement nonparametric statistics because they allow for post hoc analyses of possible group differences. If required, Games-Howell post hoc tests were applied using an uncorrected significance level of $\alpha = .05$.

Wilcoxon signed-rank tests were used for comparisons of the two different time periods being analyzed for looking behavior (individual vs. standardized).

Possible associations between performance on the action prediction question and looking behavior were evaluated using Pearson’s chi-square (for categorical variables) and Kruskal-Wallis tests (for ratio scaled variables) applying a Bonferroni corrected significance level of $\alpha' = .007$ for whole group
analyses. For comparisons in the TB condition, analyses included all children who completed the FB assessment (N = 222), while for the FB condition, only children who passed the TB action prediction question were included (N = 160).

Children who did not complete both the FB and TB condition assessment were discarded from analysis. For looking behavior, missing values were assigned to children who were not looking to any of the three task locations.

3.4.3 Language measures - analysis techniques

CCT performance was coded for mastery on a two-point scale (CCT pass criterion: at least nine correct out of ten) and, additionally, on a ten-point scale based on sum score. Effects of age, gender, and demographic variables on CCT performance were analyzed using Pearson’s chi-square and Somer’s d statistics.

RDLS assessment resulted in two variables (raw scores): receptive and expressive language, whereby the latter consisted of the sum score of content, structure, and vocabulary measures. RDLS raw scores were converted into percent values and analyzed for normal distribution using Kolmogorov-Smirnov tests (expressive RDLS: z = 1.19, n.s.; RDLS receptive: z = 1.95, p < .01). In order to address coherence of expressive language scores, Spearman’s rank and partial correlations (controlled for age) were assessed between RDLS subscales content, structure, and vocabulary. Effects of age, gender, and demographic variables on expressive RDLS score were assessed using oneway ANOVAs, with Duncan post hoc analyses for age effects, and for subscales (content, structure, and vocabulary) using Pearson’s chi-square and Somer’s d statistics with Bonferroni corrected significance level (α' = .007). For receptive RDLS score, effects of age, gender, and demographic variables were assessed using Kruskal-Wallis tests, and age effects were evaluated using Games-Howell post hoc tests. Partial correlations (controlled for age) were computed for evaluation of receptive and expressive RDLS scores association. In addition, comparisons between receptive and expressive RDLS performance within each age group were assessed using Wilcoxon signed-rank tests.

For correlation analysis of language measures, CCT sum score was converted into percent value and checked for normal distribution using Kolmogorov-Smirnov test (z = 3.06, p < .001). Associations between RDLS and CCT performance were addressed using partial correlations (controlled for age) and Wilcoxon signed-rank tests within each age group.
3.4.4 Theory of Mind Inventory - analysis techniques

In a first step, a reliability analysis was carried out to evaluate the ToMI items internal consistency based on Cronbach’s α criteria (including item difficulty and corrected item-total-correlation). Subsequently, ToMI scores of all 48 items were summed and a total mean score was computed. Missing responses for single items were coded as missing values for total mean calculation. For the rare case that parents have set two markings on response scales for single items mean scores between both markings were entered as response. Total mean scores were converted into percent values and Kolmogorov-Smirnov test assessed normal distribution (z = .86, n.s.). One way ANOVAs were used to address possible effects by age, gender, and demographic variables including post hoc analyses if applicable. In addition, effects of age and gender were evaluated on single item basis using multivariate ANOVAs. Spearman’s rank correlation coefficients were assessed for a possible effect of average time parents spent with their children on total mean ToMI score.

In order to evaluate the validity of the translated version of the ToMI used in the present study, total mean performance was compared to results of the evaluation study by Hutcheson et al. (in press). Mean scores and standard deviations from the evaluation study were entered for two to five year olds and checked for normal distribution using Kolmogorov-Smirnov test (z = 2.81, p < .001). A Kruskal-Wallis test was assessed to evaluate age effects. Comparisons of the two studies (mean total ToMI score of evaluation study vs. present study) were based on Spearman’s rank correlation (applying Bonferroni corrected α level) and Wilcoxon signed-rank tests for each age group.

In order to address the hypothesis that ToMI includes implicit and explicit components of ToM, an explorative factor analysis was carried out using principal-axis factoring method. Missing values were replaced by mean scores. In a first step, measure of sampling adequacy (MSA) was carried out for single items. Subsequently, Bartlett’s test of sphericity and Kaiser-Meyer-Olkin criterion were assessed for the remaining items to evaluate the quality of the factor analysis. Due to the large number of items, the Kaiser-criterion was not applied (which identified 13 factors with eigenvalue > 1) but factor extraction was based on the Cattell’s scree test resulting in the extraction of two factors. In order to increase eigenvalues and the explained total variance, and in regard to the association between possible latent implicit and explicit ToM components, an oblique rotation method (promax λ = 4) was applied. Item-factor correlations were taken from structure matrix and items were either assigned to correlate more with one single factor or equally with both factors, respectively. Variables containing standardized estimated regression factor scores were separately computed. Assigned items were qualitatively inspected for associations in regard to contents for further interpretation. In a first step, ToMI items were clustered into eight dimensions (items were
assigned to one or more cluster): Basis of belief formation, beliefs affecting behavior, means underlying actions, interpretation of signals (i.e., facial expressions), understanding of linguistic expressions (i.e., single verbs or abstract expressions), empathy, pretend, and desires. Single items were subsequently assigned to either factor based on correlation coefficients in structure matrix. Finally, items and associated clusters were inspected for qualitative differences between ToMI factors.

In a next step, Kolmogorov-Smirnov tests were assessed to evaluate normal distribution of variables containing standardized estimated regression factor scores (factor 1: z = .76, n.s.; factor 2: z = 1.4, p = .04). For factor 1, an univariate ANOVA was used to evaluate possible age and gender effects and Duncan post hoc analyses were computed if applicable. For factor 2, possible age and gender effects were addressed using Kurskal-Wallis tests and a Games-Howell post hoc analysis was carried out for multiple comparisons between age groups. Comparisons of factors within each age group were based on Wilcoxon signed-rank tests.

3.4.5 Associations between measures - analysis techniques

In a first step, associations between measures of ToM competence were addressed. Again, the data set was filtered for children who passed the TB control condition and for whom the ToMI has been returned (N = 137). Oneway ANOVAs were assessed to evaluate possible effects of action prediction performance in FB task (three-point scale coding) on total mean ToMI scores, ToMI factor 1, and ToMI factor 2 (Levene’s test for equality of variances: ω ≤ 1.03, n.s.). Subsequently, age was defined as covariate and included in oneway ANOVAs and mean values of ToMI measures were computed separately for FB action prediction responses within each age group. For possible effects of direction of first look and summed longest look in FB task on ToMI measures oneway ANOVAs were assessed. Comparisons of time spent looking to either location in FB task with ToMI measures were based on Spearman’s rank correlation coefficients due to looking variables not being normally distributed (see above), applying a Bonferroni corrected level of significance.

Associations between CCT and TB action prediction performance only included children who had completed CCT assessment (N = 218). For comparisons of CCT performance and FB condition measures, data was further filtered for children who had passed the TB action prediction question (N = 197). Associations between action prediction and CCT mastery was addressed using Somer’s d statistic with action prediction as dependent variable. In addition, a binary logistic regression
applying a stepwise inclusion method was carried out to evaluate the possible predictive role of CCT mastery on FB task performance and vice versa, based on the controversy about the relation of ToM and language in development. Merging incorrect and distractor responses resulted in a binary distribution of FB prediction which was included as dependent variable and age and CCT mastery were entered as independent predictors for the first binary logistic regression model, while in the second model, CCT mastery was entered as dependent variable. Omnibus test evaluated explanatory power of regression model based on log-likelihood ratio statistic (log$\Delta$) and chi-square test. Nagelkerke’s corrected $R^2$ indicates explained variance and Wald tests and odds ratio coefficient of predictors are reported. Comparisons of CCT mastery and direction of first and summed longest look in FB task were based on Pearson’s chi-square coefficients, while associations with time spent looking and shifting behavior to each location was addressed using Mann-Whitney-U tests, applying a Bonferroni corrected level of significance.

Associations of developmental language level, measured by receptive and expressive RDLS (and subscales structure, vocabulary, and content), and performance on action prediction question in FB task included only children who completed RDLS assessment (N = 212) and was further filtered for children who passed the TB prediction question for analyses concerning FB condition measures (N = 190). For statistical analyzed, Somer’s d coefficients were computed with prediction response as dependent variable, applying a Bonferroni corrected level of significance for whole group analyses. Comparisons of direction of first and summed longest look in FB task and RDLS performance used Somer’s d statistics and Bonferroni corrected $\alpha$ level for whole sample analyses. Spearman’s rank and partial correlation coefficients (controlled for age) were assessed for association between RDLS performance and time spent looking and percent of shifts to each location in FB and TB condition.

For a comparison of language measures and ToMI scores, all children were included who completed task assessment and for whom the ToMI was returned (N = 153). First, partial correlation analyses controlled for age were carried out including CCT sum, receptive, and expressive RDLS scores on one hand and total mean ToMI, factor 1, and factor 2 scores on the other. A multivariate ANOVA with age as covariate was carried out to evaluate effects of CCT performance (based on CCT sum score) on ToMI measures. Secondly, multiple linear regression analyses applying stepwise inclusion method were carried out to evaluate predictive effects of ToMI scores on language measures, and vice versa, taking into account controversial hypotheses on the developmental association of ToM and language. Corrected $R^2$ coefficients as an indicator of explained variance and results of ANOVAs are reported as measures of the models explanatory power. In addition, $\beta$ and t coefficients as well as p values indicate significance of each predictor.
4. Results

Results will be presented separately for each measure, starting with the demographic information of the study sample. In a next step, interrater-reliability of False Belief (FB) task scoring precedes the presentation of explicit and implicit FB task performance, respectively, followed by the analysis of a possible association of FB task measures. Results of Complement Comprehension Task (CCT) and Reynell Developmental Language Scales (RDLS) are then outlined and a correlation analysis of both language measures was carried out. Afterwards, Theory of Mind Inventory (ToMI) scores are compared to data of an American evaluation study by Hutchins et al. (in press), followed by a factor analysis in order to extract possible latent variables. Finally, ToM measures (explicit and implicit FB task performance and ToMI scores, respectively) and language tests (CCT and RDLS, respectively) were analyzed for possible associations.

4.1 Demographic data and study participation

The fourteen participating daycare institutions included children between one through six years of age (nurseries: one to three years of age; kindergartens: three to six years of age, respectively) in different group sizes (ranging from ten to twenty children, respectively) and different number of groups in individual institutions (one to six, respectively). Daycares were situated in eight different urban districts of Bremen of different funding backgrounds (private, parent organizations, ecclesiastic, and workers’ welfare federal organization). Of the total sum of 567 children in all daycares, parents of 224 children agreed for their children to participate in the study (approximately 40 percent, ranging from 15 to 100 percent, respectively) of which 70 percent filled in the ToMI and 78 percent returned the demographic questionnaire. In regard to the demographic questionnaire, the study sample was fairly homogenous (see Table 1): all children lived with their biological parents (100 percent, 90 percent of parents being married or in a stable relationship), the majority of parents hold an apprenticeship or university diploma (combined approximately 82 percent), most fathers had full-time jobs (84 percent) while mothers usually worked part-time (66 percent), the average annual household income varied around medium, 36 percent of tested children had older siblings (average number of siblings was one, ranging from zero to four), and the majority of children were raised monolingual (75 percent). On average, parents reported to spent approximately 47 hours per week with their child, not counting the time the child is sleeping, though there was a large variation within age groups (see Table 1). A Kolmogorov-Smirnov test revealed this item not to be normally
distributed (z = 1.77, p < .01). Overall, parents estimated their children to be typically developed compared to their peers (scales ranged from 1/delayed to 3/advanced with 2 meaning appropriate compared to peers).

Table 1: Demographic data of study sample, separately displayed for gender, age groups, and percent of age groups.

<table>
<thead>
<tr>
<th>demographic information</th>
<th>age group</th>
<th>2.0 m/f %</th>
<th>2.5 m/f %</th>
<th>3.0 m/f %</th>
<th>3.5 m/f %</th>
<th>4.0 m/f %</th>
<th>4.5 m/f %</th>
<th>5.0 m/f %</th>
</tr>
</thead>
<tbody>
<tr>
<td>family status (N = 174)</td>
<td>married/ relationship single other</td>
<td>10/9 (92)  0/0</td>
<td>9/12 (85)  0/0</td>
<td>20/13 (97) 0/0</td>
<td>1/7 (98)   0/0</td>
<td>7/10 (77)  0/0</td>
<td>5/10 (92)  0/0</td>
<td>9/17 (90)  0/0</td>
</tr>
<tr>
<td>level of education (mother N = 172; father N = 162)</td>
<td>no degree</td>
<td>0/1 (6)  0/0</td>
<td>0/1 (3)  0/0</td>
<td>0/0  0/0</td>
<td>0/0  0/0</td>
<td>0/1 (4)  0/0</td>
<td>0/2 (8)  0/0</td>
<td>0/0  0/0</td>
</tr>
<tr>
<td></td>
<td>&quot;Realschulabschluss&quot;</td>
<td>0/1 (6)  0/0</td>
<td>0/1 (3)  0/0</td>
<td>0/0  0/0</td>
<td>0/0  0/0</td>
<td>0/1 (4)  0/0</td>
<td>0/2 (8)  0/0</td>
<td>0/0  0/0</td>
</tr>
<tr>
<td></td>
<td>&quot;Abitur&quot;</td>
<td>0/1 (4)  0/0</td>
<td>0/1 (2)  0/0</td>
<td>0/0  0/0</td>
<td>0/0  0/0</td>
<td>0/1 (5)  0/0</td>
<td>0/2 (10) 0/0</td>
<td>0/3 (8)  0/0</td>
</tr>
<tr>
<td></td>
<td>&quot;Lehre&quot;</td>
<td>0/1 (4)  0/0</td>
<td>0/1 (2)  0/0</td>
<td>1/0 (3) 0/0</td>
<td>1/0 (3) 0/0</td>
<td>0/1 (10) 0/0</td>
<td>0/0  0/0</td>
<td>0/3 (6)  0/0</td>
</tr>
<tr>
<td>University</td>
<td>3/1 (18)</td>
<td>3/5 (20)</td>
<td>5/4 (27)</td>
<td>8/1 (29)</td>
<td>3/7 (44)</td>
<td>5/3 (38)</td>
<td>4/6 (75)</td>
<td>5/6 (41)</td>
</tr>
<tr>
<td>occupation (mother N = 166; father N = 162)</td>
<td>at home</td>
<td>2/2 (22)</td>
<td>0/3 (10)</td>
<td>3/2 (15)</td>
<td>1/1 (9)</td>
<td>1/0 (5)</td>
<td>0/4 (18)</td>
<td>2/4 (20)</td>
</tr>
<tr>
<td></td>
<td>student</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td>part-time</td>
<td>8/4 (62)</td>
<td>7/8 (61)</td>
<td>13/12 (75)</td>
<td>13/5 (71)</td>
<td>8/8 (61)</td>
<td>1/2 (18)</td>
<td>1/0</td>
</tr>
<tr>
<td></td>
<td>full-time</td>
<td>0/2 (12)</td>
<td>1/4 (19)</td>
<td>4/0 (10)</td>
<td>1/3 (20)</td>
<td>0/2 (9)</td>
<td>2/0 (25)</td>
<td>0/1 (3)</td>
</tr>
<tr>
<td>annual household income (N=146)</td>
<td>low</td>
<td>1/0 (5)</td>
<td>2/2 (22)</td>
<td>1/2 (11)</td>
<td>2/0 (6)</td>
<td>1/4 (22)</td>
<td>1/3 (25)</td>
<td>1/1 (11)</td>
</tr>
<tr>
<td></td>
<td>low to medium</td>
<td>6/4 (59)</td>
<td>5/3 (45)</td>
<td>12/7 (62)</td>
<td>8/2 (44)</td>
<td>6/4 (50)</td>
<td>1/4 (30)</td>
<td>2/10 (48)</td>
</tr>
<tr>
<td></td>
<td>medium to high</td>
<td>2/4 (36)</td>
<td>3/2 (27)</td>
<td>5/0 (13)</td>
<td>5/1 (24)</td>
<td>2/4 (28)</td>
<td>1/2 (20)</td>
<td>3/5 (41)</td>
</tr>
<tr>
<td>language (N=173)</td>
<td>monolingual</td>
<td>11/6 (84)</td>
<td>9/11 (87)</td>
<td>14/11 (75)</td>
<td>12/6 (67)</td>
<td>6/7 (59)</td>
<td>4/5 (41)</td>
<td>4/7 (69)</td>
</tr>
<tr>
<td></td>
<td>bilingual</td>
<td>1/2 (16)</td>
<td>0/4 (13)</td>
<td>6/3 (25)</td>
<td>6/3 (33)</td>
<td>0/4 (13)</td>
<td>0/4 (13)</td>
<td>0/4 (13)</td>
</tr>
<tr>
<td>older siblings (N=172)</td>
<td>yes</td>
<td>6/2 (38)</td>
<td>3/8 (45)</td>
<td>8/4 (35)</td>
<td>8/3 (39)</td>
<td>3/6 (40)</td>
<td>7/6 (60)</td>
<td>0/5 (21)</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>6/6 (62)</td>
<td>6/6 (62)</td>
<td>12/10 (65)</td>
<td>10/6 (61)</td>
<td>0/5 (21)</td>
<td>5/7 (79)</td>
<td>4/4 (31)</td>
</tr>
<tr>
<td>average time spent with child per week (N=167)</td>
<td>Mean</td>
<td>52/45</td>
<td>47/48</td>
<td>52/47</td>
<td>52/46</td>
<td>41/54</td>
<td>30/35</td>
<td>52/52</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>3.0/2.9</td>
<td>4.4/3.6</td>
<td>3.6/5.2</td>
<td>3.9/5.7</td>
<td>6.1/6.6</td>
<td>11.0/6.8</td>
<td>3.5/4.9</td>
</tr>
<tr>
<td></td>
<td>Stddev</td>
<td>10/8</td>
<td>14/14</td>
<td>16/19</td>
<td>16/17</td>
<td>18/23</td>
<td>22/24</td>
<td>10/21</td>
</tr>
<tr>
<td>developmental status (N=172)</td>
<td>Mental</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.5/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
</tr>
<tr>
<td></td>
<td>motor</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
</tr>
<tr>
<td></td>
<td>autonomy</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
</tr>
<tr>
<td></td>
<td>language</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
<td>2.0/2.0</td>
</tr>
</tbody>
</table>
4.2 Performance in the False Belief task

The False Belief (FB) task assessed children's verbal and looking behavior at different time points throughout the story course. Two children did not complete FB task assessment and were excluded from analysis, resulting in N = 222 children. To assure reliability of looking behavior assessment, an interrater-reliability analysis was conducted. Analysis of children's verbal response to the action prediction question will be presented first. Looking behavior analyses will be outlined separately. Finally, possible associations between verbal and looking behavior will be presented.

4.2.1 Interrater-reliability analysis

A second rater coded 40 percent of the video sequences for analysis one (individual lengths of time periods), unaware of task condition. For categorical variables (first look, summed longest look), raters agreed at least to 86 percent whether the child was looking to the correct, the incorrect or the distractor location (Cohen's $\kappa \geq .8$, $p < .001$). Duration of look to each location was highly correlated between raters ($r \geq .8$, $p < .001$), thus, interrater analysis speaks for almost perfect agreement and good reliability of the looking behavior related variables.

4.2.1 Action prediction question

Figure 7 shows percent of children verbally predicting the protagonist to return at the correct, incorrect, and distractor location in the FB and the TB condition separately for each age group. Interestingly, despite an age effect for mastery of prediction in the FB condition, older children seem to fail the explicit question in the TB condition (FB: $\chi^2 = 48.09$, $d = .18$, $p < .001$; TB: $\chi^2 = 57.97$, $d = -.24$, $p < .001$). A further investigation of older children’s performance in the TB condition revealed a distinct response pattern: The majority of children who answered the protagonist to return at the incorrect location gave a consistent and reasonable explanation of their choice and showed a corresponding pattern in control questions. Reasons ranged from the toy puppet to be stiff and, therefore, it was not able to look down at the scenery from its balcony, to notions that the experimenter has not been acting on the puppet and, thus, they believed the mouse to sleep or to be blinded by sun but not observing the scenery. Based on these reasonable line of arguments and taking into account correct explanations and responses in the FB condition, it can be assumed that children possibly possess a higher order ToM that allows them to take into account not only the mental representations of the story characters but also the experimenter’s point of view. For these
children responses to the action prediction question in the TB condition were recoded into a new variable, dissolving the age effect ($\chi^2 = 17.35$, $d = -.05$, n.s.; see Figure 7). Since the TB task serves as a control condition to assure general belief understanding, only children who scored correct in the recoded TB variable were included in further analysis of verbal FB performance, resulting in a sample size of $N = 201$.

Regarding FB action prediction, there was a significant increase with age of children passing the explicit test question ($\chi^2 = 51.95$, $d = .217$, $p < .001$), over 36 percent of four year and six month olds gave correct responses while only the oldest group reached a level over 50 percent correct. Interestingly, 29 percent (7 children) of the two year olds passed the FB prediction question, performing significantly better than two year and six month olds ($\chi^2 = 7.73$, $p < .05$) and performance pattern was comparable to four year olds (28 percent correct). Condition order effect analysis revealed a significant pattern for two year olds ($\chi^2 = 9.95$, $p < .01$): 50 percent of two year olds who had received the TB condition first passed the FB action prediction question while two year olds who had received the FB condition first failed the verbal response. Condition order effects were absent for all other age groups in single comparisons ($\chi^2 \leq 3.01$, n.s.). A comparison of children’s choice of location revealed a significant difference between response options: Distractor location was chosen significantly less compared to incorrect ($\chi^2 = 124.46$, $p < .001$) and correct location ($\chi^2 = 49.99$, $p < .001$). There was no significant difference between age groups for distractor vs. incorrect ($\chi^2 = 7.58$, n.s.) and distractor vs. correct contrasts, respectively ($\chi^2 = 8.07$, n.s.). Choice frequency of the remaining two locations differed across age groups: two year and six month to three year and six month olds showed a significant preference for the incorrect location ($\chi^2 \geq 9.14$, $p < .01$) while this difference was not significant for two year to four year and six month olds ($\chi^2 \leq 9.14$, n.s.) and five year olds showed the significant reverse pattern ($\chi^2 = 7.53$, $p < .01$).

Comparison of gender and FB prediction performance for all age group revealed a significant association ($\chi^2 = 6.13$, $p < .05$, uncorrected) though single comparisons within each age group did not turn out significant ($\chi^2 \leq 3.78$, n.s.). There was a trend for girls to outperform boys in age groups three year and six month and older. There was no significant effect of demographic variables on FB and TB task performance for any age group, respectively.
Figure 7: Percent of correct responses to verbal action prediction question in FB task separately displayed for gender in each age group and task condition (FB and TB). The two upper panels show the actual response to the action prediction questions in FB and TB condition, respectively. For the lower panel, responses in TB condition were recoded for children who gave a reasonable, higher-order ToM reasoning implicating explanation of their choice.
4.2.2 Looking behavior

Implicit ToM was measured in four different ways: direction of first look, direction of summed longest look, time spent looking to each location, and percent of shifts to each location. First, results of the categorical measures of direction of initial and summed longest look will be presented. Secondly, ratio scaled variables of time spent looking and shifts to each location are outlined. For the following analyses, data of children failing the TB prediction question were excluded resulting in a sample size of N = 160. It was expected that children who did not give the correct action prediction in the TB condition would also display a different looking pattern, irrespective of response explanation. For direction of summed longest look there was a significant effect of TB prediction response (z = -4.52, p < .001) and a marginal effect on direction of first look (z = -1.90, p = .05), therefore, the TB prediction filter for further analyses was applied. Analyses based on individual time periods are reported and, subsequently, results are compared to outcomes of the standardized, four seconds time period analyses.

4.2.2.1 First look and summed longest look

There was no significant effect of age or gender on first look and summed longest look in the FB and the TB condition, respectively. A comparison of choice of location resulted in significant effects for first looks and summed longest looks in both conditions, respectively: In regard to initial look in the FB condition, children looked significantly less to the distractor compared to correct location ($\chi^2 = 27.51, p < .001$) and significantly more to the correct compared to the incorrect location ($\chi^2 = 11.84, p < .007$) while the comparison of distractor vs. incorrect location revealed no preference ($\chi^2 = 3.77, n.s.$). Comparisons of direction of first look in the TB condition only revealed a significant difference for distractor vs. correct ($\chi^2 = 17.33, p < .001$) and incorrect ($\chi^2 = 27.17, p < .001$) while there was no significant preference between choice of correct vs. incorrect location ($\chi^2 = 1.31, n.s.;$ see Figure 8 upper panels). Summed longest look analysis of both conditions resulted in a significant choice of correct over distractor location (FB: $\chi^2 = 18.98, p < .001$; TB: $\chi^2 = 46.29, p < .001$) while the comparison of incorrect vs. distractor location did not reach significance (FB: $\chi^2 = 3.85, n.s.;$ TB: $\chi^2 = 5.59, n.s.$). Choice of correct vs. incorrect location was significant for TB and marginally not significant for FB condition (FB: $\chi^2 = 6.13, p = .01$; TB: $\chi^2 = 22.43, p < .001$) and; see Figure 8 lower panels).
Results

Figure 8: Percent of children looking initially and in sum longest to correct, incorrect, and distractor location, respectively, separately displayed for gender and task condition, only including children who passed the TB verbal prediction question.

Note: FB = False Belief condition; TB = True Belief condition; **: p < .001; *: p < .007; #: p = .01

There was a significant association of condition order and summed longest look in the TB condition ($\chi^2 = 6.55, p < .05$) reflected by an equal number of children looking longest to the incorrect and the distractor location if they received FB first (17 percent and 18 percent, respectively) while for the reverse condition order more children looked longer to the incorrect location (33 percent incorrect vs. 9 percent distractor). However, controlling for condition order did not reveal differences in the above mentioned effects. There were no significant effects due to demographic differences. There were no effects on direction of first or summed longest look by demographic variables, respectively.

Comparison of different time periods for analysis of summed longest look (individual versus standardized) revealed no significant differences (FB: $z = -0.17$, n.s.; TB: $z = -0.09$, n.s.).
4.2.2.2 Looking time and shifts to each location

Looking time and shift variables differed significantly from normal distribution, respectively (z ≥ 1.6, p < .01). Kruskal Wallis tests revealed no significant effect of age on percent of time spent looking to correct and distractor location in FB and TB condition (χ² ≤ 12.33, n.s.), percent of shifts to correct and distractor location in FB and TB condition (χ² ≤ 11.15, n.s.) as well for incorrect location in TB (χ² = 2.84, n.s.), but there was a significant age effect for time spent looking to incorrect location (FB: χ² = 16.61, p < .05; TB: χ² = 14.06, p < .05) and shifts to incorrect location in FB (χ² = 17.54, p < .01). For shifts to incorrect location in FB condition, Games-Howell tests revealed a marginally significant differences between two year olds and three year and six month olds (p = .05) and a significant difference between two year olds and five year olds (p < .05), two year olds shifted more to the incorrect location. For time spent looking to incorrect location, Games-Howell tests became significant for the comparison between five year olds and two to four year olds in FB condition (p < .05) and only a marginally significant difference between four year and six month olds and five year olds in TB condition (p = .06), that is, the oldest group looked less to the incorrect location, respectively.

Mann-Whitney-U tests revealed a significant gender effect for percent of shifts to correct location in FB condition (z = -2.08, p < .05, uncorrected), females shifted their look less to correct location (percent of shifts to correct: 12 percent (males) vs. 9 percent (females)). There were no further gender effects for time spent looking to each location and shifts to incorrect and distractor location, respectively (z ≥ -.69, n.s.). There were no effects due to demographic variables.

There was no significant effect for condition order for either task condition on time spent looking (z ≥ 1.28, n.s.) and percent of shifts to each location (z ≥ -1.68, n.s.).

Comparisons of time spent looking to each location using Wilcoxon signed-rank test revealed a significant preference of correct and incorrect compared to distractor location in the FB condition (correct vs. distractor: z = -4.62, p < .001; incorrect vs. distractor: z = -4.38, p < .001) and in the TB condition (correct vs. distractor: z = -7.78, p < .001; incorrect vs. distractor: z = -4.77, p < .001). Children also significantly chose correct over incorrect location in the TB condition (correct vs. incorrect: z = -4.90, p < .001). However, there was no significant preference in this regard in the FB condition (correct vs. incorrect: z = -0.51, n.s.). Separate analysis for each age group, in addition, revealed five year olds to spend significantly more time looking to the correct location compared to incorrect in FB (z = -2.76, p < .01). In TB condition, preference for correct over incorrect location did not became significant for two year and four year and six month olds (z ≥ -60, n.s.). Differences
between time spent looking to incorrect compared to distractor remained significant only for two and three year olds in FB condition ($z \geq -3.1$, $p < .01$) and for two, three, and four year and six month olds in TB condition ($z \geq -2.84$, $p < .05$). There was no gender effect on time spent looking to either location. Figure 9 displays percent of time spent looking to each location for the FB and the TB condition for each age group, respectively.

Figure 9: Mean time spent looking (error bars indicate SE) to correct, incorrect, and distractor location in each age group for False and True Belief condition, respectively.
Note: TB = True Belief condition.
In regard to shifting behavior, Wilcoxon signed-rank tests revealed a significant preference in FB condition for incorrect location compared to correct ($z = -4.16, p < .001$) and distractor location ($z = -4.69, p < .001$). In TB condition, there were significantly more shifts to correct compared to incorrect ($z = -4.55, p < .001$) and distractor location ($z = -5.75, p < .001$). Within each age group, preference of incorrect over correct location in FB condition remained significant only for two to three year and six month olds ($z \geq -2.88, p < .05$) and difference between shifts to incorrect compared to distractor location remained significant only for two to two year and six month olds ($z \geq -3.24, p < .05$). Figure 10 displays percent of shifts to each location for each age group in the FB and the TB condition, respectively.

![Figure 10: Percent of shifts (and SE indicated by error bars) to correct, incorrect, and distractor location in each age group in False and True Belief condition, respectively. Note: TB = True Belief condition.](image-url)
Comparisons of different target periods being analyzed for looking behavior revealed no significant
difference between data of the standardized method compared to the analysis of individual time
periods. There was a tendency for time spent looking to distractor location in the FB condition to be
smaller in the former measure (individual vs. standardized analysis: \( z = -1.986, p = 0.05 \), uncorrected),
however, this data did not affect the results of the above described analysis.

### 4.2.3 Association of verbal and nonverbal False Belief task performance

In a first step, associations between action prediction performance and looking behavior was
analyzed for the TB condition including all children who completed the task (\( N = 222 \)). There was a
significant association between TB action prediction performance and direction of summed longest
look (\( \chi^2 = 29.39, p < .001 \)) but not for direction of first look (\( \chi^2 = 3.74 \), n.s.). Over all age groups,
children looked longest to the respective location that they chose for their action prediction
response in TB condition (correct location: 61 percent; incorrect location: 51 percent) only summed
longest look to distractor location was equally associated with verbal prediction of incorrect and
distractor location (44 percent, respectively). Kruskal-Wallis tests revealed a significant effect of TB
action prediction on time spent looking to correct and distractor location (\( \chi^2 \geq 13.26, p < .001 \)) and a
trend for time spent looking at incorrect location (\( \chi^2 = 8.43, p = .015 \)). For percent of shifts, there also
was a significant effect of TB prediction on shifts to correct and distractor location (\( \chi^2 \geq 9.97, p < .007 \))
but not for incorrect location (\( \chi^2 = 4.58 \), n.s.). Overall, children tended to spent more time
looking and to shift more frequently to the respective location they chose as action prediction
response.

Comparison of verbal and nonverbal performance in FB condition only included children who passed
the TB action prediction question (\( N = 160 \)). There was a trend for a significant association between
FB action prediction performance and direction of summed longest look (\( \chi^2 = 13.01, p = .01 \)) but not
for direction of first look (\( \chi^2 = 2.72 \), n.s.). For correct and distractor location, children looked longest
to the respective location that they chose for their verbal prediction response (correct: 65 percent;
distractor: 100 percent). The majority of children who looked longest to the incorrect location tended
to prefer the correct over the incorrect location for action prediction (46 percent correct vs. 36
percent incorrect prediction). Kruskal-Wallis tests only revealed a trend for a significant association
between action prediction response and time spent looking to correct location (\( \chi^2 = 7.28, p = .026 \).
Figure 11 displays percent of time spent looking to each location separately for each age group and
response to action prediction question.
Results

Figure 11: Mean time spent looking to each location (and SE indicated by error bars) separately displayed for performance on the FB action prediction question for each age group.
Note: TB = True Belief condition.

In sum, children's correct action prediction performance in FB task was associated with longer looks to the correct location (summed longest look and time spent looking measures, respectively) in both task conditions.

4.3 Language measures

4.3.1 Knowledge of complements

Six children did not complete the Complement Comprehension Test (CCT) and, therefore, were excluded from analysis resulting in a sample size of N = 218 children. Analyses were carried out separately for CCT mastery (dichotomous) and sum score (ordinal scaling). There was a significant effect of age on CCT performance (age x CCT mastery: $\chi^2 = 107.34$, p < .001; age x CCT sum score: $\chi^2 = 208.80$, d = .62, p < .001). The majority of children tended to pass the CCT around the age of four years (≥ 70 percent correct; see Figure 12). Girls and boys did not differ significantly in their CCT performance ($\chi^2 = .67$, n.s.). Children who had older siblings failed CCT more frequently compared to children who had no older siblings ($\chi^2 = 4.13$, p < .05, uncorrected), however, this effect did not hold for the comparison with CCT sum score ($\chi^2 = 15.40$, n.s.).
Results

4.3.2 Developmental language level

Twelve children did not participate in Reynell Developmental Language Scales (RDLS) assessment and, therefore, were discarded from further analysis (N = 212). In addition, two children did not complete the content subscale, therefore, expressive RDLS analysis was based on a sample size of N = 210 children. Kolmogorov-Smirnov tests revealed structure and vocabulary subscales as well as receptive RDLS not to be normally distributed (z ≥ 1.68, p < .01). Subscales of expressive RDLS score, content, vocabulary, and structure, were significantly correlated (p ≥ .71, p < .001), even when controlled for age (r ≥ .40, p < .001). There was a significant effect of age of expressive RDLS performance (F (6,203) = 66.56, p < .001). Duncan post hoc tests revealed a significant improvement of expressive language score for each age group (p < .05). Gender did not exert a significant effect on expressive RDLS performance (F (1,208) = 3.53, n.s.), but there was a tendency for girls to outperform boys on the vocabulary measure (χ² = 17.29, d = .20, p = .008). Kruskal-Wallis tests revealed a significant effect of age on receptive language performance (χ² = 142.45, p < .001) but not for gender (χ² = 1.91, n.s.). Games-Howell post hoc tests revealed a significant difference on receptive RDLS score between age groups (p < .05). There was no significant effect of demographic variables on receptive and expressive RDLS performance, respectively.

Expressive and receptive RDLS scores showed a strong positive correlation (p= .84, p < .001; controlled for age: r = .70, p < .001). Wilcoxon signed-rank tests within each age group revealed a significant difference between receptive and expressive RDLS performance for two year and six month to five year olds (z ≥ -2.65, p < .01) but not for two year olds (z = -.76, n.s.). That is, except for two year olds, children tended to score higher on receptive compared to expressive RDLS scale (see Figure 12).

4.3.3 Association of CCT and RDLS scores

Partial correlation analysis controlling for age revealed a significant positive association of the CCT sum score with expressive (r = .33, p < .001) and receptive RDLS performance (r = .39, p < .001). However, single comparisons within each age group between language measures using Wilcoxon signed rank tests revealed a differentiated pattern: two to two year and six month olds as well as four to five year olds showed a significant difference between CCT sum score and expressive RDLS performance (z ≤ -2.21, p < .05) while performance on CCT and receptive RDLS differed only for two to three year and six month olds and five year olds (z ≤ -2.10, p < .05; see Figure 12).
4.4 Theory of Mind Inventory

4.4.1 Demographic variables and Theory of Mind Inventory mean scores

Of the total study sample, 70 percent returned the Theory of Mind Inventory (ToMI; N = 156). A reliability analysis including all 48 items of the ToMI resulted in good internal consistency (Cronbach’s α = .95). Item difficulty ranged from .23 to .98 and corrected item-total-correlation varied from .16 to .71, however, deletion of single items did not fundamentally increase internal consistency (Cronbach’s α if single item deleted ranged from .94 to .95). Total mean ToMI scores were checked for normal distribution (Kolmogorov-Smirnov z = .86, n.s.) and a oneway ANOVA revealed a significant effect of age (F (6,149) = 14.78, p < .001). Duncan post hoc tests revealed three significantly differing subgroups: two to two year and six month olds showed lower scores than three to three year and six month olds and four year olds and older children reached significantly higher scores (p < .05). Figure 13 displays increase of mean score in ToMI with age. There was no significant effect of gender or demographic variables on mean total ToMI performance. Correlation analysis of
average time parents spent with their child and total mean ToMI score did not reveal any significant effect \( (p = .01, \text{n.s.}) \).

On single item basis, there was no significant age effect for item 31 ("My child can pretend that one object is a different object ") and item 33 ("My child understands that, when I show fear, the situation is unsafe or dangerous"), while a multivariate ANOVA revealed a significant gender effect \( (F \geq 4.32, p < .05, \text{uncorrected}) \) for items 14 ("Appearances can be deceiving [...] my child would understand that it was not the object that changed, but rather his or her ideas about the object that changed"), 17 ("My child understands that people can lie to purposely mislead others"), and 38 ("My child understands that when a person promises something, it means the person is supposed to do it"). Comparison of mean values revealed girls to score higher on items 17 and 38, while boys reached higher scores on item 14.

**4.4.2 Comparison of ToMI scores with data from evaluation study by Hutchins et al. (in press)**

Mean total scores and standard deviations for age groups two to five years (twelve-month steps; group sizes varied between 7 and 20 children) were taken from the ToMI evaluation study by Hutchins et al. (in press) and compared to results in the present study \( (N = 156) \). The variable containing mean total ToMI scores of Hutchins et al. (in press) was not normally distributed \( (z = 2.81, p < .001) \) and there was a significant effect of age \( (\chi^2 = 144.21, p < .001) \). Comparison of ToMI performance in the evaluation and the present study showed a significant positive correlation \( (p = .47, p < .001) \) but Wilcoxon signed-rank tests for each age group revealed a significant difference for two year and six month to three year olds \( (z \leq -2.57, p < .05) \) and for four to four year and six month olds \( (z \leq -2.50, p < .05) \). Figure 13 displays mean scores and standard errors of mean on ToMI in the evaluation study by Hutchins et al. (in press) and the present study for each age group.
4.4.3 Factor analysis of Theory of Mind Inventory items

For a further investigation of ToMI performance for possible latent implicit and explicit ToM factors a factor analysis was carried out (principal-axis factoring method). Item 31 ("My child can pretend that one object is a different object") was excluded due to low adequacy for factor analysis indicated by the measure of sampling adequacy (MSA = .555). Bartlett’s test of sphericity ($\chi^2 (1081) = 3363.51$, $p < .001$) and the Kaiser-Meyer-Olkin criterion (KMO = .874) for the remaining 47 items proofed good sampling adequacy (MSA: .695 to .939). Based on Cattell's scree test, two factors were extracted. Non-orthogonal promax rotation ($k = 4$) revealed factor 1 to explain 29.4 percent and factor 2 to account for 7.4 percent of total variance with an acceptable correlation between both factors ($r = .61$). Medium and strong correlations ($r = .5$ to .8) were found for sixteen items on factor 1 and for seven items on factor 2, respectively. Seventeen items correlated equally with both factors and seven items displayed a medium to strong correlation with both factors ($r \geq .5$; for structure matrix and figure depicting qualitative item clusters see Appendices G1 and G2). A qualitative inspection of items correlating with the respective factor revealed factor 1 to subsume mostly questions regarding an understanding of belief formation, beliefs affecting behavior, means underlying actions, and specific linguistic expressions, subsumable under a general ability to apply situational knowledge.
Results

Factor 2, in contrast, focused on the understanding of linguistic expressions such as mental state verbs and the conscious reflection on beliefs and desires. Therefore, factor 1 will be referred to as "situational knowledge" and factor 2 was labeled "linguistic knowledge".

Factor 1 "situational knowledge" was normally distributed (z = .76, n.s.) while Kolmogorov-Smirnov test was marginally significant for factor 2 "linguistic knowledge" (z = 1.4, p = .04). For factor 1 "situational knowledge", an univariate ANOVA revealed a significant main effect of age (F (6, 142) = 19.82, p < .001) but not for gender (F (1, 142) = .01, n.s.). Post hoc comparisons revealed four homogenous subgroups (p < .05) according to chronological full age groups. Accordingly, Kruskal-Wallis tests revealed a significant effect of age (χ² = 27.50, p < .001) but not of gender (χ² = .64, n.s.) on factor 2 "linguistic knowledge". Games-Howell post hoc comparisons indicated a significant difference for two to two year and six month olds vs. four to five year olds (p < .05). Comparison of scores factor 1 and factor 2 within each age group revealed a significant difference for two to two year and six month olds, four and five year olds (z ≤ -1.97, p < .05), possibly due to larger variance of factor 1 "situational knowledge" compared to factor 2 "linguistic knowledge" with age (factor 1: s² = .91; factor 2: s² = .77).

4.4.4 Theory of Mind Inventory and the False Belief task

Associations between ToM measures were based on comparisons of verbal and looking behavior in FB task and ToMI scores including total mean and factor values. For comparisons with FB action prediction performance, only children were included who passed the recoded TB prediction question and for whom parents had returned the ToMI (N = 137). There was a significant effect of performance on the action prediction question on total mean ToMI score (F (2, 135) = 3.39, p < .05, uncorrected) and factor 1 "situational knowledge" (F (2, 135) = 4.10, p < .05, uncorrected), but not for factor 2 "linguistic knowledge" (F (2, 135) = 1.15, n.s.). However, age has been identified to be a significant predictor for both, FB task performance and ToMI outcome. Therefore, age was included as a covariate leading to the disappearance of the verbal FB task response effect on total mean ToMI score (F (2, 22) = 1.79, n.s.) and factor 1 "situational knowledge" (F (2, 134) = 1.12, n.s.). Within each age group, inspection of mean scores of ToMI measures separately for action prediction response in FB task revealed the following patterns: total mean ToMI scores did not vary fundamentally between FB responses and showed an increase according to chronological age groups. Mean scores of factor 1 "situational knowledge" were lower for two to three year and six month olds compared to four year to five year olds irrespective of FB action prediction performance. For factor 2 "linguistic knowledge",
factor scores increased with age irrespective of FB action prediction performance, except for two year olds: mastery of FB action prediction was associated with higher factor scores that were comparable to four year olds scoring. Figure 14 displays factor scores on either ToMI factor for FB action prediction response for each age group.

Figure 14: Mean scores (error bars indicate SE) on ToMI factor 1 "situational knowledge" and ToMI factor 2 "linguistic knowledge" for action prediction response in False Belief task, separately displayed for age groups.
Note: FB = False Belief condition; TB = True Belief condition; ToMI = Theory of Mind Inventory.

For comparisons of looking behavior in FB task and ToMI scores, only children were included who passed the TB action prediction question and for whom ToMI was returned (N = 101). There was no significant effect of direction of first look and summed longest look in FB condition on total mean ToMI score (FB first look: F (2,98) = .34, n.s.; FB summed longest look: F (2,98) = .88, n.s.), factor 1 "situational knowledge" (FB first look: F (2,98) = .17, n.s.; FB summed longest look: F (2,98) = 1.11, n.s.) and factor 2 "linguistic knowledge" (FB first look: F (2,98) = 1.62, n.s.; FB summed longest look: F (2,98) = 1.30, n.s.), respectively. Correlation analyses revealed no significant association between time spent looking to either location (correct, incorrect, distractor, respectively) and total mean ToMI score (p = -.14 to .07, n.s.), factor 1 "situational knowledge" (p = -.09 to .08, n.s.) and factor 2 "linguistic knowledge" (p = -.14 to .14, n.s.), respectively. Interestingly, correlation coefficients
indicated a trend for time spent looking to the correct location in FB condition to be negatively associated with ToMI measures, though not reaching significance (p = -.14 to -.09, n.s.). Shifting behavior in FB task, too, was not significantly associated with total mean ToMI score (p = -.13 to -.02, n.s.), factor 1 "situationa knowledge" (p = -.14 to -.02, n.s.) and factor 2 "linguistic knowledge" (p = -.10 to .02, n.s.), respectively.

Taken together, factor 2 "linguistic knowledge" was positively associated with action prediction performance in FB task for two year olds though not reaching significance in whole group analyses. Looking behavior in FB task, on the other hand, did not reveal any significant associations with ToMI measures.

4.5 Comparison of Theory of Mind and language measures

4.5.1 False Belief task performance and knowledge of complements

First, CCT mastery was found to be negatively associated with TB action prediction response (d = -.33, p < .001) but not with recoded TB prediction (d = -.06, n.s.). There was a significant positive relation of CCT mastery and FB action prediction performance (d = .28, p < .001; filtered for passing recoded TB prediction). Within each age group there were more children passing CCT compared to FB action prediction question, except for the youngest age group (29 percent FB prediction correct, but all failed CCT). However, this effect became only significant for four year and six month olds in single comparisons (d = .50, p < .01). Binary logistic regression revealed a significant predictive role of age (t = 22.48, p < .001, odds ratio coefficient = 2.38, R² = .18) but not CCT mastery (t = 3.52, p = .06) for FB prediction performance (Omnibus test: logΔ = 212.97, χ² = 25.82, p < .001). Age explained 18 percent of variance FB action prediction performance. In turn, a second binary logistic regression model revealed a significant predictive role of age (t = 56.13, p < .001, odds ratio coefficient = 7.34, R² = .54) and but not FB action prediction (t = 3.11, p = .08) on CCT mastery (Omnibus test: -2LL = 166.79, χ² = 100.75, p < .001), resulting in an explained variance of 54 percent. For analyses of looking variables, data was filtered for passing the TB prediction question (N = 160). Directions of first look and summed longest look in TB and FB condition were not affected by CCT performance, respectively (χ² ≤ 3.64, n.s.). However, mastery of CCT showed a significant effect for percent of time spent looking to the incorrect location in FB condition (z = -3.42, p < .001). Children looked less to the incorrect location if they had passed the CCT (see Figure 15).
4.5.2 General language ability and False Belief understanding

There was a negative association of TB action prediction and expressive (d = -.19, p < .001) and receptive RDLS score (d = -.22, p < .001), but not for recoded TB prediction (expressive: d = -.03, n.s.; receptive: d = -.05, n.s.). Filtered for correct recoded TB prediction, there was a significant positive effect of RDLS performance on FB action prediction (expressive RDLS: d = .16, p < .001; receptive RDLS: d = .17, p < .001). For expressive RDLS subscales, only structure (d = .23, p < .001) and vocabulary (d. = .17, p < .001) were significantly associated with FB prediction. However, single comparisons within each age group revealed only the older age groups to show a significant positive association between FB action prediction and receptive (four year and six month to five year olds: d = .26, p < .05) and expressive RDLS scores (five year olds: d = .23, p < .05; see Figure 16).
Results

Figure 16: Mean scores (error bars indicate SE) on RDLs expressive and receptive subscales in percent for action prediction response in False Belief condition, separately displayed for age groups.
Note: FB = False Belief condition; TB = True Belief condition; RDLs = Reynell Developmental Language Scales.

Comparisons of RDLs performance and direction of first and summed longest look in FB task only included children who passed the TB action prediction question and who completed RDLs assessment (N = 174). Whole sample comparisons revealed no significant association between RDLs scores and direction of first and summed longest look in FB and TB condition, respectively.

Correlation analyses revealed a significant negative association of shifts to incorrect location in FB condition and receptive (ρ = -.30, p < .001) and expressive RDLs score (ρ = -.26, p < .001), respectively. Controlling for age, however, showed no significant correlations between RDLs performance looking behavior in FB condition but a significant association of RDLs measures and time spent looking to correct (expressive RDLs: r = .19, p < .05; receptive RDLs: r = .22, p < .05) and incorrect location in TB condition (expressive RDLs: r = -.19, p < .05; receptive RDLs: r = -.19, p < .05), respectively.

Taken together, language measures were negatively associated with TB action prediction performance and showed a positive relation to FB action prediction response, though not reaching significance for all age groups. In addition, looking behavior to incorrect location in FB condition showed a negative correlation with CCT and RDLs performance.
4.5.3 Theory of Mind Inventory and language competence

Correlation analyses controlled for age revealed a significant positive association of CCT sum performance and total mean ToMI scores ($r = .22$, $p < .01$), and factor 1 "situational knowledge" ($r = .27$, $p < .01$), but not for factor 2 "linguistic knowledge" ($r = .10$, n.s.). For RDLS, receptive and expressive performances were significantly correlated with all ToMI measures ($r \geq .32$, $p < .001$). A multivariate ANOVA with age entered as covariate revealed a significant effect of CCT sum performance on total mean ToMI (F (10,141) = 2.54, $p < .01$) and factor scores on "situational knowledge" (F (10,141) = 2.97, $p < .01$). Figure 17 displays different trends across age groups regarding scores on factor 1 "situational knowledge" and factor 2 "linguistic knowledge" and performance on CCT.

![Figure 17: Mean estimated regression factor scores (error bars indicate SE) on ToMI factor 1 "situational knowledge" and ToMI factor 2 "linguistic knowledge" for performance in CCT, separately displayed for each age group.](image)

Note: CCT = Complements Comprehension Test; ToMI = Theory of Mind Inventory.

In order to evaluate predictive effects of ToMI scores on language measures performance and vice versa, multiple linear regression analyses were carried out.

For total mean ToMI, a one-step regression model ($R^2 = .41$, F (1,148) = 103.65, $p < .001$) revealed expressive RDLS performance as a significant predictor ($\beta = .64$, $t = 10.18$, $p < .001$) explaining 41 percent of variance while excluding age, CCT sum score, and receptive RDLS. Accordingly, a one step regression model ($R^2 = .18$, F (1,148) = 32.67, $p < .001$) revealed expressive RDLS performance to be a significant predictor of factor 2 "linguistic knowledge" ($\beta = .43$, $t = 5.72$, $p < .001$) while there was no
significant effect of age, CCT sum, and receptive RDLS performance. A two step regression model resulted for factor 1 "situational knowledge": in the first model, expressive RDLS accounted for 49 percent of variance (β = .70, t = 11.90, p < .001; F (1,148) = 141.56, p < .001) while in the second step, expressive RDLS (β = .57, t = 7.03, p < .001) and CCT sum performance (β = .18, t = 2.25, p < .05) together explained 50 percent of variance (F (2,147) = 75.27, p < .001). Age and receptive RDLS score were excluded due to no significant contribution to explanation of variance.

In turn, multiple linear regression models for language measures resulted in two-step models, respectively. For CCT sum, the first step included age as a significant predictor (β = .68, t = 11.47, p < .001) explaining 46 percent of variance (F (1,151) = 131.48, p < .001) while in the second model, 49 percent of variance was explained by age (β = .54, t = 7.22, p < .001) and factor 1 "situational knowledge" (β = .23, t = 3.12, p < .01) as significant predictors (F (2,150) = 74.41, p < .001). Accordingly, in a first step, age accounted for 63 percent of variance of receptive RDLS performance (β = .80, t = 16.01, p < .001; F (1,149) = 256.21, p < .001) and, in a second step, age (β = .63, t = 10.46, p < .001) and factor 1 "situational knowledge" (β = .26, t = 4.35, p < .001) explained 67 percent of variance (F (2,148) = 153.01, p < .001). Linear regression models for expressive RDLS performance included age in a first step (β = .81, t = 17.10, p < .001), accounting for 66 percent of variance (F (1,149) = 292.47, p < .001), and in a second model, age (β = .62, t = 11.09, p < .001) and factor 1 "situational knowledge" (β = .31, t = 5.63, p < .001) together explained 72 percent of variance (F (2,148) = 192.15, p < .001). For language measures, multiple linear regression models always excluded factor 2 "linguistic knowledge" because it was lacking a significant contribution to the models explanatory power. Table 2 contains results of multiple linear regression analyses for ToMI scores and language measures.
### Results

Table 2: Results of multiple linear regressions for ToMI variables and language measures (N = 153).

<table>
<thead>
<tr>
<th>dependent</th>
<th>predictor/s</th>
<th>step 1</th>
<th>step 2</th>
<th>excluded variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>total mean ToMI</td>
<td>expressive RDLs</td>
<td>β</td>
<td>.64</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>10.18</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>&lt;.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ToMI factor 1 “situational knowledge”</td>
<td>expressive RDLs</td>
<td>β</td>
<td>.70</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>11.90</td>
<td>7.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>CCT sum</td>
<td>β</td>
<td>-</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>-</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>-</td>
<td>&lt;.05</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.49</td>
<td>-</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>141.56</td>
<td>-</td>
<td>75.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ToMI factor 2 “linguistic knowledge”</td>
<td>expressive RDLs</td>
<td>β</td>
<td>.43</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>5.72</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>32.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CCT sum</td>
<td>Age</td>
<td>β</td>
<td>.68</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>11.47</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>ToMI factor 1</td>
<td>β</td>
<td>-</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>-</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>-</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.46</td>
<td>-</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>131.48</td>
<td>-</td>
<td>74.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>receptive RDLs</td>
<td>Age</td>
<td>β</td>
<td>.80</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>16.01</td>
<td>10.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>ToMI factor 1</td>
<td>β</td>
<td>-</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>-</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.63</td>
<td>-</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>256.21</td>
<td>-</td>
<td>153.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>expressive RDLs</td>
<td>Age</td>
<td>β</td>
<td>.81</td>
<td>.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>17.10</td>
<td>11.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>ToMI factor 1</td>
<td>β</td>
<td>-</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>-</td>
<td>5.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sign.</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>.66</td>
<td>-</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td>299.47</td>
<td>-</td>
<td>192.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>-</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: CCT sum = Complements Comprehension Test sum score; RDLs = Reynell Developmental Language Scales; ToMI = Theory of Mind Inventory
Results

In sum, language measures and ToMI scores were strongly associated. Regression models found expressive RDLS performance to be a good predictor for ToMI scores and an additional weak contribution of CCT sum score for factor 1 "situational knowledge". However, besides a strong explanatory power of age, factor 1 "situational knowledge" proofed to be a significant predictor for all language measure performances.
5. Discussion

5.1 Data interpretation and integration into a theoretical framework of ToM development

The present study aimed at an investigation and evaluation of the implicit-explicit distinction of ToM in development. Recent studies applying anticipatory looking measures in false belief (FB) tasks provided empirical support for the existence of two subcomponents of ToM (i.e., Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import & Connolly, 2001; Ruffman, Garnham & Rideout, 2001; Southgate, Senju & Csibra, 2007; Low, 2010). Early developing implicit reasoning capacity was claimed to show in children’s looking behavior, coded for direction of first look, summed longest look, and duration of visual fixating each location while later emerging explicit ToM was believed to become evident in declarative tasks and to be associated with language acquisition (i.e., Garnham & Perner, 2001; Milligan, Astington & Dack, 2007; de Villiers, 2007). The present study design included a three options FB task, a parental questionnaire on children’s ToM development (Theory of Mind Inventory (ToMI); Hutchins et al. in press), a test for children’s complement comprehension (Complement Comprehension Test (CCT); de Villiers & Pyers, 2002), and receptive and expressive language level (Reynell Developmental Language Scales (RDLS); Reynell, 1985). Key research questions were whether anticipatory looking patterns could be replicated in a three options FB task, whether performance on the declarative prediction question in the FB task and on language measures are associated, and whether the ToMI captures implicit and explicit aspects of ToM. In addition, different coding methods of children’s looking behavior in the FB task were hypothesized to measure the same mental process and should lead to according response tendencies.

5.1.1 The "retelling" tendency in initial orientation

Taking together, the present data did not replicate earlier findings which demonstrated looking behavior to be a valid and reliable measure of children’s implicit ToM understanding. However, coding methods and analyses of looking behavior resulted in different patterns. In accordance to earlier studies, the present results challenge the hypothesis that direction of first look serves a sensitive measure of implicit ToM. Across all age groups, children tended to look first to the location where the story protagonist hid the target object in the beginning, regardless of task condition. Thus, the proposal by Low (2010) that the initial look could reflect the fast application of implicit knowledge was not confirmed. Earlier studies have promoted two alternative interpretations of
direction of first look: a self-correcting mechanism or a tendency to "retell" the story with the eyes (i.e., Garnham & Ruffman, 2001; Low, 2010). Despite the general hypothesis that children would look to the corresponding location where they expect the protagonist to reappear, the first visual orientation may be influenced by (working) memory processes. In terms of a self-correcting mechanism this would mean that children initially focus on the location where the object really is relying on the most recent memory trace of the object’s location, probably using this as a starting point for their expectation guided looking response, or triggered by the experimenter mentioning the object’s name in the anticipation prompt, and only later look to the correct location based on their expectation in the FB condition. If children’s looking behavior is guided by a tendency to "retell" the story with their eyes they might either look to the starting point of the object (the white box) or they might only represent the protagonist's first action with the target object (ending with putting the object in one box). That is, children would look initially to the distractor, followed by the box where the object was hidden by the protagonist, and then to the object’s final location or looking behavior might only include the ending points of the performed actions with the object (leaving out the introductory sequence with distractor). While direction of first look would be incorrect in FB and correct in true belief (TB) condition if children rely on the most recent memory trace of the object’s location and only later apply a self-correcting mechanism, the process of retelling the story with the eyes would lead to children’s initial look to be correct in FB and incorrect in TB condition (if only end points of actions are considered). In the present study, children displayed the latter pattern, thus, supporting the hypothesis that they might have a tendency to "retell" the story with their eyes. Taking a closer look at the specific looking pattern, the majority of children who displayed a correct first look in FB condition subsequently shifted their gaze to the incorrect location (69 percent) and the reverse pattern was found for TB condition (68 percent). However, considering children’s next gaze shift revealed a preference to return to the correct compared to staying at the incorrect location in the FB condition (39 percent vs. 11 percent) while for TB condition the difference between looking to correct versus incorrect location was not as pronounced (28 percent vs. 20 percent). Still, an analysis of children's gaze shifting behavior over the whole time period only showed a distinct preference for the correct location in the TB condition while children tended to shift equally often to each location in the FB condition. Looking pattern in the TB condition supports the hypothesis that children frequently redirect their eyes to the location where they expect something to happen but FB performance draws a less clear picture. The more uniformly distributed looking pattern in FB condition may be due to children either lacking confidence in their build-up expectation and, thus, they try to focus equally on each location, they do not develop an anticipatory response, or a potential explanation does not translate to looking behavior resulting in random gaze
shifts, respectively. Additionally, children’s looking behavior may be guided by different processes, that is, there may be an initial tendency to "retell" the story with their eyes which shows in the first number of gaze fixations and shifts and subsequently or parallel an expectation based strategy leads children to look to a specific location.

Beside the fact that direction of first look did not capture the application of an implicit ToM reasoning strategy there is a further methodological constraint to the first look measure: Looking behavior was analyzed in a predefined time period, that is, there was a fixed onset for video analysis (first mentioning of target object in anticipation prompt). Some children tended to shift their gaze between locations before analysis onset while others did not focus on any location. As a result, measure of first look sometimes might have only included a look at the beginning of the analysis period but not the actual initial look.

Assuming that children may rely on a memory trace of the association between the protagonist and the target object required a further evaluation of a possible association bias in the FB story course. Garnham and Perner (2001) pointed out that there may be a strong association between the protagonist and the location where he originally put the object in the FB condition while this connection is not as strong in the TB condition. In the TB condition, the protagonist reappears on the balcony and observes the transfer of the target object before he returns into his house. In contrast, in the FB condition the protagonist is last seen with the target object and then enters his house through the respectively colored door. That is, in FB condition children may not apply an implicit reasoning strategy but simply rely on this association. In order to rule out this explanation, a control version of the FB condition was carried out in the present study. Fourteen children (seven male) aged three to five years and six month received the control FB condition (with counterbalanced order): After the protagonist Sam had found the target object in the white box (distractor) and stored it in one of the remaining containers Sam entered his house through the back door. That is, the experimenter moved the puppet around the model, not using one of the colored doors as in the standard FB condition. Moreover, Sam reappeared on his balcony before he moved behind the model again to have a short nap. The protagonist’s reappearance on the balcony was introduced to avoid Sam to be last seen with the target object and to parallel the story course of the TB condition. Nonparametric comparisons of children’s verbal and looking behavior in the FB task revealed no significant difference between control and standard FB condition (z ≤ -1.45, n.s.). Thus, it can be concluded that children did not rely on a simple (visual) association strategy in the FB task.
5.1.2 Coding effects on anticipatory looking analyses and the disregarded distractor

As direction of first look seems to be associated with a tendency to "retell" the story with the eyes, the possible application of an implicit ToM in looking behavior may only show after this initial look, for instance, in direction of summed longest look or time spent looking to each location. In terms of the categorical measure of direction of summed longest look, children displayed the expected pattern of looking longest to the correct location in either task condition. However, it remains questionable whether summed longest look can be interpreted to reflect implicit ToM for several reasons: First, the comparison between correct versus incorrect location was not significant in FB condition \( (p = .01) \). Secondly, summing up the time children spent looking to each location and scoring only the summed longest look might lead to a loss of information, for instance, degree of deviation is not represented by the measure. The additional analysis of time spent looking to each location allows for a more profound interpretation of behavioral patterns. Overall, children displayed a preference to look longer to the correct location compared to incorrect and distractor location in the TB condition but there was no significant difference of time spent looking to correct and incorrect location in the FB condition for two- to four-year olds. These results support the idea that summed longest look is strongly biased by the reduced information content and seems inappropriate as a valid measure of the existence and application of implicit ToM knowledge. Interestingly, only the oldest age group displayed the expected looking pattern in spending more time looking to the correct location compared to incorrect and distractor location irrespective of task condition. In this regard, the hypothesis that children's looking behavior indicates an implicit ToM understanding under the age of four years and different analysis methods do not affect results needs to be rejected.

For the present study, an additional location was included in order to increase validity and interpretability of behavioral patterns. This third location depicted a picture of the target object on the outside serving as a visual distractor and it was always the starting point of the target object, that is, where the protagonist first retrieved the object from, irrespective of task condition. This procedure allowed for an interpretation of possible underlying response strategies according to Samson et al. (2007) and Garnham and Ruffman (2001), that is, reality-based reasoning, a simplified mentalizing strategy (relying on the visual cue), or a seeing equals knowing rule. If children rely on the former strategy they would always choose the location where the object really is, the visual distractor would be chosen based on simplified processing, while children would randomly choose the distractor and the incorrect location when relying on the seeing equals knowing rule. In addition, the target object was associated with each box throughout the story course of each task condition, taking into account the possible constraint of the FB task design by Garnham and Ruffman (2001).
Interestingly, in all measures of looking behavior children displayed a clear preference for the correct or incorrect location over the distractor across all age groups. Visual orientation appeared not to be affected by the distractor and, thus, looking responses cannot be explained by a simplified mentalizing strategy or a seeing-equals knowing rule.

It is also important to note that there was no effect due to length of time period being analyzed for looking behavior. Thus, studies including a predefined time period or relying on individual response times remain comparable, yet, an explanation why the particular analysis period was chosen should be included in future experiments.

5.1.3 Control function of the True Belief condition

Filtering methods of the study sample need to be mentioned. In the present study, children displayed a distinct response pattern in the TB action prediction question. There was a trend for older children to fail this control condition which serves as a measure of children’s general understanding of and paying attention to the story, the question itself, and a general appreciation that one will act according to one’s perspective or belief that coincides with reality. It has been argued that a TB control condition allows for a better interpretation of children’s FB performance than simple comparison to chance level (Carpenter et al., 2002). In the present study, older children appeared to apply a higher order reasoning strategy and, based on their line of arguments, a recoded response variable was computed and used as a filter for further action prediction analyses. For looking behavior analyses, study sample was filtered for the actual TB action prediction response because older children, who answered this question incorrectly though presumably possessing the required knowledge, did still exert a different looking pattern in the TB condition than those who gave a correct action prediction. That is, the higher order reasoning strategy had older children take into account the anticipated aim and mental state of the experimenter, causal relation of story aspects, or physical laws, building an according expectation and looking to the respective location. With TB action prediction performance being used as a filter condition, TB and FB looking patterns were compared. As can be seen in the analysis of direction of first look, performance in the FB condition should be qualified and is in fact only interpretable by taking into account initial looking behavior in TB condition. For these reasons, the respective filters for further analyses were applied accepting the resulting reduction of sample size.

An interesting question is whether the introduction of a visual distractor and the assessment of the TB control task affected the probability distribution of FB task performance and later interpretations.
As noted above, the TB condition appears to play an important role for the interpretation of children's response patterns. In addition, a three options design in each condition leads theoretically to a specific response probability pattern (33 percent correct vs. 66 percent for the remaining two locations). However, the resulting probability distribution of the combined task adjustments in the present study remains unclear. Here, one of the incorrect response options combined a deceptive feature and the starting point for the stories. The protagonist performed an action with the target object at each location to assure object-location associations; yet, study results might imply that the distractor was less attractive than the other incorrect location. As outlined above this may have been caused by children only memorizing the end points of performed actions. Thus, the assumption of an equal probability distribution between the incorrect and the distractor location appears to be inappropriate and remains unclear. The TB control condition, on the other hand, allows for an interpretation of the FB condition performance compared to chance level. But what exactly is chance level if a three options design is applied with the before mentioned features? The exact probability for correct responses and resulting chance level performances appear to be highly dependent on children's appreciation of the two incorrect locations, still, correct responses in both task conditions (FB and TB) can be taken as an accurate indicator of ToM ability. By all means, the TB condition serves as a good control for the interpretation of behavioral data and it should always be considered as a standard adjustment and verification of basic assumptions on FB task performance (i.e., general understand of the story and related questions).

5.1.4 Processing of the False Belief action prediction question

There was no significant gender effect on looking behavior but on performance in the FB action prediction question: For three year olds and older children girls tended to outperform boys. This replicated data of earlier studies that report a slight advantage for girls on social tasks (Charman, Ruffman & Clements, 2002; McClure, 2000; Walker, 2005). However, gender has been shown to only account for one to three percent of variation in cognitive development (Ardila, Rosselli, Matute & Inozemtseva, 2011) and for ToM, the effect has been found not to be stable and to vanish or even invert throughout development (Barbu, Cabanes & Le Maner-Idrissi, 2011). Thus, interpretations in regard to sex differences should be handled with caution. As the gender effect in the present study did not show up in single comparisons within age groups, it did not receive further attention.

Performance in the FB action prediction question was assumed to be an indicator of explicit ToM knowledge. There was a significant age effect on verbal FB task performance replicating earlier
Discussion

studies. Yet, there was a slow increase of passers from younger to older age groups and no clear-cut, age-dependent shift from failing to passing the action prediction question. This unexpected pattern might be caused by specific task demands. On one hand, adding a distractor location to the FB task might have increased story complexity and possibly led to higher working memory load, executive control, and linguistic processing. Interestingly, Wimmer and Perner (1983) also applied a three-options design in their original FB task and found the majority of children to pass the test question only around the age of six years. Modified versions applied in later studies, mostly including only two response options, reported mastery of FB prediction between three to four years of age when linguistic demands were reduced (Call & Tomasello, 1999; Lewis & Osborne, 1990; Schick, de Villiers, de Villiers & Hoffmeister, 2007; van Cleave & Gauker, 2010), memory aids were provided (Mitchell & Lacohée, 1991), and need for inhibitory control was minimized (Carpenter et al., 2002; Southgate et al., 2007). Furthermore, a meta-analysis revealed age of success on the FB task to vary considerably with individual factors such as family structure, and applied ToM task (Milligan, Astington & Dack, 2007). In the present study, no effects of demographic variables were found which may be caused by the fairly homogenous sample. Working memory capacity and inhibitory control have been shown to be strongly associated with FB task performance (Carlson, Moses & Breton, 2002; Sabbagh, Moses & Shiverick, 2006; Rothmayr, Sodian, Hajak, Döhnel, Meinhardt & Sommer, 2011; Henning, Spinath & Aschersleben, 2011; van der Meer, Groenewold, Nolen, Pijnenborg & Aleman, 2011) and might have impacted the present data. Children witnessed the enacted story and though it was made sure that they understood relevant elements (i.e., names of the protagonist Sam and his friend Katie, name of the target object and its picture on the white box, references between boxes and colored curtains, memory questions on where the object has been put by Sam and by Katie, respectively), the present FB task required children to play close attention and to appreciate and to keep in mind different mental states (that is, Sam’s false representation of the object’s location, the actual location of the object known by the child, and Sam expressing the desire to return to the object). As measures of executive functioning were not included in this study, possible associations and effects could not be controlled.

On the other hand, mastery of the FB prediction question might have been affected by children being required to delay and to hold their response. That is, children observed the transfer of the target object, received the anticipation prompt, but only 10 seconds later the actual prediction question was posed. The development of an expectation about the protagonist’s reappearance may elicit an immediate tendency to share this response, for instance, by pointing at or verbally indicating the respective location. In order to communicate the expectation children might have tried to initiate a joint attention episode but the experimenter was unresponsive by looking down to the floor.
Therefore, children could have held back this executive tendency and, in turn, had more time to think about their response. Garnham and Perner (2001) found children who required more time for answering the FB action prediction question to produce more errors compared to their spontaneously responding peers. A possible explanation for this behavioral pattern is the so called temporal stacking hypothesis (Garnham & Perner, 2001): Children develop competing, temporally stacked expectations based on different reasoning strategies. For instance, early responses may result from implicit ToM knowledge while later theories are built on explicit reasoning paths. Thus, timing of response execution effects the selection of the applied reasoning strategy. A distinct effect of response latency was also reported by Atance, Bernstein and Meltzoff (2010): Until the age of three years and six months children tended to pass the FB action prediction question more frequently when they did not respond immediately, while longer response latencies led to more incorrect responses in older children. In the present study the required response delay caused by the anticipation period might have replicated these effects. While older children might have "overthought" their response and lost confidence in their reasoning process, it remains unclear which processes were at work in younger children. A striking effect was found for two year olds: The youngest group performed on the same level as four year olds in the FB task, that is, approximately 30 percent passed the FB action prediction question when controlled for TB performance. Interestingly, two year old passers had received the TB condition first, yet, reducing this effect to a condition order bias appears not a conclusive explanation. Fifty percent of two year olds who had received the TB condition first failed the FB action prediction question and together with children who started with the FB story, most two year olds did not master the FB task. The TB condition demanded children to either rely on their own knowledge of the target object’s actual location or to take into account the protagonist’s true belief in order to perform correctly. The subsequent FB story required children to appreciate the protagonist’s incorrect belief and to inhibit their own knowledge about the true state of affairs. Mastery of both action prediction questions, thus, speaks for a general understanding of the story and the test questions. Two year old passers of action prediction questions did not show different looking patterns or performances on language measures compared to other two year olds. A sampling bias cannot be ruled out but two year old passers were from different day care institutions and urban districts of Bremen and there was no gender effect. A possible explanation for the performance pattern might be that these children have applied an implicit reasoning strategy to the prediction question or they might have shown a weak from of explicit knowledge. Two year olds form the youngest group of the present study sample and they usually are also the youngest children in most day care institutions. Assuming that children were raised at home before they were accepted at a daycare, they supposedly have encountered various
social situations at home with their parents during their first two years of life. Language might not have played a substantial role in these early years, so social interaction was mostly carried out on a nonverbal level. In this regard, these experiences may have trained children in using implicit strategies to explain behavior and to gather knowledge about dependencies between actions, underlying mechanisms, and their outcome. When children entered a day care institution, however, language started to play a central role in their life. They were required to communicate their desires and needs to the teacher and other children, methods of interaction learned in early parent-child situations did not suffice any more. Accordingly, children have been shown to begin to produce sentences between the ages of two to three years (see chapter 2.5.3). In regard to ToM ability, the implicit capacity might be shadowed by the application of verbal processing strategies from this day on. In terms of the present study, this would mean that two year olds rested their response on early developed, implicit reasoning strategies while two year and six month olds and older children were confronted with immature verbal processing strategies. Thus, the majority failed the prediction question and performance increased only slowly with age and advanced coherence of implicit and explicit reasoning strategies.

5.1.5 Dual process models on ToM development

Traditional dual process models claim for implicit ToM to be associated with non-declarative, spontaneous, unconscious behavior while explicit, analytical, consciously accessible strategies are developed later in childhood and in the reasoning process, respectively. However, basic assumptions and specifications of different dual process models vary markedly and require particular consideration when interpreting study results. In this regard, Barrouillet (2011) has formulated some fundamental research questions as a guideline to distinguish predictions of dual process models:

- Where do these two systems come from and how do they develop?
- Are these two systems developmentally associated and do they follow the same developmental course?
- Is there a hierarchical organization of the respective processes, in that implicit knowledge is replaced or overridden by explicit reasoning strategies throughout development, or do both knowledge bases coexist and reach an advanced level of reasoning?

Traditional models of cognitive development strongly promote the progression from early, intuitive thinking to more sophisticated, analytical reasoning strategies. Models by Karmiloff-Smith (1992) and
Dienes and Perner (1999), respectively, allow for the coexistence of implicit and explicit knowledge but the latter strategy always forms a higher level of reasoning that is applied later in development. At first sight, neither model would originally predict younger children to master the FB action prediction question. Implicit ToM develops earlier and relies on insufficient cues and processing capacities to produce a correct prediction, according to both models, while mature ToM mostly consists of logical, analytic reasoning strategies (though implicit processing may still be available and function on an advanced level, too). In contrast, Reyna and Brainerd (2011) claim for intuition to control most advanced reasoning processes. Their fuzzy-trace theory structurally resembles dual process models of cognitive development in that it postulates the existence of two reasoning systems, intuition or gist based thinking on one hand and analytic or verbatim based processing on the other. While verbatim representations comprise the detailed and literal form of a specific situation, gist representations cover the subjective, fuzzy impressions associated with a state of affairs. Both forms of representations operate in parallel and develop during childhood, but in contrast to traditional dual process theories, the fuzzy-trace model takes intuition to be the default mode of adult reasoning processes (Reyna & Brainerd, 2011; see Figure 18). Importantly, both processes grow from an immature form to an advanced, insightful reasoning strategy, both being available to the adult mind and not one replacing the other. In contrast to other approaches, mature intuition is not a primitive reasoning mechanism but it reflects a deep understanding of a situation, based on mentally stored knowledge, values, emotion, and so on which constitute a meaning to the individual.
Figure 18: The fuzzy-trace model of reasoning according to Reyna and Brainerd (2011). The situation or problem cues the retrieval of certain memory traces that have been stored in fuzzy representations including emotional and experiential associations. Mental representations are computed based on the actual situation and impacted by activated memories. Inhibition and monitoring factors operate to hold back and control thoughts and actions that would negatively affect the reasoning process (gray box). The final decision or reasoning about the situation or problem can be based on verbatim or gist based representations with the latter presents the default setting. Variability in reasoning outcome is based on the cued and accessed knowledge and values, their previous implementation, and possible inhibition or monitoring effects.

In terms of the present study data, explanations of the distinct performance patterns in the FB action prediction question can be integrated into each of the three models: Taking two year olds performance, early implicit representations may not have passed through the redescriptions process yet (Karmiloff-Smith, 1992), the correct response may rely on a filtering process and recognition of statistical patterns in the implicit-explicit model by Dienes and Perner (1999), or a gist representation is applied, taking into account early learned situational associations (Reyna & Brainerd, 2011). The slow increase of FB task mastery, in turn, may have been affected by the income of explicit reasoning strategies or verbatim representation. As noted above, children may be confronted with and required to use language based communication when entering a daycare institution. This demand leads to higher processing loads, including the appreciation of more details and information, interdomain integration, handling conscious reflection, acceptance of attitudes and holders of mental states, and possibly even verbally accessing this process. Implicit or gist based knowledge relies on representation adjuncts (in the representational redescriptions model; Karmiloff-Smith, 1992), abstract situational regularities constituting procedural knowledge (according to the implicit-explicit model; Dienes & Perner, 1999), or fuzzy, impressionistic associations (fuzzy-trace model; Reyna & Brainerd, 2011) and can be applied fast and spontaneously. This does not imply that implicit
processing is only at work in non-declarative, spontaneous actions but it may just as well underlie
handling of more complex or delayed responses (as promoted by the fuzzy-trace model). Declarative
tasks may lead to higher processing loads on part of explicit or verbatim representations with longer
decision times leading to the integration of more and more details and, additionally, implicit and
explicit or verbatim and gist representations, respectively, may compete for control of behavior.
Newly developed explicit or verbatim strategies may initially override implicit or intuition based
representations, however, over time children become more proficient in handling both forms of
knowledge and possibly even integrate them to a coherent picture. The response delay in the
present study might just have affected this integrative process in that children slowly became able to
handle the information flood in the explicit or verbatim reasoning strategy.

5.1.6 Basic ToM accounts: theory, simulation, mentalizing but not mechanism

Turning to basic theories of ToM functioning, present study data appear to be compatible with
Theory theory, simulation, and psychoanalytic mentalizing accounts but conflict with the concept of
the development of a specialized ToM Mechanism. The latter theory implies that ToM capacities
develop with the maturation of a specialized cortical module or selection processor and though
simply generalizations are possible, mastery of the FB task requires a certain maturational level of
the relevant structures (see chapter 2.2.3). Advances in ToM reasoning rely on the accumulation and
integration of rules and causal relations in Theory theory (see chapter 2.2.1). Inborn or early acquired
knowledge grows to a complex theoretical construct of social interchange, however, the theory
allows for the above given explanation that two year olds possibly act upon implicitly learned
abstract situational associations. Interestingly, the simulation account and the psychoanalytic
mentalizing model appear to result in similar predictions. Interpersonal experience between the
infant and its parents generates psychological dynamics that are found in any social relationship later
in life with inter-subjectivity consisting of two basic components: the experience of another person’s
mind (that is, identification) and the awareness of oneself as a different person (differentiation;
Georgieff, in press; see chapter 2.2.4). Accordingly, the simulation account takes the individual to
interpret behavior from a first-person perspective either with or without a representation of the self
(see chapter 2.2.2). Both accounts fit nicely with the fuzzy-trace model in that young children may
rely on intuitive, emotionally marked representations influenced by early experience developed
dynamics, yet, according to above outlined arguments, dual process models are also compatible with
both accounts.
5.1.7 The Theory of Mind Inventory as a valid informant measure

An additional approach to ToM capacity was the assessment of the Theory of Mind Inventory (ToMI), an informant, parental questionnaire. In an evaluation study, Hutchins et al. (in press) claim the ToMI to cover a wide range of of ToM abilities and the measure was shown to have good psychometric properties. In the present study, the ToMI was translated to German language and compared to results of the evaluation study. Present empirical data replicated the general trend of age related increase of ToMI scores as in the evaluation study, yet, mean scores of two age groups differed significantly. It is important to note that the ToMI evaluation study sample was separated into full age groups, that is, covering 12 months of age while the present study applied a six month classification criterion. Thus, differences may have been caused by insensitivity to developmental increases in ToM abilities within a twelve months period and, additionally, findings of the evaluation study were based on considerably smaller samples (mean sample size of 22 children in present study vs. 13 children in evaluation study). In this regard, the replication of a general age trend plus the good internal consistency of the German version lead to the acceptance of the ToMI as a measure of ToM ability in the present study.

In terms of validity of the ToMI, Hutchins et al. (in press) have argued for accuracy of the informants, that is, the parents to play an important role. In their study, they found a significant effect of time parents reported to spend with their children per day and thus, they recommended to apply a minimum threshold criterion of five hours per day spent with the child for the inclusion of ToMI data (Hutchins et al., in press). This selection criterion should assure that parents gather enough knowledge about their child’s abilities and spontaneous reactions in various situations to form a stable and reliable report in the ToMI. In the present study, the particular question was reformulated into how many hours parents spent with their child per week based on the fact that during the week, children possibly are most of the time in the day care institution but during weekends families would spend markedly more time together. Translating the recommended threshold to the present data would result in a selection criterion of a minimum of approximately 36 hours parents and children spending together per week (around four hours per weekday plus approximately eight hours on Saturdays and Sundays, respectively). Correlation analyses revealed no effect of indicated time spent with the child per week and ToMI scores while there was a large variation of time spent variable within but not between age groups. Some reports on time spent per week were strikingly high leading to the assumption that this measure might have been insensitive and, thus, it might not be a good selection criterion for data analyses. Additional inspection of single items and total ToMI scores revealed different extreme values in single age groups. However, the handling of extreme or missing
values of informant questionnaires appears to be a tricky task. Outliers may be caused by random or systematic factors, that is, either the value accurately reflects the personality of the respondent or responses may be affected by wording of the particular question, motivational and situational factors, or specific response tendencies or strategies (Zijlstra, van der Ark & Sijtsma, 2011). As parents have to report perceptions on their children’s capacities, a rating bias may additionally be caused by the lack of specific knowledge for reliable indication, limitations of ToM competence on part of the parents, or a social desirability bias. The same reasons hold true for missing values, though it seems reasonable to assume that most missing values resulted from inattentiveness during completion of the questionnaire (in particular when taking into account that it was considerably long). However, Zijlstra et al. (2011) pointed out that it is in fact impossible to accurately identify causes of extreme values in real data and they suggest that researchers needs to come to a decision whether or not to remove extreme values. In the present study, analyses included all children for whom the ToMI had been returned for the following reasons: There was a large variation in age of children for whom the ToMI was filled in and ToM is known to be associated with age. However, ToM must not follow a linear pattern but it may rely on a nonlinear growth curve. Grimm, Ram and Hamagami (2011), for instance, promote the consideration of nonlinear developmental models for the interpretation of complex change processes. Taking into account the above outlined results of the FB task, namely, two year olds were just as good as four year olds on the action prediction question, based on the fact that the present study included typically developed children, and considering limitations of the time spent per week variable, the exclusion of extreme values was claimed rather to lead to misclassifications or additional biases and, thus, it would not result in an accurate reflection of children’s competences. Missing values for single items were excluded from total mean ToMI score calculation, however, factor extraction required the replacement by mean values to avoid a reduction of sample size.

5.1.8 Interpretation of "situational knowledge" and "linguistic knowledge" ToMI factors

An exploratory factor analysis was carried out to address the hypothesized existence of latent factors in ToMI measure. Possible effects due to applied extraction method were evaluated by contrasting results of two broadly accepted statistical techniques, that is, principal-axis and principal-component factoring. The former extraction strategy was chosen for the present analysis as both methods led to the extraction of two latent factors based on Cattell’s scree test with comparable structures and principal-axis factoring allows for residual variance which might be caused by other latent factors not included in the model. If the ToMI captures multiple ToM capacities, bearing in mind that additional
Discussion

developmental factors play a role for ToM, then the principal-component method would possibly underestimate these differential effects. In addition, an oblique rotation method was applied for theoretical and methodological reasons. On one hand, factor rotation is intended to result in an easier interpretable model structure by increasing eigenvalues and total explained variance. Considering the basic theories on ToM competence discussed in the present study, on the other hand, leads to the conclusion that any extracted factor will be related to ToM and, thus, related to any other extracted factor. In this regard, an oblique rotation appeared to integrate the need for an interpretable model while allowing for the extracted factors being correlated. It has to be noted, though, that oblique rotations increase the level of subjectivity in the interpretations of analysis outcomes. In any case, though factor analyses generally form a highly interpretable method they allow for an interesting and promising insight into internal, oblique structures of cognitive functions. The exclusion of item 31 ("My child can pretend that one object is a different object") for adequacy reasons was not surprising because this particular item evaluates acts of pretense which are learned at an early age and should be found in all study participants. The final extraction of two factors nicely fit with the hypothesized implicit-explicit distinction of ToM. Labeling of factor 1 "situational knowledge" and factor 2 "linguistic knowledge" was based on a thorough inspection of associated items looking for a possible underlying and common structure. Though both factors were intercorrelated and various items showed an equal association with both constructs, qualitative inspection allowed for a reasonable dissociation: "situational knowledge" was focused on a general understanding of situational factors and their interdependencies while "linguistic knowledge" appeared to reflect a more precise picture of single constructs (i.e., the concept of desire, including an understanding of associated linguistic expressions, expectations, and links desire guided actions). Thus, the former factor covers aspects that have been described in fuzzy or implicit reasoning strategies, that is, the application of procedural knowledge and impressionistic, abstract situational regularities. A more detailed, consciously accessible thinking process and possibly even linguistically driven representations seems to be associated with items subsumed under factor 2 "linguistic knowledge". However, this interpretation was compromised by the finding that scores on factor 2 "linguistic knowledge" of two year old passers of the FB action prediction question were comparable to those of four year old passers. In terms of above outlined interpretation of two year olds to possibly apply an implicit reasoning strategy, it was rather assumed that two year olds score higher on "situational knowledge" but not on "linguistic knowledge". Further analysis revealed "situational knowledge" to be strongly associated with expressive RDLS performance and even scores on CCT, yet, direction of influence remained unclear (either one served as a significant predictor in regression models for the respective other measure). On part of "linguistic knowledge", a significant predictive
role of expressive RDLS performance was revealed but not vice versa. It needs to be mentioned that there was a main effect of age on both factors taking into account that the study sample covered a wide age range and, thus, a long developmental period.

These associations between the derived ToMI factors and language measures, FB performance, and children's chronological age led to a reevaluation of the earlier interpretation. The allocation of each ToMI factor to implicit and explicit aspects of ToM reasoning, respectively, might not fit perfectly because results might be strongly affect by the study sample's wide age range included in analysis. Items of factor 1 included early developing competences such as appreciation of general attitudes and emotions that may lead to certain actions or patterns, followed by appearance-reality distinctions, incorrect and correct beliefs, and advanced levels of ToM, that is, concepts of secrets, promises, and intentions as well as understanding sarcasm, humor, and metaphors. Humor understanding, for instance, has been found to develop in stages with the ability to handle riddles, puns and to identify jokes around the age of seven years (Semrud-Clikeman & Glass, 2010; Sullivan, Winner & Hopfield, 1995). Factor 1 might cover ToM competencies that follow a developmental order as scores of all 16 items that were highly and exclusively correlated (r ≥ .5, p < .001) with this factor showed a trend to increase with age. On part of factor 2, items were also correlated with age, yet, half of them showed particularly high scores across all age groups (item 5 "My child understands the word ‘want’", item 7 "My child understands that when people frown, they feel differently than when they smile", item 9 "My child understands that when people get what they want, they will be happy", and item 41 "My child understands that different people have different personalities"). Thus, factor 2 appears to include items that on the one hand form a general understanding of social signals and concept structure and on the other hand relate linguistic expressions to these aspects.

The above outlined dual process models of ToM development speak for the coexistence of two distinguishable processing routes in advanced reasoning, namely, explicit and implicit strategies. Assuming that implicit ToM capacities precede explicit strategies implies that any social situation or problem triggers explicit and implicit representations in parallel after a certain age or developmental progression. In this regard, items of factor 2 may include both processes, early acquired knowledge on basic ToM understanding (reflected by high scores on items 5, 7, 9, and 41, see above) and verbatim based, age dependent explicit reasoning strategies (covered by items measuring linguistic concepts of social phenomena). It has to be noted that the lacking age effect for item 5 ("My child understands the word ‘want’") is not surprising because ‘want’ expressions have been shown to be used considerably early compared to other mental state terms (de Villiers & de Villiers, 2000; Perner, Sprung, Zauner & Haider, 2003). Though ‘want’ (german 'wollen') takes the general grammatical structure of mental state expressions in German, desire understanding has been claimed to be easier
understood than beliefs because they do not represent a real fact but only a desired outcome in the future, thus lacking the notion of truth (Perner, Sprung, Zauner & Haider, 2003). Items of factor 1, on the other hand, might share gist based, implicit reasoning processes which as particularly predicted by fuzzy-trace theory should follow a developmental course and are also applied for more complex social problems or situations. Studies on typical child development support the view that even preschoolers appreciate inconsistencies underlying humorous pictures (Loizou, 2006) and showed even earlier signs of verbal humor understanding (Johnson & Mervis, 1997), thus, processes originally thought to reflect higher order ToM capacities may be processed during preschool years in an implicit manner.

5.1.9 False Belief performance and language level

In regard to the association between FB task performance and language development, the present study results speak for a two-way interaction. Measures of children’s language level (CCT and RDLS, respectively) revealed an expected age effect. The majority of children passed the CCT around the age of four years (70 percent correct) when applying the strict criterion by de Villiers (see chapter 3.3.3). Interestingly, mastery of CCT was associated with less time spent looking to the incorrect location and number of shifts to incorrect location was negatively associated with receptive and expressive RDLS scores, respectively. As present findings speak for looking behavior to possibly reflect a tendency to "retell" the story with the eyes an association of language measures and looking behavior in FB task would not be expected. Still, it is possible that children displayed an expectation guided behavior after the retelling gaze movements but whether this expectation was derived by implicit or explicit reasoning processes remains unclear. Taking into account that younger children failed the CCT, the observed effect on FB looking behavior appears to apply only for children around the age of four years. The possible later showing expectation driven looking behavior might have included explicit reasoning which could be associated with linguistic processing routes (de Villiers, 2007; see chapter 2.4.2). That is, children might have first acquired an understanding of sentential complements of mental state expressions on a linguistic level and used this knowledge as a scaffold to build a correct explicit expectation in the FB task, thus, looking less to the incorrect location.

Regression analysis revealed no significant predictive effect of CCT score on FB action prediction performance and vice versa when controlled for age but there was a trend for a positive association. Furthermore, there were significant positive correlations between FB action prediction performance
and CCT, receptive and expressive RDLS scores but single comparisons within each age group became significant only for older age groups (around the age of four years and six months).

5.1.10 The complex affair of ToM and language revisited

In sum, ToM ability and developmental language level appear to have a distinct and reciprocal association. ToMI factor 1, labeled "situational knowledge", expressive RDLS, and CCT performance were shown to take a predictive role for each other in regression models. However, this may be a misleading, spurious association caused by the strong developmental aspect included in either measure. Language acquisition is highly dependent on age dependent maturation and "situational knowledge" appears to measure implicit or gist based ToM representations which also become more advanced with age. Expressive RDLS performance served as a significant predictor for ToMI factor 2, labeled "linguistic knowledge", but not vice versa. It seems reasonable to assume that "linguistic knowledge" includes the parallel application of implicit and explicit ToM reasoning strategies with the latter requiring language based (and probably consciously accessible) processing. In accordance, FB action prediction performance and developmental language level were only positively associated in older children who presumably apply explicit ToM knowledge (beside implicit reasoning strategies). In other words, correct FB prediction of older children may be guided by a verbatim based process while mastery of two year olds could rely on an implicit, unconscious process. The weak association between FB task performance and ToMI scores might be explained by the former method only capturing one aspect of ToM while ToMI assessed a wide range of social competences in everyday social situations and both measures relied on different modes of examination (direct test in experimental setting vs. informant measure relying on parents’ perceptions). Higher scores on ToMI for two year old passers of FB action prediction question might be interpreted to reflect an advanced level of implicit but not explicit ToM knowledge compared to their peers. Thus, explicit or verbatim based ToM might rely on linguistic processes and require a certain language level while early implicit or intuition based reasoning develops independently of language competence. Furthermore, implicit ToM might rather play a role for the acquisition of mental state expressions, taking into account results of earlier studies applying expectation-violation task designs (Onishi & Baillargeon, 2005; Träuble, Marinović & Pauen, 2010) and a recent study on the predictive effect of joint attention on mental state language (Kristen, Sodian, Thoermer & Perst, 2011). In consideration of earlier outlined theories on the association of ToM and language, both models appear to apply to a certain amount (see chapter 2.4). The so called parasitic hypothesis (Saxe & Baron-Cohen, 2006), namely that ToM development takes a linguistic vehicle to represent and express mental content
Discussion

(Astington & Jenkins, 1999; de Villiers, 2007; Papafragou, Cassidy & Gleitman, 2007), might be true for explicit ToM reasoning and even seems to be implicated in the suspiciously labeled verbatim based thinking in the fuzzy-trace model (Reyna & Brainerd, 2011). However, assuming that implicit and explicit reasoning processes work in parallel after a certain age and considering early implicit competences reported in infants (i.e., Kristen, Sodian, Thoermer & Perst, 2011; Onishi & Baillargeon, 2005; Träuble, Marinović & Pauen, 2010), the communicative hypothesis (Saxe & Baron-Cohen, 2006) seems to apply to the general association between language and ToM competences. Implicit reasoning capacities might drive the acquisition and understanding of language as a communication method including mental state expressions (note, this process most not only apply for childhood) while, in turn, language serves as a scaffold for analytical, conscious reflection - thus, supporting the view for a reciprocal facilitation promoted by Slade and Ruffman (2005).

5.1.11 An integrative dual process model of ToM development and processing

Taken together, results of the present study lead to the formulation of an integrative model of the three dual process approaches to ToM development (see Figure 19), namely, the representational redescription model (Karmiloff-Smith, 1992), the theory of implicit and explicit knowledge (Dienes & Perner, 1999, 2002), and the fuzzy-trace account (Reyna & Brainerd, 2011). During early child-parent interactions, children acquire implicit, unconscious reasoning strategies for understanding and predicting behavior which serve as a scaffold for learning the semantics of spoken language. Implicit ToM is stored as subjective impressions and associated with specific cues. The implicit default mechanisms are supplemented with explicit processing routes that develop together with language acquisition. The implementation and growth of explicit reasoning requires children at a certain age (which presumably varies depending on individual capacities, experiences, and developmental progress) to prefer explicit over implicit processing, yet, the latter strategy is only shadowed but still applied. Explicit ToM relies on linguistic input for its development and it allows for conscious and verbal access, however, it does not have to be processed linguistically after its acquisition. The knowledge base of explicit ToM contains detailed information of a situation with strong associations to particular (linguistic) markers. Any socially relevant information activates certain cues and is parallel processed in an implicit and an explicit manner, respectively, and after explicit ToM has reached a certain level of proficiency, both processing routes compete for control of behavior. Interestingly, dual-process models described above do not clearly state whether the distinction of two components of ToM applies only to information processing or actually resulting in two distinct knowledge bases. In the present thesis, terms like representation, processing routes, reasoning
strategies, and knowledge bases have been used interchangeably lending on formulations used in the original dual-process literature (Dienes & Perner, 1999; Karmiloff-Smith, 1992; Reyna & Brainerd, 2011). In the present model, ToM knowledge is believed to consist of a shared pool that includes explicitly and implicitly marked representations, respectively. Yet, any marked memory trace may be transferred to the other level, possibly by a redescription process, resulting in some shared representations or universal ToM knowledge.

Figure 19: The integrative dual process model of ToM. Socially relevant information cues memory traces of implicitly and explicitly stored ToM knowledge which leads to the parallel application of implicit and explicit processing strategies. Implicit reasoning is driven by intuition, consciously not accessible, and leads to the implementation of impressionistic memory traces in turn. Explicit processing is guided by analytical thinking, allows for conscious and verbal access, and accordingly drives the generation of detailed memory traces. Implicitly marked ToM knowledge can be transferred to explicit knowledge by a redescription process and vice versa, resulting in a general ToM knowledge base. In regard to ToM development (gray box), children first acquire implicit reasoning strategies that they apply from early on, serving as a scaffold for language development. Explicit ToM processing is driven by linguistic progress, comes in later and is taken as preference mode to allow for the implementation and growth of this processing route though implicit reasoning still takes places. Having reached a certain level of explicit knowledge, both processing routes become equal and compete for control of behavior.
5.2 Present study's limitations, shortcomings, and consequences for future directions

The present study was confronted with some limitations caused by external circumstances and potential methodological and conceptual constraints should be considered when interpreting present data.

5.2.1 Universal pitfalls of behavioral studies

Generally, studies on cognitive development rely on parent's and children's compliance, motivation, and a basic interest in research and applied tests. Recruitment of the present studies sample took place at different kindergartens and nurseries excluding public daycare facilities who did not grant permit for assessment. The final sample consisted of children who started to regularly visit daycare between two and three years of age, who were willing to participate, and for whom parents agreed to participate, only including welfare, private, ecclesiastic, or parent organizations. In addition, the present sample was fairly homogenous with most parents having a high educational level and living in a stable relationship or being married. Thus, present study results should be carefully interpreted bearing in mind these potential sampling and sample biases.

As has been noted above, there is convincing evidence for an association of executive functions such as working memory and inhibitory control and ToM assessment (Carlson, Moses & Breton, 2002; Sabbagh, Moses & Shiverick, 2006; Rothmayr, Sodian, Hajak, Döhnel, Meinhardt & Sommer, 2011; Henning, Spinath & Aschersleben, 2011; van der Meer, Groenewold, Nolen, Pijnenborg & Aleman, 2011). For economic reasons, the present study included assessment of three direct tasks during one session with a total duration of approximately 50 minutes per child. Sessions also included short phases for free play to have children recover from demanding assessments where they had to pay close attention. Chosen session duration was considered to represent the maximum time children of the target age range can sustain their attentional focus. The inclusion of more than one testing session would allow for repeated measures of ToM and language capacity and thus, yielding more reliable results, on one hand and the additional assessment of other cognitive competences to account for possible influences on ToM performance on the other. In addition, a longitudinal study design could capture intra- and inter-individual differences and developmental trajectories in ToM reasoning across different ages.
5.2.2 Consideration of methodological constraints

Methodological caveats need to be considered when evaluating results of the ToMI. Beside the fact that informant measures assess cognitive abilities in an indirect manner and always are confronted with the allegation to be influenced by other factors (i.e., response strategies, wording of questions), analyses in the present study included extreme values, a strong age effect, and were based on a newly developed German version. That is, translation of the original English ToMI might have led to different wording of questions (i.e., due to syntactical differences between languages, or finding equivalents for puns) and thus, measured slightly different aspects of ToM. However, statistical comparisons of present results and scores of the evaluations study by Hutchins et al. (in press) revealed only minor differences for single age groups which could reflect differences between study samples. Based on good internal validity and a replication of the general trend in ToMI score increase with age, the German version seems to be a valid measure of ToM capacity and to accord to the English questionnaire. Handling of extreme values and so called statistical outliers in questionnaire based data forms a fundamental issue in cognitive sciences (Zijlstra, van der Ark & Sijtsma, 2011) and requires further discussion and research. As mentioned earlier, extreme values were not discarded from analyses as they might reflect inter- and intra-individual differences in ToM development. Results of the conducted factor analysis also might have been affected by inclusion of outliers, applied methods of extraction and rotation, and most importantly the wide age range. In order to evaluate whether the extracted factors could possibly reflect the hypothesized distinction between implicit and explicit components of ToM it would be desirable to apply the ToMI to larger samples of children of the same age. For the present study, different extraction methods were compared and resulted in quite similar factor patterns, thus, supporting the defined factors. However, interpretation of factor analysis results should always be handled with care, bearing in mind the problem of objectiveness.

5.2.3 Specific limitations of the present study and perspectives for future research

A central concern is whether looking behavior can really capture implicit processing or ToM knowledge. Though there is convincing evidence for children's looking behavior to represent exploratory preferences (Spelke, 1985), surprise in expectation-violation tasks (i.e., Onishi & Baillargeon, 2005; Träuble, Marinović & Pauen, 2010), or anticipation in FB tasks (i.e. Clements & Perner, 1994; Ruffman, Garnham, Import & Connolly, 2001; Low, 2010), present data did not replicate earlier findings. Originally, implicit ToM was claimed to show in spontaneous, non-
declarative actions. However, the FB task puts children in an experimental setting, having them observe and not actively participate in an acted out story, measuring so called spontaneous looking behavior in fixed, predefined time periods. Some of these constraints have been approached in other studies: For instance, active helping designs require children take an active role in the story course by having them first observe an action and then they can decide whether they want to help the counterpart in reaching a goal or not. Recent studies found children as young as 18 months to succeed in those tasks (Buttelmann, Carpenter & Tomasello, 2009; Warneken & Tomasello, 2006), thus, providing further support for an early ability to infer mental states (i.e., goals, beliefs).

Interestingly, ToM anticipation studies presented earlier were all carried out in English-speaking countries. Taking into account present study results from a German-speaking sample, there appears to be a chance for possible culture or language effects. Thus, it would be desirable to parallel present study design with an assessment in an English-speaking sample and additionally to include measures of differences between cultures (i.e., educational system, exposure to mental state language and content, linguistic specifications).

However, application of implicit reasoning strategies must not only be driven by spontaneous, non-declarative situations but any socially relevant information might be processed in parallel on implicit and explicit ToM routes (see Figure 19). Early acquired implicit ToM might serve as a default for information processing and shows in declarative and non-declarative actions, respectively. Present study results might have been affected by the three options FB task posing higher demands, causing children to apply a coping strategy, that is, retelling the story with the eyes. This coping strategy, in turn, could have shadowed any implicit ToM driven anticipation behavior. The introduction of an additional incorrect response option appears to have increased FB task complexity, yet, it did not receive as much attention during assessment as the remaining two locations. Children's general lack of interest in this distractor location either speaks for children not to rely on a simplified mentalizing strategy (purpose of the deceptive feature) or results from having the object always start at this location. The fixed starting point was chosen to allow for an easy interpretation in terms of response strategies according to Samson et al. (2007), however, it might have led to another constraint: It was hypothesized that the third location did have a less relevant role in the story course in the study by Garnham and Ruffman (2001) because the object was never inside the respective container. In the present study, the protagonist always found the object first in the distractor box and, thus, performed a similar action at either location. Still, children might tend to only represent end points of performed actions which would result in leaving out the distractor in the mental reasoning process. As the deceptive feature did not seem to affect children’s responses, future studies should
include a third location in FB task assessment with randomized order of hiding actions of the target object.

In terms of the possible parallel application of implicit and explicit reasoning strategies it would be interesting to assess children’s looking behavior during the FB prediction prompt and to relate it to earlier gaze shifting behavior. Beside the possibility that children could display diverging responses on the visual and the verbal or pointing level during action prediction, this strategy would also allow for a superior statistical analysis: As has been outlined above, time spend looking to each location proved to be a more valid and reliable measure than categorical scoring methods. Verbal responses originally allow only for categorical coding. The problem of comparing low scaled verbal responses and high scaled looking behavior in FB tasks has been noted by Low (2010). Measuring children’s looking behavior after each prompt in the FB task (anticipation and action prediction, respectively) could possibly resolve this methodological constraint.

In regard to the conclusions drawn on two year olds performance in the FB task discussed above it would be interesting to study children who were raised at home. If children acquire implicit reasoning strategies in early child-parent interactions and may apply this knowledge even in declarative contexts as long as explicit strategies have not been developed then the observed trend for mastery of the FB action prediction in the present study would be expected to be even more pronounced in home raised children.

The reasonable application of a filter for better interpretation of FB task performance, that is, only including children who passed the TB action prediction question, led to a considerable reduction of the total study sample and sizes of single age groups. Therefore, statistical tests for comparisons of behavioral patterns within each age group had less power and might not always have achieved significance. In addition, between group analyses might have been affected by different group sizes and response probabilities in behavioral measures, for instance, the majority of children under the age of four years failed the CCT while the reverse pattern applied for older age groups. Whole sample analyses were controlled for multiple comparisons by introducing a Bonferroni corrected level of significance. However, interpretation of results on single age group level should always consider these methodological limitations.

Finally, the present behavioral study does not yield implications for neural correlates and functional organization of measured cognitive processes. There is a controversial debate in cognitive sciences on the neural bases of ToM processes (Liu, Sabbagh, Gehring & Wellman, 2009; Mitchell, 2009; Saxe, Whitfield-Gabrieli, Scholz & Pelphrey, 2009) that requires further research. Interestingly, Dienes and
Perner (1999) drew an analogy between implicit and explicit ToM and dorsal and ventral visual processing routes, respectively. Whether or not this theoretically guided approach to the neural underpinnings of ToM and the courageous attempt to bridge the gap between different cognitive functions is a step in the right direction remains an open question for future research. However, even the seemingly established functional dissociation of visual perception just recently became subject to revision and controversial discussions (Milner & Goodale, 2008; de Haan & Cowey, 2011). Studies using neuroscientific methods (i.e., functional Magnet Resonance Imaging, Electroencephalography) should be applied to study neural networks underlying ToM processes in adults and children, bearing in mind the methodological limitations of each single or combined technique.

5.3 Final conclusions

In sum, present study results speak for the existence of two distinct ToM processing routes in typically developing children. Looking behavior in FB task did not serve as an indicator of implicit reasoning capacity, thus, not replicating earlier studies on ToM development. Children rather displayed a tendency to "retell" the story with the eyes. This finding might have resulted from higher task demands by introducing a three options design to the FB measure. However, a distinct performance pattern on the forced choice action prediction question in the FB task and scores on latent factors derived from a factor analysis of ToM1 measure served as empirical support for an integrative dual process model of ToM acquisition. Language was shown to have a reciprocal developmental association with ToM components.

The combination of different assessment techniques, methods, and analyses procedures, as well as development, evaluation, and consideration of different models and theories in interdisciplinary approaches are required for a deeper understanding of ToM development. The present study findings and resulting implications serve as a further step in this direction.
6. References


On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach


On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach


On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach


Science, 311, pp. 1301-1303.

Natural theories of mind: Evolution, development and simulation of everyday mindreading (pp. 19-


A. Whiten (ed.). Natural theories of mind: Evolution, development and simulation of everyday 

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of 

World Medical Association (2004). Declaration of Helsinki. Ethical Principles for Medical Research 
Involving Human Subjects. The World Medical Association (WMA).

186-212.
7. Appendix

Please note, as the study has been carried out in Germany, parent information and consent forms as well as the approval letter of the ethics committee of the University of Bremen are provided in German language.

The demographic questionnaire, the verbatim of the False and True Belief stories, the Complement Comprehension Test (CCT) are provided in English language, German versions are available on inquiry. The Theory of Mind Inventory (ToMI) is attached in English and in German language.
Appendix A

Approval letter of the ethics committee of the University of Bremen

Ihr Vorhaben: "On the relation between implicit and explicit Theory of Mind and linguistic competence"

Sehr geehrte Frau Gießmann,

die Ethikkommission der Universität hat auf ihrer Sitzung am 10.03.2009 festgestellt, dass aus ihrer Sicht keine Bedenken gegen die Durchführung des o.g. Forschungsvorhabens bestehen.

Mit freundlichen Grüßen

[Signature]

I.A. - Brüning
Appendix B1

**Parental information form of Theory of Mind study**

Lieber Eltern,

die Abteilung für Neuropsychologie und Verhaltensneurobiologie am Zentrum für Kognitionswissenschaften (ZKW) der Universität Bremen führt derzeit ein Projekt zur Untersuchung des Zusammenhangs von Kommunikationsverhalten, sozialen Fähigkeiten und Sprachkompetenz im Kindesalter durch.

**Worum geht es?**

Die sogenannte *Theory of Mind* beschreibt die Fähigkeit sich in andere hineinversetzen zu können und bildet eine wichtige Grundlage unseres sozialen Verhaltens. Ohne diese grundlegende Kompetenz wäre unser soziales Miteinander eher zufällig und wenig interaktiv. Bisher wurde vermutet, dass Kinder erst ab dem 4. Lebensjahr beginnen, diese Fähigkeit zu zeigen. Die *Theory of Mind* ist aber weit komplexer als im ersten Moment angenommen wurde. Sie setzt sich aus vielen verschiedenen Komponenten zusammen, zum Beispiel aus der Fähigkeit andere Perspektiven einzunehmen, sich empathisch einzufühlen, zu verstehen was Wünsche und Gefühle sind, und vieles mehr. Bis heute ist unklar, wann genau sich diese einzelnen Fähigkeiten entwickeln und wodurch sich eine *Theory of Mind* eigentlich bildet. Mit der hier vorgestellten Studie werden zwei Ziele verfolgt:

Zum einen soll die *Theory of Mind* mit ihren vielen Komponenten untersucht werden, um herauszufinden, ob es hier bestimmte Phasen oder Stufen bei der Entwicklung gibt und in welchem Alter diese auftreten.


**Was wird gemacht?**


Zusätzlich soll ein Elternfragebogen Aufschluss darüber geben, welche sozialen Kompetenzen das Kind im Alltag zeigt und wie stark diese ausgeprägt sind.
Appendix B1 continued

**Warum wenden wir uns an Sie?**

Wir möchten Sie und Ihr Kind gerne einladen, an dieser Studie teilzunehmen. Um das oben beschriebene Projekt umsetzen zu können sind wir auf Ihre Unterstützung und die Ihres Kindes angewiesen und wir würden uns sehr freuen, wenn wir Sie für unser Projekt gewinnen können. Das Projekt soll helfen, die Bedeutung der Förderung von sozialen Fähigkeiten im Kindesalter deutlich zu machen und Licht in die komplexe Theory of Mind Forschung zu werfen. Wir würden Ihnen gerne beim nächsten Elternabend das Projekt und die einzelnen Tests genauer vorstellen und stehen Ihnen jederzeit für Fragen zur Verfügung.


Falls Sie Fragen zu dem Projekt haben sollten, können Sie sich gerne an die Projektleitung wenden:

Dipl.-Psych. Charlotte Herzmann  
Tel: 0421 – 218 – 68744  
Email: C.Herzmann@uni-bremen.de

Wenn Sie Ihr Einverständnis zur Teilnahme an der Studie erteilen möchten, dann benötigen wir eine von Ihnen unterzeichnete Einwilligungserklärung, die Sie bei dem Elternabend bekommen.

Wir bedanken uns herzlich für Ihr Interesse und hoffen auf Ihre Unterstützung!
Appendix B2

Parental consent form of Theory of Mind study

Einwilligungserklärung

für die Teilnahme am Projekt „Zusammenhang zwischen impliziter und expliziter Theory of Mind und Sprachkompetenzen“.

Hiermit erkläre ich,
Name, Vorname: ___________________________________________________________
Telefon: ______________________________________________
Adresse: __________________________________________________________________


Name des Kindes: ____________________________________________________
Geburtsdatum: ________________________________


Eine schriftliche Elterninformation habe ich erhalten.

_________________________   _____________________________   _____________________________
Ort, Datum                 Unterschrift: Sorgeberechtigte/r         Unterschrift: Untersucher
Appendix C

Verbatim of False Belief task

italic text: directions for experimenter
normal text: story to be told and acted out
bold italic text: target questions

material:
- two mouse puppets (Sam and Kathy)
- cardboard model house with three doors and colored curtains
- three boxes (one in front of each door; green, yellow, white with picture)
- target object (candy or cheese)
- camera behind the house model

Version 1: False Belief story

This is Sam. Do you know what animal Sam is? He is a mouse. Sam says: “Hello!”
Show Sam.
And this is Sam’s house. His house has three doors and he can use either one to enter his house, see? He can go in here to get inside and he can come out here again and now he’s entering the house from here.
While explaining, move Sam over the scene and let him enter and reappear through the doors (e.g., entering on one side, coming out in the middle, and then using the last one to get inside again). Make clear that Sam can use each door to get inside and to get out as well.
In front of each door there is a box.

Comprehension 1 questions: If Sam wants to go to the green box where will he come out?
Always ask the question and check the child’s respond. The child has to get 2 out of 3 correct, otherwise repeat it until the child’s seems to understand.

He comes out of the green door. Sam says: “Oh, I wonder what is inside. Well, there’s nothing inside.”
Sam comes out of the green door and takes a look into the green box which is empty.
If Sam wants to look into the yellow box which door will he use? He uses the yellow door.
Sam says: “Oh, look, what’s in here? It’s empty as well.”
Sam comes out of the yellow door and looks into the yellow box which is empty.
And if Sam wants to check the white box where does he have to come out? He comes out of the white door. Sam says: “I want to know what’s in here. Oh, look, there’s a picture on the box. Looks like cheese. Let’s have a look. Wow, there’s really cheese in here, see?”
Sam comes out of the white door and looks at the white box before opening it and finds a piece of cheese.
Sam takes the cheese and walks in front of his house. Sam says: “Oh, I really like cheese. And you?” Sam is now eating from his cheese. Soon he has enough so he walks to one of the colored boxes. Sam says: “I will put my cheese here and get something later on.”
Sam puts the cheese inside the green/yellow box.
Sam says: “I am tired, I will go inside and take a nap.”
Sam enters the house through the door behind the box with the cheese in it and makes snoring sounds.
Appendix C continued

**Memory 1 question:** Where did Sam put the cheese?
*If the child answers correctly, go on with the story. If the child fails, Sam reappears though the door related to the target box and looks inside again to check. Then he goes inside and memory 1 is asked again.*

Now, Sam is asleep. Oh look, who is this? This is Kathy mouse. Kathy says: “Hello! What’s your name?” Kathy is taking a walk. Kathy says: “Oh, there is a box. I wonder what is inside. Let’s have a look.”

*Kathy appears on the scene and shown to the child. She walks up and down in front of the house then she moves to the target box. She opens it and looks inside.*

Kathy says: “Hey, there’s a cheese in here. Great, I like that!” Kathy takes the cheese. Kathy says: “You know what? I don’t want to eat it right now. I put it here in the yellow/green box.”

Look, Kathy puts it in here.

*Move Kathy to the other colored box and let here put the cheese inside.*

Kathy says: “I have to leave. Goodbye!”

*Put Kathy away, out of the child’s sight. Make a yawning sound.*

Oh, did you hear that? I think Sam is waking up. Sam is hungry for cheese. *

**Anticipation prompt:** I wonder where he is going to look for his cheese.

*Leave Sam behind the scene. While saying that Sam is waking up look down to the floor and leave your eyes there until the prediction question is asked. Count approximately 5 seconds before speaking again.*

Now Sam wants to get his cheese.

**Prediction question:** Where will he look for his cheese?

**Explanation:** Why do you think that he is going to look there?

**Memory 2 question:** Where did Sam put the cheese first?

**Memory 3 question:** And where is the cheese now?

*After asking the questions, Sam reappears through the correct door and looks inside the box.*

Sam says: “Now I want my cheese and I remember that I left it in here. Oh no, it’s gone. Where is my cheese? Can you help me? – Oh there it is? Let’s check that. Oh yes, great! Thank you!”

**Comprehension 2 question:** Did Sam see that Kathy took the cheese and put it in here?

*Sam takes the cheese and goes away.*

---

**Version 2: true belief task**

Look, this is Sam again in front of his house. He can use each door to enter his house, do you remember? He can go in here to get inside and he can come out here again and now he’s entering the house from here.

*While explaining, move Sam over the scene and let him enter and reappear through the doors (e.g., entering on one side, coming out in the middle, and then using the last one to get inside again). Make clear that Sam can use each door to get inside and to get out as well.*

In front of each door there is a box.
Comprehension 1 questions: If Sam wants to go to the green box where will he come out?
Always ask the question and check the child’s respond. The child has to get 2 out of 3 correct, otherwise repeat it until the child’s seems to understand.
He comes out of the green door. Sam says: “Oh, I wonder what is inside. Well, there’s nothing inside.”
Sam comes out of the green door and takes a look into the green box which is empty.
If Sam wants to look into the yellow box which door will he use? He uses the yellow door.
Sam: “Oh, look, what’s in here? It’s empty as well.”
Sam comes out of the yellow door and looks into the yellow box which is empty.
And if Sam wants to check the white box where does he have to come out? He comes out of the white door. Sam says: “I want to know what’s in here. Oh, look, there’s a picture on the box. Looks like a candy. Let’s have a look. Wow, there’s really a candy in here, see?”
Sam comes out of the white door and looks at the picture on the white box before opening it and finding a candy.
Sam takes the candy and walks in front of his house. Sam says: “Oh, I really like candy. And you?” Sam wants to keep the candy for later, so he walks to one of the colored boxes. Sam says: “I will put my candy here and get something later on.”
Sam puts the candy inside the green/yellow box.
Sam says: “The sun is shining today so I will go to my balcony.”
Sam enters the house through the door behind the box with the candy in it. He reappears on the balcony, on top of the house.
Look, Sam is on his balcony now. Sam says: “It’s really nice and I can see all the boxes from here!”

Memory 1 question: Where did Sam put the candy?
If the child answers correctly, go on with the story. If it fails, Sam goes down again and reappears though the door related to the target box and looks inside again to check. Then he goes inside, reappears on the balcony and memory 1 is asked again.
Oh look, there is Kathy mouse again. Kathy says: “Hello! Oh, hi Sam!” Sam says: “Hi Kathy!” Kathy is taking a walk. Kathy says: “Oh, there is a box. I wonder what is inside. Let’s have a look.” (talking to herself)
Kathy appears on the scene and talks to Sam. Then, she walks up and down in front of the house then she moves to the target box. She opens it and looks inside.
Kathy says: “Hey, there’s a candy in here. Great, I like that!” Kathy takes the candy. Kathy says: “You know what? I don’t want to eat it right now. I put it here in the yellow/green box.”
Look, Kathy puts it in here.
Move Kathy to the other colored box and let here put the candy inside.
Kathy says: “I have to leave. Goodbye Sam!” Sam says: “Goodbye Kathy!”
Kathy leaves.
Sam says: “I have to go to the bathroom now.” Look, Sam is going inside to use the bathroom.
Sam disappears behind the scene. Make sounds of water.
Now Sam is done. And Sam is hungry for candy.
Anticipation prompt: I wonder where he is going to look for his candy.
Leave Sam behind the scene. While saying that Sam is ready look down to the floor and leave your eyes there until the prediction question is asked. Count approximately 5 seconds before speaking again.
Appendix C continued

Now Sam wants to get his candy.

Prediction question: Where will he look for his candy?
Explanation: Why do you think that he is going to look there?
Memory 2 question: Where did Sam put the candy first?
Memory 3 question: And where is the candy now?
   After asking the questions, Sam reappears through the correct door and looks inside the box.
Sam says: "Now I want my candy and I saw that Kathy took it and put it in the yellow/green box.
There it is, great!"

Comprehension 2 question: Did Sam see that Kathy took the candy and put it in here?
Sam says: "Now I will eat it. Thank you and see you later!"
   Sam takes the candy and goes away.
Appendix D1

Theory of Mind Inventory - English version

ToMI: Theory of Mind Inventory (Hutchins et al, in press)

INSTRUCTIONS: The purpose of this measure is to learn about caregiver’s ideas regarding children’s thoughts and feelings. Please read each statement carefully and indicate the degree to which you believe each statement is true for your child. Sometimes, you may feel that you don’t know for sure whether a statement is true or not. When you feel this way, reflect upon your experiences with your child and try to decide, given everything you know about this child, how certain you are that the statement is true or not true. There are no right or wrong answers so try to respond as honestly and thoughtfully as possible. Your answers are completely confidential. To indicate your response, make a vertical hash mark along the appropriate point on the continuum. If you feel undecided as to the most appropriate answer, make a hash mark somewhere underneath ‘undecided’ but keep in mind that you can lean toward either end of the continuum. For example:

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

If you have more definite feelings that the statement is true or not true, make a hash mark along the continuum that reflects those feelings. For example:

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

To express total confidence in your answer, put a hash mark through the dot at either end of the continuum. For example:

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>
Appendix D1 continued

1. My child understands that when someone puts on a jacket, it is because he/she is cold.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. If it were raining and I said in a sarcastic voice “Gee, looks like a really nice day outside,” my child would understand that I didn’t actually think it was a nice day.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. My child recognizes when someone needs help.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. My child understands that when someone says they are afraid of the dark, they will not want to go into a dark room.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. My child understands the word ‘want’.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix D1 continued

6. My child understands that people can be wrong about what other people want.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

7. My child understands that when people frown, they feel differently than when they smile.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

8. My child understands the word ‘think’.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

9. My child understands that when people get what they want, they will be happy.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

10. If I put my keys on the table, left the room, and my child moved the keys from the table to a drawer, my child would understand that when I returned, I would first look for my keys where I left them.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D1 continued

11. My child understands that peoples’ personalities basically don’t change from day to day.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12. My child understands that to know what is in an unmarked box, you have to see or hear about what is in that box.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13. My child understands the word ‘know’.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

14. Appearances can be deceiving. For example, when seeing a candle shaped like an apple, some people first assume that the object is an apple. Given the chance to examine it more closely, people typically change their mind and decide that the object is actually a candle. If my child was in this situation, my child would understand that it was not the object that changed, but rather his or her ideas about the object that changed.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Appendix D1 continued**

16. If I said “Let’s hit the road!” my child would understand that I really meant “Let’s go!”

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17. My child understands that people can lie to purposely mislead others.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18. My child understands that when someone makes a ‘guess’ it means they are less certain than when they ‘know’ something.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

19. My child understands that when someone is thinking about a cookie, they cannot actually smell, eat or share that cookie.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20. My child understands that people can smile even when they are not happy.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix D1 continued

21. My child understands the difference between when a friend is teasing and when a bully is making fun.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

22. My child understands that people don’t always say what they are thinking because they don’t want to hurt others’ feelings.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

23. My child understands the difference between lies and jokes.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

24. My child understands that if two people look at the same object from a different standing point, they will see the object in different ways.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

25. My child understands the word ‘need’.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>
On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach

Appendix D1 continued

26. My child understands that people often have thoughts about other peoples’ *thoughts*.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

27. My child understands that people often have thoughts about other peoples’ *feelings*.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

28. My child understands whether someone hurts another on purpose or by accident.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

29. My child understands the word ‘desire.’

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

30. My child recognizes when others are happy.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix D1 continued

31. My child can pretend that one object is a different object (for example, pretending a banana is a telephone).

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

32. My child recognizes when a listener is not interested.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

33. My child understands that, when I show fear, the situation is unsafe or dangerous.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

34. My child understands the word ‘if’ when it is used hypothetically as in, “If I had the money, I’d buy a new house.”

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>

35. My child understands that when a person uses his/her hands as a bird, that the person doesn’t actually think it is a real bird.

<table>
<thead>
<tr>
<th>Definitely Not</th>
<th>Probably Not</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
</table>
**Appendix D1 continued**

36. My child knows how to make up stories to get what he/she wants.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

37. My child understands that in a game of hide and seek, you don’t want the person who is ‘it’ to see you.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

38. My child understands that when a person promises something, it means the person is supposed to do it.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

39. My child is able to put himself/herself in other people’s shoes and understand how they feel.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

40. My child understands that when someone shares a secret, you are not supposed to tell anyone.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D1 continued

41. My child understands that different people have different personalities.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

42. If I said “What is black, white and ‘read’ all over? It’s a newspaper!” my child would understand the humor in this play on words.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

43. My child is able to show me things.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

44. My child is able to pay attention when I show him/her something.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

45. My child understands the word ‘believe’.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D1 continued

46. When we like others, we are likely to interpret their behavior in positive ways and when we don’t like others, we are likely to interpret their behavior more negatively. My child understands that previous ideas and/or opinions of others can influence how we interpret their behaviors.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

47. My child understands that two people can see the same image and interpret it differently. For example, when looking at this image, one person might see a rabbit whereas another might see a duck.

48. My child understands that if Bruce is a mean boy and John is a nice boy, Bruce is more likely than John to engage in malicious or hurtful behaviors.

<table>
<thead>
<tr>
<th>Definitely</th>
<th>Probably</th>
<th>Undecided</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D2

Theory of Mind Inventory - German version

Fragebogen zur Einschätzung der kindlichen Theory of Mind
(Übersetzung des englischsprachigen Theory of Mind Inventory von Hutchins et al., in press)

Instruktion: Ziel dieses Fragebogens ist es, mehr über die Gefühle und Gedanken von Kindern durch die Einschätzung ihrer Eltern zu erfahren.
Bitte lesen Sie jede der folgenden Aussagen gründlich durch und geben Sie an, wie stark diese Behauptung auf Ihr Kind zutrifft. Manchmal kann es vorkommen, dass Sie sich nicht sicher sind, ob eine Aussage zustimmen sollen oder nicht. In einem solchen Fall versuchen Sie sich bitte an Situationen zu erinnern, in denen Ihr Kind ein solches Verhalten gezeigt haben könnte und entscheiden Sie dann, wie sicher Sie sind, dass die Aussage zutrifft oder nicht. Es gibt keine falschen oder richtigen Antworten. Stützen Sie Ihr Urteil auf Ihr gesamtes Wissen über Ihr Kind und Ihre gemeinsamen Erlebnisse und antworten Sie bitte ehrlich und nach reiflicher Überlegung. Um Ihre Einschätzung anzugeben platzer Sie bitte einen vertikalen Strich an der entsprechenden Stelle auf der Antwortska, die sich unter jeder Aussage befindet.
Wenn Sie unsicher sind, ob eine Aussage zutrifft oder nicht, setzen Sie ihre Markierung bitte unter „weiß ich nicht“. Sie können hier auch eine kleine Zustimmungs- oder Ablehnungstendenz angeben, indem Sie ihren Markierungs-Strich eher in die eine oder die andere Richtung der Skala verschieben.
Zum Beispiel:

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

Wenn Sie meinen, dass eine der Aussagen tendenziell mehr oder weniger zutrifft, markieren Sie bitte die Skala an der Stelle, die am Besten zutrifft. Zum Beispiel:

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

Wenn Sie sich absolut sicher sind, dass eine Aussage zutrifft oder nicht zutrifft, machen Sie bitte Ihren Markierungs-Strich auf dem Punkt rechts bzw. links von der Skala. Zum Beispiel:

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>
Appendix D2 continued

1. Wenn sich jemand eine Jacke anzieht, versteht mein Kind, dass dieser Person wohl kalt ist.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Mein Kind erkennt, wenn jemand in Not ist und Hilfe braucht.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Mein Kind versteht, dass wenn jemand Angst im Dunkeln hat, er nicht gerne in dunkle Räume geht.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Mein Kind kennt die Bedeutung von dem Verb ‘wollen’.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Mein Kind weiß, dass man sich über die Wünsche anderer Menschen irren kann.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix D2 continued

8. Mein Kind kennt die Bedeutung von dem Verb ‘denken’.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

9. Mein Kind weiß, dass Menschen eher glücklich sind wenn sie bekommen was sie wollen.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

11. Mein Kind begreift, dass sich ein Mensch nicht von einem Tag auf den anderen ändert (bezogen auf sein Verhalten und seine Persönlichkeit).

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiss ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

147
### Appendix D2 continued

14. Der Schein kann trügen. Zum Beispiel würden die meisten Menschen eine Kerze, die aussieht wie ein Apfel, zuerst für einen Apfel halten. Wenn sie genauer hinschauen stellen sie fest, dass das Objekt in Wirklichkeit doch eine Kerze ist und ändern ihre vorherige Meinung. Mein Kind würde in dieser Situation verstehen, dass nicht das Objekt selbst sich verändert hat sondern nur die Meinung über das Objekt.

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>

15. Wenn ich meinem Kind eine Müsli Schachtel zeige, die mit Keksen gefüllt ist und es dann frage: „Was würde jemand denken, was in dieser Schachtel ist, wenn er nicht reingezuckert hätte?“, würde mein Kind antworten, dass die andere Person Müsli in der Schachtel erwartet würde.

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>

16. Wenn ich sage „Jetzt hauen wir uns mal auf’s Ohr!“ weiß mein Kind, dass ich eigentlich meine „Lass uns schlafen gehen!“

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>

17. Mein Kind weiß, dass manche Menschen bewusst lügen um andere in die Falle zu locken oder in die Irre zu führen.

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>

18. Wenn jemand eine Vermutung äußert, begreift mein Kind, dass diese Aussage weniger zuverlässig ist als wenn jemand etwas wirklich weiß.

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>

19. Mein Kind begreift, dass wenn man an einen Keks denkt, man diesen Keks nicht wirklich riechen, essen oder teilen kann.

<table>
<thead>
<tr>
<th>Stimme</th>
<th>Eher</th>
<th>Weiß ich</th>
<th>Eher ja</th>
<th>Stimme</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht zu</td>
<td>nicht</td>
<td>nicht</td>
<td></td>
<td>zu</td>
</tr>
</tbody>
</table>
## Appendix D2 continued

20. Mein Kind weiß, dass man lächeln kann ohne glücklich zu sein.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

22. Mein Kind begreift, dass man manchmal nicht sagt was man denkt um die Gefühle anderer Menschen nicht zu verletzen.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

23. Mein Kind kennt den Unterschied zwischen Lüge und Spaß.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

24. Wenn zwei Menschen ein Objekt von zwei unterschiedlichen Seiten betrachten, kann mein Kind nachvollziehen, dass sie das Objekt verschieden wahrnehmen.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

25. Mein Kind kennt die Bedeutung von dem Verb „brauchen“.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

26. Mein Kind weiß, dass man oft über die Gedanken anderer Menschen nachdenkt.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>
### Appendix D2 continued

27. Mein Kind weiß, dass man oft über die Gefühle anderer Menschen nachdenkt.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

28. Mein Kind kann unterscheiden, ob man jemanden absichtlich oder aus Versehen verletzt hat.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

29. Mein Kind kennt die Bedeutung von dem Verb „wünschen“.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

30. Mein Kind erkennt wenn andere glücklich oder fröhlich sind.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

31. Mein Kind kann so tun als ob ein Spielzeug etwas ganz anderes ist (zum Beispiel, so tun als ob eine Banane ein Telefon wäre).

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

32. Mein Kind bemerkt es, wenn man nicht richtig zuhört.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

33. Wenn ich ängstlich bin, weiß mein Kind, dass die Situation unsicher oder gefährlich ist.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiβ ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>
### Appendix D2 continued

34. Mein Kind versteht das Wort „wenn/falls“, wenn es hypothetisch genutzt wird, zum Beispiel „Wenn ich das Geld hätte, würde ich mir ein großes Haus kaufen“.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

35. Wenn jemand mit seine Händen einen Vogel nachmacht, weiß mein Kind, dass diese Person nicht wirklich denkt, dass dies ein Vogel sei.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

36. Mein Kind kann Geschichten erfinden um zu bekommen was es will.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

37. Mein Kind weiß, dass es beim Verstecken spielen darum geht, möglichst nicht gefunden zu werden.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

38. Mein Kind weiß, dass wenn man ein Versprechen gibt, man es auch halten sollte.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

39. Mein Kind kann sich in andere hinein versetzen und nachvollziehen wie sie sich fühlen.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

40. Mein Kind weiß, dass wenn man ein Geheimnis teilt, man dieses nicht weiter erzählen soll.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Weiß ich nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>
Appendix D2 continued

41. Mein Kind weiß, dass Menschen unterschiedlich sind.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

42. Wenn ich sage „Was ist gelb-braun gestreift und macht ‘mus-mus’? Eine Biene, die rückwärts fliegt!“ versteht mein Kind, dass ich ein Wortspiel gemacht habe.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

43. Mein Kind zeigt mir Dinge.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

44. Mein Kind ist aufmerksam.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

45. Mein Kind kennt die Bedeutung von dem Verb “glauben” (nicht im religiösen Sinne).

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

46. Wenn wir jemanden mögen, dann neigen wir dazu das Verhalten dieser Person positiv zu bewerten. Wenn wir aber jemanden nicht mögen, dann bewerten wir dessen Verhalten eher negativ. Mein Kind versteht, dass Vorurteile und vorgefasste Meinungen über andere unsere Wahrnehmung beeinflussen können.

<table>
<thead>
<tr>
<th>Stimme nicht zu</th>
<th>Eher nicht</th>
<th>Ich weiß nicht</th>
<th>Eher ja</th>
<th>Stimme zu</th>
</tr>
</thead>
</table>

152
Appendix D2 continued

47. Mein Kind kann nachvollziehen, dass zwei Menschen das gleiche Bild ansehen können und es unterschiedlich interpretieren. Zum Beispiel sehen manche Menschen im folgenden Bild einen Hasen, während andere eine Ente sehen.

48. Wenn Moritz ein gemeiner Junge ist und Philip ein liebes Kind, weiß mein Kind, dass Moritz sich eher bösertig und heimtückisch verhalten wird als Philip.
Appendix E

Complement Comprehension Test (CCT)

Picture pairs were presented one after the other and target sentences and questions were accompanied by pointing to the respective picture. Children received one point for each correctly answered item and scored as passing if at least nine out of ten were correct.

Example:

A) The woman said there is a spider in the bathtub (point to 1). But look, it was only hair (point to 2). What did she say was in the bathtub (point back to 1)?

B) The teacher said the girl had a bug in her hair. But it really was a leaf. What did she say the girl had in her hair?

C) The girl said she was reading book. But she really was playing cards. What did the girl say she was doing?

D) The woman told her husband she saw a ghost. But look, it was really a sheet. What did she say she saw?

E) The girl said she was getting dressed. But look, she was really watching TV. What did the girl say she was doing?

F) The girl told the teacher she drew a face. But it was just a scribble. What did the girl tell her teacher she drew?

G) The girl said she found a dollar on the ground. But it was just a piece of paper. What did she say she found?

H) The woman said she was eating an egg. But really it was a ball. What did she say she was eating?

I) The girl said there was a bug in her cereal. But it was just a raisin. What did she say it was?

J) She said her dad had a cut. But look, it was just ketchup. What did she say he had?
Appendix F

Demographic questionnaire

1. Who is filling in this questionnaire and the „Theory of Mind Inventory“?
   Mother □   Father □   Caregiver of the child □   Both parents □

2. Name of the child: ________________________________

3. Date of birth:_____________________  4. gender:   male □   female □

5. Please indicate family status:  6. The child is living with
   Married / Relationship □   the biological parent(s) □
   Single parent □   adoptive parent(s)/ caregiver(s) □
   Other □   Other □   whom? ____________

7. Please indicate the highest educational status of the parents:

<table>
<thead>
<tr>
<th>none</th>
<th>Mother</th>
<th>Father</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school (German: Realschule)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University qualification (German: Abitur)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apprenticeship diploma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University degree</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Please indicate parents' occupational status:

<table>
<thead>
<tr>
<th>Full time</th>
<th>Mother</th>
<th>Father</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In educational program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At home</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. What is your annual combined household income (on average before taxes)?
   Less than $ 25.000 □   between $ 25.000 and $ 65.000 □
   Between $65.000 and $ 130.000 □   more than $ 130.000 □

10. How many hours, on average, do you spend with your child during the week (not counting time the child is sleeping)?   approx. ___ hours

11. How many siblings does the child have and which position does it occupy?
   Number of siblings ____   child position 1. □   2. □   3. □   4. □   _ □

12. Which languages are spoken at home?
   English □
   Another language □   which one? __________________________
   Bi-/ multilingual □   which ones? __________________________

13. Did your child ever receive language training?   □ No   □ Yes, when? ________
14. If you were to estimate the developmental status of your child in comparison to its peers, where would you position your child in the following categories?

<table>
<thead>
<tr>
<th></th>
<th>More advanced than peers</th>
<th>Appropriate to its age</th>
<th>Delayed to its peers</th>
<th>I don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix F continued
### Structure matrix of ToMI factor analysis

<table>
<thead>
<tr>
<th>ToMI Item</th>
<th>ToMI factor 1</th>
<th>ToMI factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.275</td>
<td>0.269</td>
</tr>
<tr>
<td>2</td>
<td>0.556</td>
<td>0.444</td>
</tr>
<tr>
<td>3</td>
<td>0.288</td>
<td>0.400</td>
</tr>
<tr>
<td>4</td>
<td>0.511</td>
<td>0.492</td>
</tr>
<tr>
<td>5</td>
<td>0.263</td>
<td>0.691</td>
</tr>
<tr>
<td>6</td>
<td>0.514</td>
<td>0.610</td>
</tr>
<tr>
<td>7</td>
<td>0.304</td>
<td>0.560</td>
</tr>
<tr>
<td>8</td>
<td>0.626</td>
<td>0.629</td>
</tr>
<tr>
<td>9</td>
<td>0.403</td>
<td>0.604</td>
</tr>
<tr>
<td>10</td>
<td>0.551</td>
<td>0.452</td>
</tr>
<tr>
<td>11</td>
<td>0.466</td>
<td>0.442</td>
</tr>
<tr>
<td>12</td>
<td>0.323</td>
<td>0.269</td>
</tr>
<tr>
<td>13</td>
<td>0.307</td>
<td>0.639</td>
</tr>
<tr>
<td>14</td>
<td>0.487</td>
<td>0.491</td>
</tr>
<tr>
<td>15</td>
<td>0.660</td>
<td>0.480</td>
</tr>
<tr>
<td>16</td>
<td>0.372</td>
<td>0.122</td>
</tr>
<tr>
<td>17</td>
<td>0.616</td>
<td>0.344</td>
</tr>
<tr>
<td>18</td>
<td>0.532</td>
<td>0.612</td>
</tr>
<tr>
<td>19</td>
<td>0.580</td>
<td>0.654</td>
</tr>
<tr>
<td>20</td>
<td>0.461</td>
<td>0.249</td>
</tr>
<tr>
<td>21</td>
<td>0.447</td>
<td>0.338</td>
</tr>
<tr>
<td>22</td>
<td>0.597</td>
<td>0.320</td>
</tr>
<tr>
<td>23</td>
<td>0.675</td>
<td>0.447</td>
</tr>
<tr>
<td>24</td>
<td>0.598</td>
<td>0.460</td>
</tr>
<tr>
<td>25</td>
<td>0.506</td>
<td>0.646</td>
</tr>
<tr>
<td>26</td>
<td>0.647</td>
<td>0.427</td>
</tr>
<tr>
<td>27</td>
<td>0.544</td>
<td>0.498</td>
</tr>
<tr>
<td>28</td>
<td>0.386</td>
<td>0.329</td>
</tr>
<tr>
<td>29</td>
<td>0.627</td>
<td>0.655</td>
</tr>
<tr>
<td>30</td>
<td>0.726</td>
<td>0.485</td>
</tr>
<tr>
<td>31</td>
<td>0.211</td>
<td>0.399</td>
</tr>
<tr>
<td>32</td>
<td>0.367</td>
<td>0.490</td>
</tr>
<tr>
<td>33</td>
<td>0.729</td>
<td>0.539</td>
</tr>
<tr>
<td>34</td>
<td>0.451</td>
<td>0.706</td>
</tr>
<tr>
<td>35</td>
<td>0.485</td>
<td>0.338</td>
</tr>
<tr>
<td>36</td>
<td>0.479</td>
<td>0.261</td>
</tr>
<tr>
<td>37</td>
<td>0.652</td>
<td>0.311</td>
</tr>
<tr>
<td>38</td>
<td>0.430</td>
<td>0.477</td>
</tr>
<tr>
<td>39</td>
<td>0.714</td>
<td>0.384</td>
</tr>
<tr>
<td>40</td>
<td>0.388</td>
<td>0.518</td>
</tr>
<tr>
<td>41</td>
<td>0.520</td>
<td>0.165</td>
</tr>
<tr>
<td>42</td>
<td>0.055</td>
<td>0.466</td>
</tr>
<tr>
<td>43</td>
<td>0.101</td>
<td>0.493</td>
</tr>
<tr>
<td>44</td>
<td>0.557</td>
<td>0.685</td>
</tr>
<tr>
<td>45</td>
<td>0.564</td>
<td>0.155</td>
</tr>
<tr>
<td>46</td>
<td>0.607</td>
<td>0.351</td>
</tr>
<tr>
<td>47</td>
<td>0.387</td>
<td>0.282</td>
</tr>
<tr>
<td>48</td>
<td>0.506</td>
<td>0.646</td>
</tr>
</tbody>
</table>
Appendix G2

**Qualitative item clusters of ToMI factor analysis**

<table>
<thead>
<tr>
<th>ToMI factor 1</th>
<th>cluster</th>
<th>ToMI factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;situational knowledge&quot;</td>
<td></td>
<td>&quot;linguistic knowledge&quot;</td>
</tr>
<tr>
<td>17.</td>
<td>My child understands that people can lie to purposely mislead others.</td>
<td>41.</td>
</tr>
<tr>
<td>26.</td>
<td>My child understands that people often have thoughts about other peoples’ thoughts.</td>
<td>11.</td>
</tr>
<tr>
<td>27.</td>
<td>My child understands that people often have thoughts about other peoples’ feelings.</td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>When we like others, we are likely to interpret their behavior in positive ways and when we don’t like others, we are likely to interpret their behavior more negatively. My child understands that previous ideas and/or opinions of others can influence how we interpret their behaviors.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>My child understands that when someone says they are afraid of the dark, they will not want to go into a dark room.</td>
<td>7.</td>
</tr>
<tr>
<td>22.</td>
<td>My child understands that people don’t always say what they are thinking because they don’t want to hurt others’ feelings.</td>
<td>9.</td>
</tr>
<tr>
<td>38.</td>
<td>My child understands that when a person promises something, it means the person is supposed to do it.</td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>My child understands that when someone shares a secret, you are not supposed to tell anyone.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>If it were raining and I said in a sarcastic voice “Gee, looks like a really nice day outside,” my child would understand that I didn’t actually think it was a nice day.</td>
<td>8.</td>
</tr>
<tr>
<td>23.</td>
<td>My child understands the difference between lies and jokes.</td>
<td>18.</td>
</tr>
<tr>
<td>34.</td>
<td>My child understands the word ‘if’ when it is used hypothetically as in, “If I had the money, I’d buy a new house.”</td>
<td>25.</td>
</tr>
<tr>
<td>42.</td>
<td>If I said “What is black, white and ‘read’ all over? It’s a newspaper!” my child would understand the humor in this play on words.</td>
<td>29.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45.</td>
</tr>
<tr>
<td></td>
<td>Linguistic meaning</td>
<td></td>
</tr>
</tbody>
</table>
Appendix G2 continued

<table>
<thead>
<tr>
<th>ToMI factor 1</th>
<th>cluster</th>
<th>ToMI factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;situational knowledge&quot;</td>
<td>&quot;linguistic knowledge&quot;</td>
<td></td>
</tr>
<tr>
<td>10. If I put my keys on the table, left the room, and my child moved the keys from the table to a drawer, my child would understand that when I returned, I would first look for my keys where I left them.</td>
<td><strong>Understanding of beliefs</strong></td>
<td>35. My child understands that when a person uses his/her hands as a bird, that the person doesn’t actually think it is a real bird.</td>
</tr>
<tr>
<td>15. If I showed my child a cereal box filled with cookies and asked &quot;What would someone who has not looked inside think is in the box?&quot;, my child would say that another person would think that there was cereal in the box.</td>
<td>6. My child understands that people can be wrong about what other people want.</td>
<td>14. Appearances can be deceiving. For example, when seeing a candle shaped like an apple, some people first assume that the object is an apple. Given the chance to examine it more closely, people typically change their mind and decide that the object is actually a candle. If my child was in this situation, my child would understand that it was not the object that changed, but rather his or her ideas about the object that changed.</td>
</tr>
<tr>
<td>24. My child understands that if two people look at the same object from a different standing point, they will see the object in different ways.</td>
<td>19. My child understands that when someone is thinking about a cookie, they cannot actually smell, eat or share that cookie.</td>
<td></td>
</tr>
<tr>
<td>47. My child understands that two people can see the same image and interpret it differently. For example, when looking at this image, one person might see a rabbit whereas another might see a duck.</td>
<td>39. My child is able to put himself/herself in other people’s shoes and understand how they feel.</td>
<td></td>
</tr>
</tbody>
</table>
8. Declaration of oath

Eidesstattliche Erklärung

(gemäß § 6 Abs. 5 der Promotionsordnung Dr. phil. der Universität
Bremen für die Fachbereiche 7 - 12 vom 26.06.2000)

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

On the relation between implicit and explicit Theory of Mind and linguistic competence
- An empirical approach

selbstständig und ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen
Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken
(wörtliche oder inhaltliche Zitate) sind als solche kenntlich gemacht.

Bremen, den
Ort, Datum

Unterschrift des Verfassers