EVIDENTIALITY AND SECOND-ORDER SOCIAL COGNITION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU ARSLAN

COGNITIVE SCIENCE PROGRAMME, INFORMATICS INSTITUTE
MIDDLE EAST TECHNICAL UNIVERSITY, TURKEY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

JANUARY 2012

Approval of the Graduate School of Informatics

_____

Prof. Dr. Nazife Baykal

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Deniz Zeyrek

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____	_____
Prof. Dr. Rineke Verbrugge	Assist. Prof. Dr. Annette Hohenberger
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Cem Bozşahin	(METU, COGS)	_____
Assist. Prof. Dr. Annette Hohenberger	(METU, COGS)	_____
Prof. Dr. Rineke Verbrugge	(RUG, AI)	_____
Dr. Ceyhan Temürcü	(METU, COGS)	_____
Assist. Prof. Dr. Murat Perit Çakır	(METU, COGS)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this wok.

Name, Last name: Burcu Arslan

Signature: _____

ABSTRACT

EVIDENTIALITY AND SECOND-ORDER SOCIAL COGNITION

Arslan, Burcu

M.Sc., Department of Cognitive Science

Supervisor : Asst. Prof. Annette Hohenberger

Co-supervisor : Prof. Rineke Verbrugge

January 2012, 88 pages

In this study, the development of a second-order false belief task is investigated by considering the impact of the acquisition of Turkish evidential markers, namely –DI (direct evidence) and –mIş (inference or hearsay). A neutral version of the tasks served as a control form. 21 kindergarten children (aged 4-5 years), 47 primary school children (aged 6- 12 years) and 10 adults participated in the study. Our results revealed that there is no effect of acquisition of evidentials on false belief understanding. Together with the other studies, there is a facilitative effect of –DI (direct evidence) in understanding of stories/narratives in general rather than false belief understanding for the children at the age of 4 to 6/7. In addition to the second-order false belief tasks (FBT_2), a simple working memory task (WST), a complex working memory task (LST), a perspective taking task (PTT) and a double-embedded relative clause task (REL_2) were used in order to investigate the

developmental trend of these tasks and their possible relationship with second-order false belief understanding. Also, to the best of our knowledge this is the first time that a REL_2 task has been devised in a Turkish study. The general developmental trend was found for all tasks. Even if some significant correlations were found for FBT_2 score predicted from other tasks, analyses showed that only the contribution of age was significant. Since all of these domains are not related to second-order false belief reasoning but develop at the same time, it is not incompatible with the serial bottleneck hypothesis. In sum, the findings are matching with the modularity view that ToM is a faculty of the human mind at their own pace that does not share intrinsic content with other faculties such as language and working memory (Leslie et al., 2004). However, it develops together with those other faculties and they may constrain the expression of child's false belief understanding.

Keywords: Second-order Social Cognition, Cognitive Development, Theory of Mind (ToM), Evidentiality, Language

ÖZ

DELİLE DAYALILIK VE İKİNCİ DERECE SOSYAL BİLİŞ

Arslan, Burcu
Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Yrd. Doç. Dr. Annette Hohenberger
Tez Danışmanı: Prof. Dr. Rineke Verbrugge

Ocak 2012, 88 sayfa

Bu çalışmada ikinci derece yanlış inanç testinin gelişimi Türkçe delile dayalılık belirteçlerinin, yani -DI (doğrudan kanıt) ve -mIş (çıkarım ya da rivayet) eklerinin etkisi dikkate alınarak incelenmiştir.Testlerin nötr (geniş zaman) biçimleri kontrol yöntemi olarak kullanılmıştır. Çalışmaya, 21 anaokul (4-5 yaşlarında), 47 ilköğretim öğrencisi (6-12 yaşlarında) ve 10 yetişkin katılmıştır. Sonuçlar, delile dayalılık belirteçlerinin kazanımının yanlış inancın anlaşılması üzerinde etkisi olmadığını ortaya çıkarmıştır.Diğer çalışmalarla birlikte, 4 ile 6-7 yaşlarındaki çocuklar için –DI (doğrudan kanıt) ekinin yanlış inancın anlaşılmasından ziyade öykülerin/anlatıların anlaşılmasında kolaylaştırıcı bir etkisi bulunmaktadır. İkinci derece yanlış inanç

testine ek olarak basit bir işleyen bellek testi (WST), karmaşık bir işleyen bellek testi (LST), bir bakış açısı alma testi (PTT) ve bir çift girişik ilgi cümlesi testi (REL_2), bu testlerin gelişim eğilimini ve ikinci derece yanlış inancın anlaşılması ile olası ilişkisini incelemek amacıyla kullanılmıştır. Ayrıca, bildiğimiz kadarıyla ilk defa Türkçe bir çalışmada bir REL_2 testi tasarlanmıştır. Genel gelişim eğilimi tüm testlerde bulunmuştur.Diğer testlerden tahmin edilen FBT_2 değeri için bazı anlamlı ilintiler bulunmuş olsa da analizler sadece yaşın katkısının anlamlı olduğunu göstermiştir.Tüm bu alanlar ikinci derece yanlış inanç akıl yürütmesiyle ilgili olmadığından; ancak aynı zamanda geliştiklerinden seri dar boğaz hipoteziyle uyumludur.Sonuç olarak, bulgular zihin kuramının insan aklının kendi çapında, dil ve işleyen bellek yetileri gibi diğer yetilerle özgün içerik paylaşmayan bir yetisi olduğu birimsellik görüşüyle eşleşmektedir.Ancak, diğer yetilerle birlikte gelişir ve bunlar çocuğun yanlış inanç anlayışındaki ifadesini kısıtlayabilir.

Anahtar kelimeler: İkinci Derece Sosyal Biliş, Bilişsel Gelişim, Zihin Teorisi (ZT), Delile Dayalılık, Dil

Annem ve Babam için…

ACKNOWLEDGEMENTS

I am thankful to my supervisor, Assist.Prof.Dr. Annette Hohenberger, for her continuous support, guidance and patience throughout this study. The regular evening meetings that we had always motivated me.Her invaluable comments, feedbacks and diligent contributions made it possible to carry out this research. Also, I would like to express my special thanks to my co-advisor, Prof. Rineke Verbrugge, for leading me to study second-order social cognition and sharing her ideas throughout this study.

I would like to express my thanks to the members of the thesis jury, Prof. Cem Bozşahin, Prof. Deniz Zeyrek, Dr. CeyhanTemürcü, Dr. Murat Perit Çakır for providing important comments and feedbacks when I was constructing the experiments and when I was interpreting the results of the experiments.

I would like to thank the managers and the children of Milli Eğitim Vakfı Koleji, İlkem Koleji, METU Yuva ve Anaokulu and Sosyal Güvenlik Kurumu Anaokulu. Also, I would like to thank, Mesut Aydoğar, for putting me in touch with the managers of the Milli Eğitim Vakfı Koleji and İlkem Koleji. It would not be possible to carry out this study without them.

I would like to thank Liesbeth Flobbe for allowing me to use the drawing for the false belief tasks, and would like to thank, Duygu Özge, for allowing me to use thedrawings of single-embedded relative clause task and commenting on the double-embedded one that we constructed.

I am also grateful to my friends, Kadir Kozan and Sevilay Karahan,for their technical support. Whenever I asked for help, they helped me without hesitation.

Another special thanks go to Ülge Uğurlu and Sertaç Tezcan for supporting and motivating me all the time.

I would like to thank to my managers, Mehmet Emin Bilmez, Murat Zorluoğlu and Dr. İlker Haktankaçmaz, and also to my colleauges, Esin Özdemir, Gülsüm Belge, Arzu Kavuşmuşoğlu, and Ali Rıza Kurşunlu for their understanding and support during this study.

Lastly, I wish to express my sincerest gratitude to my parents for their constant and unconditional support and understanding. This study is exclusively dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accusative Marker
FB	False Belief
FBT	False Belief Task
FBT_2	Second-order False Belief Task
LST	Listening Span Task
METU	Middle East Technical University
MEV	Milli Eğitim Vakfı
PAST-PROG	Past Progressive
PTT	Perspective Taking Task
RC	Relative Clauses
REL_2	Double Embedded Relative Clause Task
ToM	Theory of Mind
WM	Working Memory
WST	Word Span Task

CHAPTER 1

INTRODUCTION

In daily life, we are constantly in interaction with other agents, such as co-workers, friends and family members. As a result of this interaction, we form models pertaining to the different mental states of other agents. Social cognition of individuals is shaped based on these models. The ability to understand that different agents have different mental states, such as desires, beliefs, knowledge and intentions, which can be different from one's own, is called Theory of Mind (ToM) (Premack& Woodruff, 1978).

Zero-order, first-order, second-order and higher-order reasoning are different levels of social cognition. The objects of zero-order reasoning are the rules of nature and real-life environment. For instance, if David knows "There is an apple on the table", he is applying zero-order reasoning. However, in daily life we are not just talking about world facts. Social interaction covers statements such as "David thinks Jessica knows that there is an apple on the table". In this situation David is applying first-order reasoning by attributing a mental state to Jessica.In addition to first-order reasoning, social interaction covers more complex social situations like "Jack thinks David knows that Jessica knows that there is an apple on the table". This time, Jack is applying second-order reasoning by attributing a first-order reasoning to David who attributes a mental state to Jessica. In this study we follow Verbrugge (2009) in using the term 'second-order social cognition' in the same sense as 'second-order theory of mind'. The usage of this terminology aims to investigate the theory of mind without preferring the 'theory-theory' approach to the 'simulation theory' approach.

First-order theory of mind develops between ages three and five (Wimmer&Perner, 1983). Interestingly, second-order ToM develops much later than first-order reasoning, between the ages of six and nine (Perner, 1988; quoted in Verbrugge, 2009). The reason for this gap has not been clarified yet, and attracts the curiosity of researchers who are working on theory of mind. In Verbrugge (2009), it is hypothesized that the developmental latencies between first and second-order social reasoning is due to the children's need to overcome constraints on serial processing rather than simple working memory capacity. More explicitly, 6 year-old children do have the ability to represent other's mental state about their own mental state. However, they cannot apply this because of the lack efficiency in serially applying the related mental processes (cf.Hendriks et al., 2007).

Studies of theory of mind can be grouped under three headings. These headings are referred to as (1) structures of mental states, (2) development of these structures, and (3) theoretical analysis of this development (Astington&Baird, 2005). There are different paradigms in studying the development of theory of mind. These paradigms can be grouped as verbal and non-verbal. In the following, I will briefly discuss the two verbal paradigms and then one non-verbal paradigm. One of the most widely applied verbal paradigms is the false-belief task (FBT), which has first been studied by Wimmer and Perner (1983). The main idea of the false-belief task is to examine whether children can attribute a false belief to other agents in a given story where they know the reality and the other agents do not. Mostly, the false belief task contains five types of questions that help understanding false belief. After the first part of the story has been told to the participant, a *reality control question* is asked in order to make sure that the participant understood the story. Then, the experimenter continues to tell the story. Subsequently, an *ignorance question* is asked as a control question to verify the absence of knowledge, followed by a *linguistic control question*. Finally, the *false belief question* and a *justification question* are asked to the participant. In this study we focused on the development of second-order social cognition by applying a second-order false belief task to Turkish children in the appropriate age range.

Using language comprehension tasks is another verbal paradigm in the study of the development of social cognition. These tasks generally test listeners' semantic and/or pragmatic inferences. In these tasks, the listener has to take the speaker's linguistic alternatives into account to understand the correct meaning of the sentence. In this study, a complex language comprehension task was used to test children's ability to meet the listener's expectations while the speaker gives an answer.

As regards non-verbal paradigms, strategic games are among the most common examples (Hedden et al., 2002; Flobbe et al., 2008). Since strategic games require the representation of the opponent's mental states, it is highly dependent on the different levels of theory of mind. Moreover, strategic games are applied tasks that do not directly depend on language. Because of the time constraints, the strategic games were not used in this study.

The development of theory of mind has been largely investigated and documented in the literature (for recent monographs on the topic, see Doherty, 2009; Saxe & Baron-Cohen, 2007; Apperly, 2010). However, one of the debatable issues is still how children acquire this ability. There is one influential factor as regards language development (Astington & Baird, 2004; Hollebrandse et al. 2011; Garfield et al., 2001; Schick et al., 2007; Flobbe et al. 2008): Does language have an effect on acquiring this ability, or not? Since language has different levels such as phonology, morphology, pragmatics, semantics and syntax, it is important to distinguish these while searching answers to this question. In this study, the morphological structure, in particular evidentiality markers in Turkish (in the second-order FB task) and zero vs. accusative markers (in the complex language comprehension task), and also syntactic structure, namely relative clauses, were investigated in order to understand the relationship between language and social reasoning during development.

Since evidentiality markers allow speakers to encode different sources of knowledge, it can be important in the development of social reasoning. If evidentiality markers exist in a specific language, they are marked lexically or morphologically (Aikhenvald 2004, Fitneva and Matsui 2009). For example, in English and French, the evidentiality marker is a lexical element, e.g., "apparently", "according to",

whereas in Turkish and Korean, it is a morphological one. Examples of different sources of knowledge indicated by evidential markers are direct experience or indirect experience. While direct experience refers to everything that we observed or witnessed in the past, indirect experience can be explained as hearsay or inference (Plungian, 2001; quoted in Özoran, 2009). In Turkish, it is compulsory to use evidential markers when referring to the past. There are two different suffixes used according to the source of information in past tense. The evidential marker –DI refers to the direct experience of the speaker, while the evidential marker –mIş refers to hearsay or inference:

(1) Kız gel -di.
'The girl came.' (I saw that the girl came.)
(2) Kız gel -miş.
'The girl has come.' (I heard or inferred that the girl came.)

However, there are also different usages of –mIş, even if the speaker is direct experiencer of the events. It is also used for telling stories, for pretend play, for expressing surprise and also for reporting unconscious events. In some cases the usage of –mIş can be replaced by –DI. For example, in the sentence: "1980 yılında doğ-du-m" (I was born in 1980).

The main goal of this study is to investigate the effect of the acquisition of evidential markers on the development of second-order false belief understanding in Turkish children between the ages four (kindergarten) and twelve (fifth grades). In addition, the study also aims to investigate the relations between false belief and complex language and relative clause understandings. This study also tried to investigate Verbrugge's (2009) hypothesis that the developmental latencies between first- and second-order social reasoning is due to the children's need to overcome serial processing rather than simple working memory capacity by looking at the correlations between social cognition tasks and simple and complex working memory tasks.

CHAPTER 2

LITERATURE REVIEW

In this study, the relationship between evidentiality and second-order social reasoning is investigated. In the first subsection, an overview of the development of theory of mind will be given. In the second section, the acquisition of evidentiality will be clarified. In the third section, the role of evidentiality in theory of mind research will be described. In the remaining two sections, the relationship between ToM and working memory and complex language abilities will be explained.

2.1 The Development of Theory of Mind

Disregarding the discussion whether theory of mind is innate or not (cf., Leslie et al., 2004; Gerrans, 2002), it is obvious that different levels of theory of mind and precursors of theory of mind develop with age. Already infants around 9 months of age can perceive human action as goal-directed (Gergely et al., 1995; Wellman & Philips, 2001; Woodward, 2001; quoted in Malle, 2002). 2 year olds not only have the ability to mimic an action (de Villiers, 2007), but also engage in pretend play and have an understanding of desire (Flobbe et al., 2008). However, they cannot distinguish an external goal from an internal one (de Villiers, 2007). The understanding of belief develops one year later than that of desire (Malle, 2002). While children at age 3 cannot understand a verbal false belief task, children at age 4 can understand it (Wimmer&Perner, 1983). However, Onishi and Baillargeon (2005)

studied a non-verbal version of the false belief task with toddlers. They concluded that even 15 months-old toddlers wereable to pass the false belief task. These infants looked longer at a person with a False Belief about the true location of an object than at a person with a true belief. When this result is compared with the verbal false belief task, it can be said that understanding of the verbal false belief task develops much later than the non-verbal one.

Compared to first-order false belief tasks, studies of higher-order false belief tasks are scarce in the literature. After Wimmer and Perner's (1983) seminal study of first-order false belief task, Perner and Wimmer (1985) tried again to shed light on the comprehension of second-order false belief. They concluded that this ability does not develop before the age of 6. In a very recent study of Hollebrandse et al. (2011), the ability to understand second-order verbal and non-verbal false belief tasks were investigated with 6 to 9 year old children. Their results showed that children performed better in verbal second-order false belief task than non-verbal one. When the results were compared to Onishi and Baillargeon's (2005) findings, they concluded thatunlike the first-order false belief reasoning, language facilitates second-order false belief understanding.

Flobbe et al. (2008) studied the development of second-order theory of mind in children between the ages 8 and 10 by using false belief and strategic game tasks. She used two false belief stories, a strategic game, and a language comprehension task in her study. The first false belief task was the adapted version of Sullivan et al.'s (1994) 'Birthday Puppy' story and the second false belief task was the adapted version of Hogrefe and Wimmer's (1986) first-order 'Chocolate Bar Story'. Flobbe used her own drawing during the experiment. As a strategic game, a modified version of Hedden and Zhang's (2002) matrix game was used. The strategic game was played on a computer. The participant and the computer opponent were sequentially controlling a car. There were three decision points in the game where the participant or the computer opponent had to decide to move to the next decision point or stay at the current point. In each decision point there were different or same number of marbles for each player, which represents the reward of the each participant. The participant was told to maximize her own reward and was told that

the computer opponent would try to do the same. When one of the participants decided to stop at the decision point, each participant took the rewards at that point. This required them to reason about their opponent's moves in the game. The results of the matrix game revealed that children performed much better (93%) in the phase that needs first-order reasoning than the phase that needssecond-order reasoning (57.2%). Even though adults perform better than children and better than subjects in Hedden and Zhang's (2002) study (where they showed only 60% - 70% success), they could not apply second-order reasoning reliably (75.5%). Flobbe et al.'s study revealed that succeeding in a second-order false belief task is a necessary but not sufficient condition in second-order reasoning in the strategic game. Flobbe et al. (2008) could not find any relation between the false belief task, the strategic game, and the language comprehension test, either.

More recently, Meijering et al. (2010) used the Marble Drop game with 22 adults, which is logically equivalent to Hedden and Zhang's (2002) matrix game and Flobbe et al.'s (2008) strategic game.The game is played on a computer with a computer opponent. They are presented with a marble and trapdoors in the game. After the zero- order and first-order training blocks, 8 second-order games were presented to the participants. In the second-order test game, there were four bins with payoffs for each player. During the game, players should choose one of the two trapdoors to guide the marble into the preferred bin. According to the choice of the player, the marble drops into the bin or into the next trapdoor that ends with another bin. The aim of the game was to get the highest payoff. The results revealed that participants applied second-order reasoning much better (94%) than subjects in Hedden and Zhang's (2002) matrix game and in Flobbe et al.'s (2008) strategic game which supports the idea of a facilitative effect of the context.

Liddle and Nettle (2006) studied higher-order theory of mind up to the fourth level in 10 and 11 year old children by using five stories. They found out that 10 and 11 year old children successfully understand the first-order and second-order stories, whereas they perform at chance on the fourth level. They also correlated the performance of theory of mind with teacher ratings of the child's social skills and found that it was positively correlated with these ratings. More challenging stories have been applied

in the study of adults' understanding of higher-order theory of mind. The studies reveal that adults' performance on these stories is better than chance up to level four, but after this level the error rate is very high (Liddle& Nettle, 2006).

These findings together indicate that second-order and higher-order social reasoning is a different milestone from first-order social reasoning.

2.2 Acquisition of Evidentiality

Most of the acquisition of evidentiality studies comes from Turkish and Korean whereevidential markers are marked morphologically. In Turkish, it is compulsory to use evidential markers when referring to the past. Early studies of Aksu-Koç (1988) revealed that the first productions of –DI and –mIş appear between the ages of 2 and 3. Beyond the production of these morphemes, children start to use –DI for direct evidence around at the age of 3 and a half and –mIş for inference around the age of 4 and a half (Aksu-Koç, 1988). The reason behind this late development is the different usages of –mIş. In addition to the evidential usage of –mIş, it is also used for telling stories, for pretend play and for expressing surprise. The delay in the acquisition of –mIş as an evidential marker is considered to be due to this multipurpose nature (Aksu-Koç, 1988).

Aksu-Koç (1988) conducted another experiment to investigate whether this usage occurs with the full understanding of evidentials or not. In her study, she used a doll to tell the events to the children by using –DI for direct evidence and –mIş for hearsay. She asked the children whether the doll had seen the event or had heard about it. The results of the experiment showed that even 6 year-olds could not fully understand the proper conditions for using evidentials. More recently, Aksu-Koç and Alıcı (2000) replicated the results of previous work (quoted in Papafragou et al., 2006). The reason behind the very early production of –DI and –mIş but late acquisition of its evidential usage is explained by Aksu-Koç (1988) as follows:

> Children's early lack of sensitivity to the distinction between direct and indirect experience suggests that they are more attentive to concrete, referential and objective characteristics of situations than to subjectively

> relevant distinctions such as the speaker's attitude to the proposition asserted. (p.195; quoted in Papafragou et al., 2007)

Korean also has the sentence-ending morphologically marked evidentials as Turkish. The morphemes –e and –ta, which differ in terms of the degree of the assimilated knowledge, are used for direct evidence and arelike the morpheme – DI in Turkish, and the morpheme –tay is used for hearsay like the morpheme –mIş in Turkish (Papafragou & Li, 2001). However, Korean has a different morpheme (-kwun) for the inference, unlike Turkish. WhileKorean children start to use direct evidence morphemes –e and -ta around the age of 1;9, the usage of the hearsay morpheme –tay appears before 2;5 and children productively use all of the morphemes at the age of 3 (Choi, 1995; quoted in Papafragou & Li, 2001).

These results show that although both Turkish and Korean have sentence-ending morphologically marked evidentials, Korean children's acquisition of evidentials develops earlier then Turkish children. However,it is clear that the ability to understand direct evidence for both languages develop earlier than the hearsay.

2.3 Evidentiality and Theory of Mind

Theory of mind is the ability to understand that different agents may have different mental states, such as desires, beliefs, knowledge and intentions (Premack& Woodruff, 1978). In social cognition it is important to remember where, when, and from whom information comes. This encoding is called "source monitoring" (Schacter, Kautstall, & Norman, 1997; Lindsay, Johnson, & Kwon, 1991; quoted in Ögel, 2007). Linguistic evidentiality allows us to reason about the evidence with respect to a certain piece of information. This ability is important in storing and updating information, and leads us to understand that people hold different beliefs or knowledge, which is part of theory of mind. Because of this, studying the relationship between evidentiality and theory of mind attracts researchers who want to investigate the interaction of language and thought (Gleitman & Papafragou, 2005). Since Turkish evidentials –DI and –mIş are obligatory for past reference, they are good candidates for studying this interaction.

Some cross-cultural studies revealed that different categories of theory of mind, related to different intentional states such as beliefs, desires, intentions, emotions and knowledge develop at different ages for different cultures (Wellman et al. 2006). Bayramoğlu and Hohenberger (2007) adapted Wellman and Liu's (2004) ToM scale for Turkish and conducted an experiment with Turkish children at the age of 4 and 5 to explore the cultural influences of the development of the different categories of theory of mind. They found that while Turkish children had a better performance than Western children in knowledge and emotions domains, they had a worse performance in the belief domain. They partly related these differences to the morphological structure of the evidentials in Turkish insofar as these stories contained such evidentials. After that, Özoran (2009) studied the development of evidentiality and theory of mind by using the previously adapted Turkish version of Wellman and Liu's (2004) ToM scale to investigate this possible explanation with 4 to 7-year old Turkish children. He used three different versions of the stories. As a control form in the first version, he told the stories without using evidential markers (NEUTRAL). In the two remaining versions, he told the stories by using –DI and –mIş evidentials. His research findings showed that children's performance on the ToM stories using the –DI form but not -mIş were significantly better than the neutral ones. This finding may be counted as evidence that the use of the direct evidential marker –DI facilitates reasoning about other people's mental states at that age. However, in a recent control study where the effect of the same three conditions (neutral, -DI, and –mIş) on the understanding of stories not involving theory of mind was tested with pre-schoolers, the same facilitative effect of –DI over the neutral version and –mIş was found (Gözenman, 2010). This result sheds doubt on the claim that evidential markers directly act on ToM understanding. It is compatible with the view that they generally facilitate understanding of narratives at that age.

Papafragou et al. (2006) also studied comprehension and production of the Korean evidentials –e (direct evidence) and –tay (hearsay), and also they compared Korean children's source monitoring abilities with the English children. According to their results 3- and 4-year old Korean children could not understand the –tay (hearsay) in the comprehension tasks. They also found no significant difference between Korean and English- speaking children in terms of their source-monitoring abilities. Their

findings revealed that the evidential markers do not depend on children's ability to reason about the source of information. Finally, they emphasized that non-linguistic source monitoring tasks should be used in order to evaluate the ability of children to take the source of information into consideration rather than linguistic tasks involving evidentials.

Ögel (2007) conducted an experiment with 3- to 6-year-old children in order to test the hypothesis that the evidentials are positively correlated with the non-linguistic source-monitoring abilities of Turkish children. She used three different language tasks, namely direct experience, inference and reportative markers. In the direct evidence task, children watched an event and were askedto report that event immediately (production of –DI). In the inference task, children were asked to comment on the presented events in which they did not witness (production of –mIş).Lastly in the reportative markers task, they were asked to rephrase the story, which they heard from someone else in the form of direct experience.Her findings replicated Aksu-Koç's (1988) earlier findings for the production of evidential markers. She also used two different source-monitoring tasks, namely mode-of-knowledge acquisition task (adapted from Gopnik and Graf, 1988) and the source memory task (adapted from Drummey and Newcombe, 2002). In the mode-of knowledge task six boxes were presented to the participants and they were asked to find out the contents of the boxes by looking, guessing and being told about the content and then they were expected to report how they had found out the content of each box. They found no significant relationship between the use of evidentials and source-monitoring ability. For the source memory task, ten novel facts were introduced to the children. Some of the facts were introduced by the experimenter and the others were introduced by the Puppy. The children were asked to recall the facts and the sources one week later. She concluded that there is a significant relationship between the ability to use reportative –(I)mış with the ability to recall the source of information.

2.4 ToM and Working Memory

In the literature, some researchers revealed that the development of ToM goes parallel with cognitive development (Hala et al., 2009; quoted in Özoran, 2009). Working memory, which is an active sub-module of short-term memory with its active attention device, the "central executive" (Baddeley 2003), is one of the elements of cognitive development (Gathercole 1999). Gordon& Olson (1998) (quoted in Özoran, 2009) found that children's ToM understanding is highly correlated (r=.64) with their working memory (WM) capacity. In order to investigate this correlation, two working memory tasks were added in this study. One of them is a simple working memory task, namely Word Span Task (WST). The Word Span Task is a simple verbal working memory task related to the phonological loop component of Repovs & Baddeley's (2006) model of working memory. This task was adapted to Turkish in Ünal's (2008) Master's Thesis. She conducted a study with Turkish children from grade 1 to grade 5. She found that the WST develops linearly. Özoran (2009) also investigated the effect of WST on ToM with the children from 4 to 7 years of age. He divided the data into two groups as younger (3;6 to 5;6 years) and older (5;7 to 7;5 years). His results revealed that there is no significant difference on WST score between the groups. He also investigated the relation between WST and ToM. He found that the WST was not a predictor of ToM. Despite these negative findings, the Turkish WST of Ünal will be used in the present study.

Hasselhorn et al. (2005) also studied the relation between phonological working memory and second-order false belief performance of children from 4 years to 6 years of ages. They found that there is a high developmental dependency between the children's second-order false belief performance and their phonological working memory capacity.

Since one of the aims in this study is to investigate whether the developmental latencies between first and second-order social reasoning is due to the children's need to overcome serial processing rather than simple working memory capacity, a complex working memory task, namely the Listening Span Task (LST) was also

added to the study. Different from the WST, LST performance requires attention-allocation to two different tasks, serial processing and storing of information. This task was also adapted from Ünal's (2008) Master's Thesis. In her study, she found a step-wise development of LST across age.

2.5 ToM and Complex Language

Flobbe et al. (2008) studied the relation between a sentence comprehension task and second-order ToM reasoning with children from 8 to 10 years of ages. The experimenter told two stories involving the use of indefinite or definite articles for marking the subject. After each story the participant heard one canonical Dutch sentence in which the subjects appears initially ("Een meisje ging twee keer van de glijbaan af.", in English "A particular girl went down the slide twice.") or existential Dutch sentence in which the subject appears internally ("Er ging twee keer een meisje van de glijbaan af.", in English "Twice a girl went down the slide."). The participantwas expected to judge whether the sentence was correct or not. De Hoop and Krämer (2005/2006) (quoted in Flobbe et al., 2008) argues that independent subjects are interpreted referentially. However, since the speaker chooses the marked existential word order instead of the best canonical word order, it leads the hearer to the non-referential reading which is not 'a particular girl' but 'any girl'. According to this bidirectional Optimality Theory explanation, speakers take into account the hearers' perspective when expressing the idea, and also the hearers interpret the meaning by taking into account the speaker's perspective (Blutner, 2000; quoted in Flobbe et al., 2008).The results revealed that 9-year-old children could not reason about the speaker's alternatives with regard to the use of indefinite subjects. Also, Flobbe et al. (2008) could not find any significant relationship between thesentence comprehension task and the second-order false belief task.

According to de Villiers & de Villiers (2005), the syntactical component of language is related with ToM. Generally, complement clauses (e.g.John knows that Mary loves apples) are used to investigate this relationship. Relative clauses, like complement clauses, can be used recursively. At each level of recursion they refer to a different subject or object. However, relative clauses do not necessarily involve

mental state predicates such as "knowing that" or "believing that". Using relative clauses instead of complement clauses allows us to specifically focus on the structural format of 2-way embedding. This is a purely structural parallel between 2nd order embedding in the thought domain and 2nd order embedding in the language domain.

In Özoran's (2009) Master's Thesis study a first-order relative clause task, which was adapted from Özge's (2010) PhD thesis, was used to investigate the relationship between relative clauses (RCs) and ToM scores. Özoran (2009) had found that relative clause task is a good predictor for ToM scores. In this study, we also used Özge's (2010) stimuli in our relative clause task by modifying them to double-embedded RCs, hence we call it "REL_2". Özge et al. (2009) had conducted an experiment with 37 monolingual 5 to 8 year old children to test the subject-object asymmetry in Turkish RCs. The authors found that children's performance in subject RCs (96.45%) was higher than in object RCs (66.72%). They pointed out that this asymmetry was related to morphosyntax in addition to embedding. Since our aim is to investigate children's ToM abilities and not their different abilities in subject *vs*. object RCs, only one type of RCs, namely subject RCs, were used in our task. We decided to use subject RCs since they are more straightforward to understand. Thus, we were able to focus entirely on the embedding aspect of RCs.

In addition to the relative clause task, a complex language task was constructed to investigate the relationship between pragmatic inferential abilities and ToM understanding. We named this task "perspective-taking task" (PTT). The perspective-taking task includes two questions in order to understand the participant's ability to meet the speaker's expectations when answering their questions in a given context (see methods section).

CHAPTER 3

RESEARCH QUESTIONS AND HYPOTHESES

The general research questions of this study are the following ones:

1. Is there a developmental trend in the performance of kindergarten, 1st, 3rd, and 5th graders in the studied tasks: Second-order False Belief Task (FBT_2), WST, Perspective Taking Task (PTT), Double-embedded Relative Clause Task (REL_2), LST? Adults should outperform children in all tasks.
2. Is there any facilitator effect of acquisition of Turkish evidential markers on the development of the second-order false belief task?
3. Is the perspective-taking task, in which accurate comprehension entails reasoning about the speaker's linguistic alternatives, related to the second-order false belief task?
4. Is understanding of relative clauses which contain complex syntax related to the second-order false belief task?
5. Is the acquisition of second-order social cognition a question of a processing bottleneck rather than a question of simple working memory capacity?

The hypotheses of the study related to the above research questions are the following ones:

- **H1: Main effect "age":** Since the previous research revealed that the development of second-order social cognition occurs between the ages 6 to 9, I also hypothesize a developmental trend, that is, older children (e.g., 11 year

olds) will be more successful in understanding second order false belief tasks than younger children (e.g., 4year olds). Likewise for the other tasks: WST, PTT, REL_2, and LST, we expect an age effect.

- **H2: Main effect "evidentiality":** Aksu-Koç (1988) found that full acquisition of the evidential marker –mIş occurs only after the age of 6. Özoran (2009) found a facilitator effect for -DI as opposed to the neutral version in 4-6 year old children, so I would also expect the same effect in this study.
- **H3: Interaction of "age x evidentiality":** As the evidential –mIş develops later than –DI, I expect possible differences between the understanding of the stories marked with –mIş and those marked with –DI for younger and older children, that is, older children may profit more from -mIş than younger ones.
- **H4: Sentence comprehension predicting second-order false belief:** Discourse and sentence comprehension, which considers taking into account speakers' linguistic alternatives, should be a predictor of false belief understanding**.**
- **H5**: **Double-embedded relative clauses predicting second-order false belief**: Embedding, as tested in the relative clause task, should also be a predictor of false belief understanding.
- **H6**: **Processing bottleneck:** In the same vein of Hendriks et al.'s (2007) hypothesis in Verbrugge's (2009), it is hypothesized that the developmental latencies between first- and second-order social reasoning is due to the children's need to overcome serial processing rather than simple working memory capacity. The testing of this hypothesis, however, is not as straightforward as the above hypotheses. It will be aimed at evaluating this hypothesis by means of looking at the results of various tasks, among them the "Listening Span task" which tests Complex Working Memory and the "Word Span task" which tests simple verbal working memory capacity.

CHAPTER 4

METHOD

4.1. Participants

A total of 68 (35 female, 33 male) children and 10 (5 female, 5 male) adults participated in the experiments. The adults served as a control group. Children's grades varied from kindergarten to fifth-grade, and their age range was from 4 to 12 years.

In the initial stage of the experiment, as a prerequisite of academic study, the related ethical procedures were completed prior to the identification of individual participants. This procedure involved the preparation of the application form, parent approval form, voluntary participation form, project information form, post-participation form and samples of tasks to be used in the experiment. These documents are officially required by the Middle East Technical University (METU) Research Centre for Applied Ethics. Upon receipt of approval from the Ethics Committee of the Research Centre, the first request to conduct the experiment was submitted to the METU College. After a series of consultations with the officials at the METU College, unfortunately the experiment request was rejected due to reluctance of the management. As a result, other requests were sent to Milli Eğitim Vakfı (MEV) College and İLKEM College, both of which accepted to circulate the parent approval forms among their students. The experiment was then started to be

conducted with those students whose parents provided the school with approvals. After finishing the experiments with primary school children, the study was extended to the kindergarten children. The same ethical procedures were completed and the experiments were done with the METU Kindergarten and SGK Kindergarten. The experiment with the adults was conducted upon the signature of voluntary participation forms.

The descriptive statistics related to the participants are shown in Table 1 and Table 2.

Table 1: Descriptive statistics of the children and adult groups (in years of age)

Age (in years)	N	Minimum	Maximum	Mean	Std. Deviation
Children	68	3.83	11.53	7.53	2.53
Adults	10	19.61	50.33	33.48	10.00

Table 2: Descriptive statistics of each grade

Grades	N	Minimum	Maximum	Mean	Std. Error
Kindergarten	21	3.83	5.03	4.43	.07
Grade 1	17	6.08	7.48	6.99	.09
Grade 3	15	8.53	9.50	9.01	.08
Grade 5	15	10.35	11.53	11.00	.10

4.2. Design

A within subject design was used in the experiment with the exception of the three versions of the second-order false belief task where between subject design was used. All subjects participated in the following five tests:

- word span task (WST)
- second-order false belief task(FBT_2)
- perspective taking test (PTT)
- second-order relative clause task (REL_2)
- listening span task (LST)

All of the tests were completed in one session, which varied from 25 minutes to 35 minutes. Children were tested in a quiet empty classroom at their schools. Adults were tested in a meeting room at the Ministry of Interior, Ankara. The answers of the subjects were recorded via voice recorder.

4.2.1. Word Span Task (WST)

Material

To be able to measure the working memory of the participants, Ünal's (2008) English–to-Turkish adaptation of the original WST (Pickering&Gathercole, 2001, as cited in Ünal, 2008) was used. The task consists of one-syllabic words from Turkish. The words such as "saç, tuz and yurt" (hair, salt and country) were selected considering their frequency in daily usage and easy pronunciation. There are a total of seven sets, which consist of 2 to 8 words. Each set is comprised of 3 sub-sets. An example of a set of 2 words as follows (see Appendix A for the entire material):

1. köşk muz (manor banana)
2. pil üst (battery upper)
3. buz dört (ice four)

Procedure

The words from these sets were read to the participants starting from the set of 2. After reading one set (i.e. köşk muz), the participant repeated the words in that order. If the participant makes less than two errors, i.e., any error in two of the three sub-sets of that level, the subsequent, next higher, set was read (i.e. the set of 3 words). If s/he makes two errors, the experiment was terminated. The word spanequals the correct number of words at the respective level at which the child makes less than two errors. Thus, in the analysis the word span rage varies between 0 and 8. This task is adapted from Gülten Ünal's Master Thesis (2008) with permission.

4.2.2. Second-order False Belief Task (FBT_2)

Material

The study consists of two different second-order false belief stories, namely the 'Birthday Puppy' Story and the 'Chocolate Bar' Story. Both stories were adapted from English to Turkish from Flobbe et al. (2008) with the author's permission. These stories were told to the subjects by presenting Flobbe et al.'s (2008) drawings also used with permission. The grandmother character was added in order to make the drawing more explicit in Flobbe et al.'s (2008) drawing of the Birthday Puppy Story. Figure 1 demonstrates the drawings related to the chocolate bar story. The English version of the text of the story is given below:

> John and Mary are brother and sister. Here they are in the living room. Then mother returns from shopping. Mother bought some chocolate. She gives the chocolate to John. Mary doesn't get any chocolate, because she has been naughty. John eats some of the chocolate and puts the remainder in the drawer. He doesn't give any of the chocolate to Mary. That makes Mary angry. Now John goes to help mother in the kitchen. He is helping with the dishes. Mary is alone in the living room. John is in the kitchen. Because she is angry with John, Mary hides the chocolate. She takes the chocolate out of the drawer and puts it in the toy chest. John is busy doing dishes. He throws the fruit leftovers in the rubbish bin in the garden. Through the window he sees the living room. He sees how Mary takes the chocolate out of the drawer, and puts it in the toy chest. Mary does not see John.
>
> *Reality control question:* Where is the chocolate now?
>
> *1st order ignorance:* Does John know that Mary has hidden the chocolate in the toy chest?
>
> *Linguistic control:* Does Mary know that John saw her hide the chocolate?
>
> John has finished the dishes. He is hungry. Now he wants to eat some of his chocolate. John enters the living room. He says: "Hmm, I would like some chocolate."

2nd order false belief: Where does Mary think that John will look for the chocolate?

Justification: Why does she think that?

Figure 1: The drawings used for the chocolate bar story (Flobbe et al., 2008)

Second-order embedding structures such as "Mary thinks that John thinks the chocolate is in the drawer." were not used in the stories. In this way, second-order reasoning can be tested without testing child's ability of processing second-order embedding structures.

Since this research's main goal is to investigate the effect of Turkish evidentials on the understanding of children's ToM, three different versions of the stories in the Turkish language were constructed. In the neutral version, the story was told by using present tense indicating that a direct experience of the present events by using Turkish present tense (imperfective) marker '-Iyor' and aorist marker '-Ar'. In the –DI version, the story was told by using past tense indicating a direct experience of the past events by adding the marker –DI to the verb stem. In the –mIş version, the story was told by using another past tense indicating an indirect (hearsay) experience of the past events by adding the marker –mIş at to the verb stem. In the neutral control version, the story was told by using present tense. The same modifications were applied to the 'Birthday Puppy Story', as well.

Only one of these versions was presented to each subject, that is, a between subjects design was used for this task. The three Turkish versions of the stories used in the experiment, including the drawings of the 'Birtday Puppy Story', can be found in Appendix B.

Procedure

For both stories, the drawings were shown to the participants when the stories were being told. Since Flobbe et al. (2008) stated that younger children have a higher error rate in responding to the Birthday Puppy Story than to the Chocolate Bar Story, the order of stories in the false belief task was balanced. The drawings for the stories were presented on a table. While the stories were being told, the related parts of the drawings were pointed out to the participants.

If a participant gave correct answers to the reality control, first-order ignorance, linguistic control and second-order false belief questions, the participant's score of the first story was 1. The total score for both of the false belief stories is therefore minimum 0 and maximum 2. The analysis of the justification question was done separately. Since the questions before the second-order false belief question are control questions, the prerequisite of analyzing the score is being successful of them.

4.2.3. Perspective-taking Test (PTT)

Material

The perspective-taking test includes two close-ended questions with two options. The English version of the text of the story is given below:

> Ayşe and Ali are siblings. They are talking to each other. Ali tells Ayşe that he is planning to go to the bookstore today. Ayşe wants Ali to buy a storybook. Ali goes to the bookstore and buys the book. While Ali is going back home, he sees his friend Mehmet on the road. Mehmet asks Ali what he did today.

Question: Which answer does Ali give to Mehmet?

a) Kitab-ı al-dı-m.
   Book-ACC buy-PAST-PROG
   'I bought the book'

b) Kitap al-dı-m.
Book buy-PAST-PROG
'I bought a book'

After that, Ali goes back home. Ayşe opens the door and asks Ali what he did today.

Question: Which answer does Ali give to Ayşe?

a) Kitab-ı al-dı-m.
   Book-ACC buy-PAST-PROG
   'I bought the book'

b) Kitap al-dı-m.
   Book buy-PAST-PROG
   'I bought a book'

The order of the answers to the close-ended questions provided to the subjects was balanced across participants. Since Mehmet asks a more general question to Ali, the expected answer for the first question was "Kitap aldım" rather than "Kitabı aldım". More explicitly, if a participant correctly understands that Mehmet asks the question just for general conversation, s/he will think that Ali knows that Mehmet does not know that Ali went to the bookstore to buy a storybook that Ayşe wanted and s/he will give the answer "Kitap aldım". Since Ayşe wanted Ali to buy a storybook, the expected answer for the second question was "Kitabı aldım" rather than "Kitap aldım". Again more explicitly, the reason behind the answer "Kitabı aldım" for the second question is as follows: Ali knows that Ayşe wants to know whether Ali bought the storybook that she wanted him to buy or not.

Procedure

The story was told to the participants and two closed-ended questions were asked to the participants. If the participant gave the expected answer to the two questions, s/he received a score of 2 points in the analysis. Sometimes the participants tended to change their first answer after they heard the second question. In these circumstances, their second answer was taken into consideration.

4.2.4. Second-order Relative Clause Task (REL_2)

Material

The REL_2 is related to the comprehension of relative clauses (RC) in Turkish. This task was adapted from DuyguÖzge's (2010) PhD thesis with the author's permission. In the original task, there were 32 experimental and 28 control single-embedded RCs along with their related drawings. The questions and the drawings were modified to double-embedded ones to be able to analyze the participants' second-order embedding abilities, on a par with their second-order ToM abilities. Due to time restrictions, 1 practice trial and 6 experimental items were used.

Figure 2 and Figure 3 demonstrate the drawings for one of the questions related to the REL_2. The other items used in the experiment can be found in Appendix C. The positions of the correct answers were equally distributed across the drawings (3 times in the first row and 3 times in the second row) and between right (2 times), left (2 times) and central position (2 times).

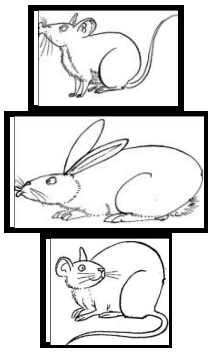


Figure 2: Example picture for the introductory figures

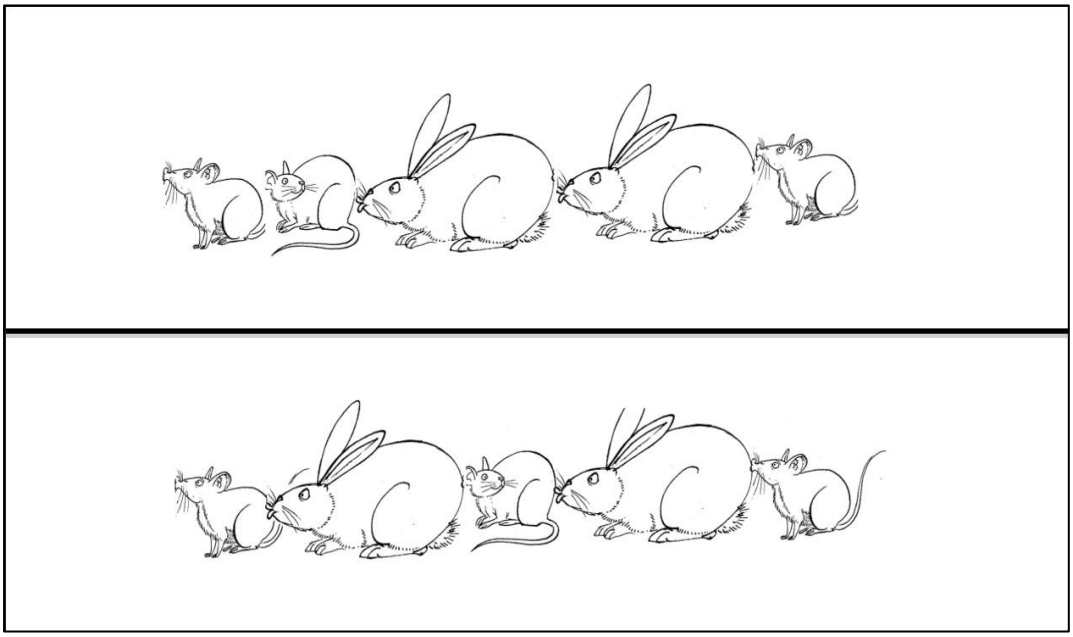


Figure 3: Picture of the question "Hangi resimde fareyi öpen tavşanı öpen bir fare var?" (In which picture there is a mouse kissing the rabbit that is kissing the mouse?)

Procedure

First, the introductory pictures (Figure 2) were shown to the participants in order to familiarize them with the animals in the action by telling the name of the animals and the actions (e.g., “this is a kissing rabbit”). After that, the pictures representing the questions (Figure 3) were shown one by one. The first and second rows of the picture were pointed out in order to make it clear that there are two separate lines of pictures by saying, “This is the first picture and this is the second picture”. In the trial session, it was explained that the participants were required to point out the row with the animals related to their answer. If they could not answer correctly, the correct animals were pointed out by the author with necessary explanations. If they could not answer the questions during the experiment, the sentences were repeated up to 4 times. If the participants correctly answered all of the questions, they scored 6.

4.2.5. Listening Span Task (LST)

Material

To be able to measure complex working memory, Ünal's (2008) English-to-Turkish adaptation of the original LST (Archibald &Alloway, 2008, as cited in Ünal, 2008) was used with the author's permission. The task consists of sets of sentences read out to the participants one by one. There are a total of five sets which consist of two to six sentences. At the level of each set size, there were also 6 sets of sentences. An example of a 3-sentence set of LST is as follows (see Appendix D for the entire material):

1. Muzlar bisiklete biner. (Bananas ride bicycles)
2. Elimiz beş parmaklıdır. (Our hands have five fingers)
3. Soğan acıdır. (Onions are hot)

Procedure

In the experiment, the sentences were told to the participants. They were expected to first judge the truthfulness of the sentences by saying "Yes" or "No". Secondly, they had to recall the last word of all the sentences told to them in the reverse order. After they gave an answer to the first sentence, the next sentence was told to them. For example, for the 2-sentence set if the first sentence is "Muzlar bisiklete biner." (Bananas ride bicycles), the participants were required to say "Hayır[1];biner". After that, if the second sentence is "Soğan acıdır.", they were required to say "Evet[2];acıdır, biner.". If the participant made less than two mistakes in a sentence set, the subsequent sentence set, which comprised one more sentence, was told to the participant. The score of the participants equaled to the number of sentence sets in which they did not make more than one mistake.

[1] 'Hayır' means 'No'
[2] 'Evet' means 'Yes'

CHAPTER 5

RESULTS

A total of 68 (35 female, 33 male) children participated in the experiments. Children's grades varied from kindergarten to fifth grade, and their age range was between 4 to 12 years (Table 2). The statistical analyses of children's responses to the five tasks are presented in this chapter. Later, the results of the adult control group will be presented. The p values are two-tailed, unless stated otherwise in which case the p-values are one-tailed.

5.1 The FBT_2

For the FBT_2, the number of subjects, the mean values and standard deviations are shown in Table 3. The maximum score for each FBT_2 is 1. Therefore the maximum score for the two stories is 2. The total FBT_2 score was taken into consideration in the rest of the analysis. Figure 4 shows the mean values of the FBT_2 score according to the grades.

Table 3: Descriptive statistics for the FBT_2

FBT_2	N	Mean	Std. Deviation
FBT_Chocolate	68	0.59	0.49
FBT_Puppy	68	0.63	0.48
Total FBT_2	68	1.22	0.91

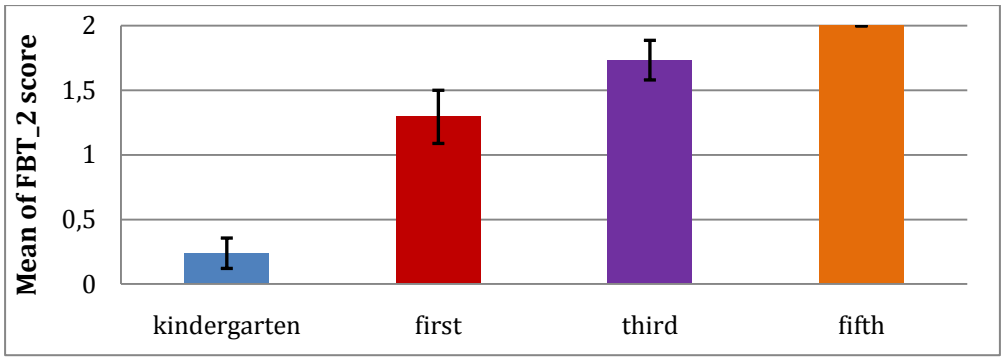


Figure 4: Mean values for the FBT_2 scores (Error bars represent SEs)

In order to analyze the developmental trend in understanding the FBT_2, the data was divided into four groups according to the participants' grades (kindergarten, 1st, 3rd, 5th grade). Table 4 shows the numbers, the mean ranks and the medians of the subjects according to their grades. The non-parametric Kruskal-Wallis Test showed that there is a significant difference in performance between the grades ($\chi^2(2) = 40.22$, $p = .000$). To be able to see the grades creating the differences, Mann Whitney Tests were used in order to compare the four age groups with each other. Since six Mann-Whitney Tests were used to test the difference across the grades, the alpha level for the Bonferroni correction was set to .008. This figure was calculated by dividing the original alpha level of .05 by the number of tests (6) conducted (.05/6=.008). According to the results, while there is a steady increase in performance, there is no significant difference between the first and third grades and between the third and fifth grades. However, there is a significant difference between kindergarten and grade one (Z= -3.73, p= .000), kindergarten and grade three (Z= -4.73, p = .000), kindergarten and grade five (Z= -5.36, p = .000), and grade one and five (Z= -2.99, p = .003).

Table 4: Number, mean rank and median of subjects for the FBT_2

Grades	N	Mean Rank	Median
Kindergarten	21	15.55	0
First	17	35.53	2
Third	15	44.37	2
Fifth	15	50.00	2
Total	68		

5.2 Conditions of the FBT_2

Table 5 shows the number of subjects, the mean values and standard deviations for the conditions of total FBT_2 score, namely –DI, -mIş and neutral and Table 6 shows the mean and standard deviation of the conditions of total FBT_2 score across the grades.

Table 5: Descriptive statistics for the conditions of the FBT_2 score

FBT_2 conditions	N	Mean	Std. Deviation
-DI	22	1.36	0.85
-mIş	23	1.22	0.90
Neutral	23	1.09	0.99

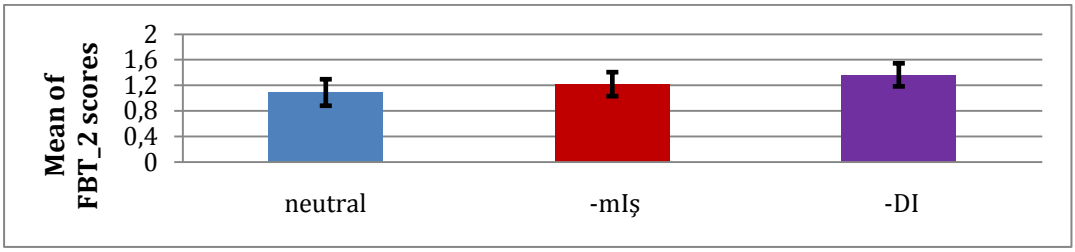


Figure 5: Mean values of conditions of the FBT_2 scores (Error bars represent SEs)

Table 6: Descriptive statistics for the conditions of the FBT_2 score across grades

Grade	Condition	Mean	Std. Deviation	N
kindergarten	neutral	0.00	0.00	7
	-mIş	0.14	0.38	7
	-DI	0.57	0.79	7
	Total	0.24	0.54	21
first	neutral	1.17	0.98	6
	-mIş	1.33	0.82	6
	-DI	1.40	0.89	5
	Total	1.29	0.85	17
third	neutral	1.60	0.89	5
	-mIş	1.80	0.45	5
	-DI	1.80	0.45	5
	Total	1.73	0.59	15
fifth	neutral	2.00	0.00	5
	-mIş	2.00	0.00	5
	-DI	2.00	0.00	5
	Total	2.00	0.00	15
Total	neutral	1.09	1.00	23
	-mIş	1.22	0.90	23
	-DI	1.36	0.85	22
	Total	1.22	0.91	68

Since the data were not normally distributed (the results of Shapiro-Wilk Test were p <.05), the Kruskal-Wallis Test was used as a non-parametric test. Generally, gender does not affect the understanding of the false belief task. In this test the effect of gender was investigated using the Kruskal-Wallis Test. Similar to the previous studies (e.g. Özoran, 2009), the result of the analysis was not significant (at the .05 level).

Even though Figure 5 shows that overall children profited from the stories told with –DI more than from those told with –mIş and the neutral version, the non- parametric Kruskal-Wallis showed that there is no statistically significant difference between the conditions –DI, mIş and neutral ($\chi^2(2) = 0.83$, p= .66). Also as shown in Figure 6, children profited from the stories told with –DI more than from those told with –mIş and neutral version until the third grade. Third-grade children profited slightly from the stories told with –DI and –mIş equally, but still more than from those told in the neutral version. Finally, since all of the fifth graders passed the FBT_2, there is no difference across the versions. However, both the non-parametric Kruskal-Wallis Test and ANOVA showed that there is no statistically significant difference between the conditions for each grade.

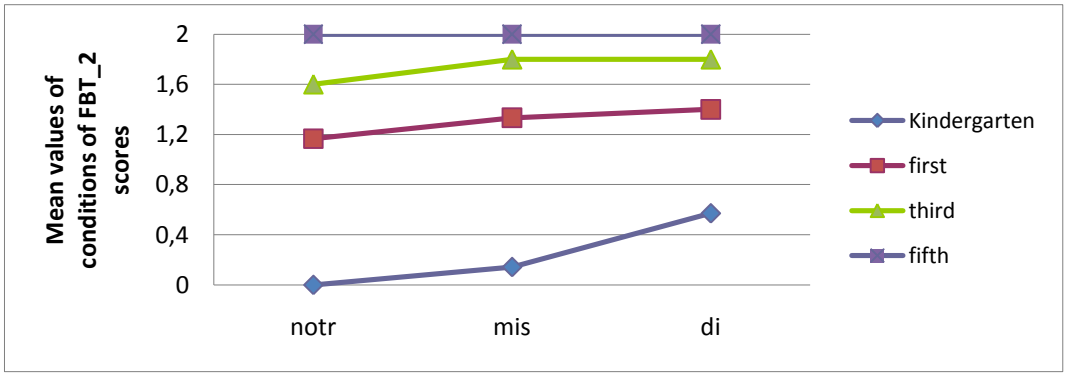


Figure 6: Mean values of conditions of the FBT_2 scores

Since it is not possible to look at the effects of two independent variables and their interaction with non-parametric tests, an independent factorial ANOVA with the factors (1) grades (kindergarten, one, three, five) and (2) FBT condition (-DI, -mIş, -neutral) was used in order to investigate the age and evidentiality interaction. Results showed that there is a significant main effect of grades (F3, 56= 30.09, p= .000,

ηp2= .617) on FBT_2 scores, however there is no significant main effect of conditions of the stories on FBT_2 scores and also as shown in Figure 7, there is no significant effect of the interaction between grades an conditions of the stories.

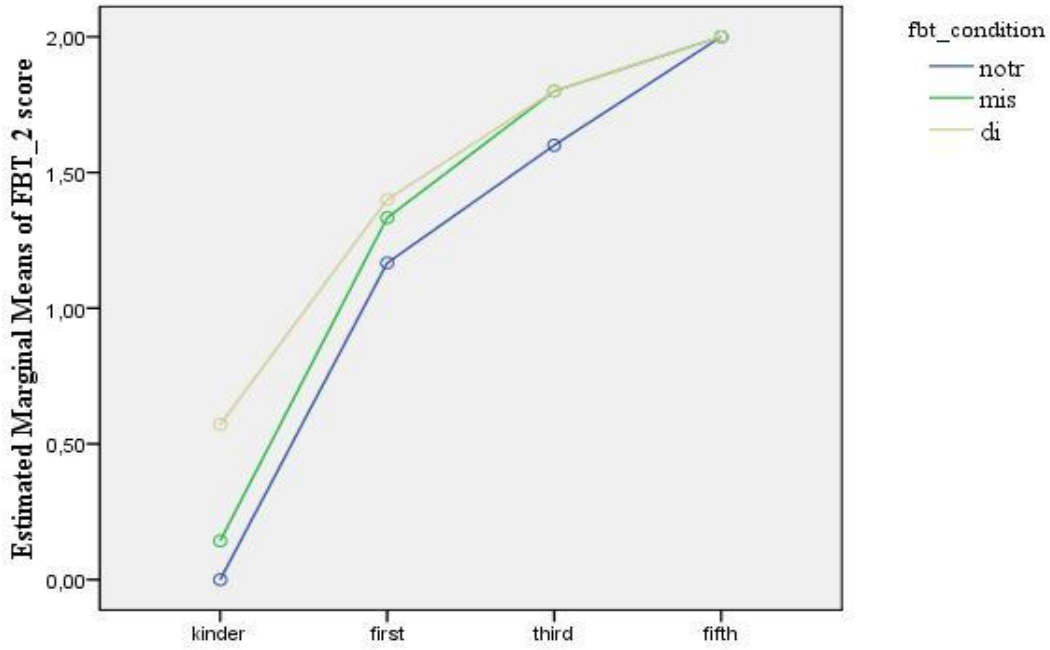


Figure 7: Interaction of the FBT_2 conditions and grades

5.3 WST

For the WST, the number of subjects, the mean value, standard deviation and median are shown in Table 7. The maximum score for WST is 8. Figure 8 shows the mean values of the WST score according to the grades.

Table 7: Descriptive statistics for the WST

	N	Mean	Std. Deviation	Median
WST	68	4.46	0.98	5

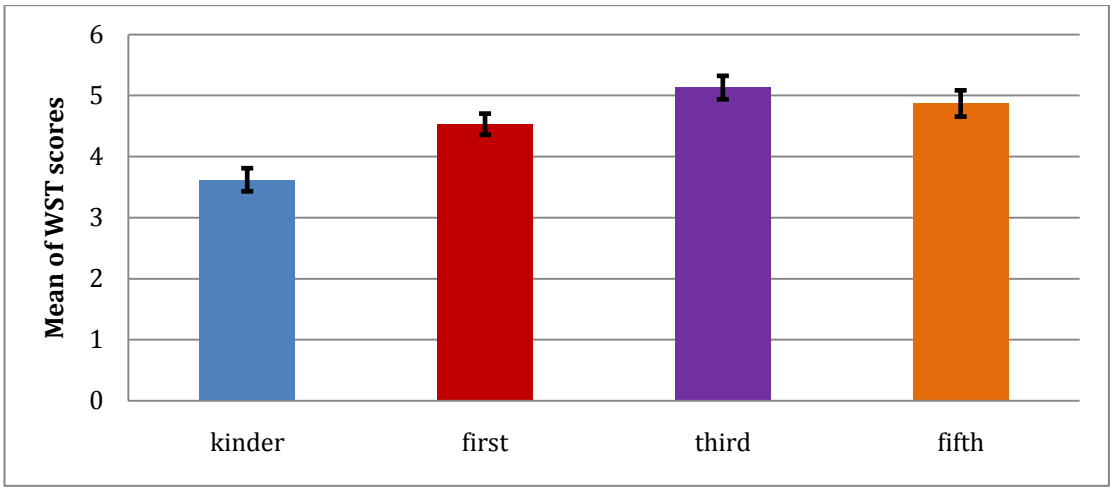


Figure 8: Mean values of conditions of the WST scores (Error bars represent SEs)

In order to analyze the developmental trend in the WST, the data was divided into four groups according to the participants' grades. Table 8 shows the number, the mean rank and the median of the subjects according to their grades. The non-parametric Kruskal-Wallis Test showed that there is a significant difference between the grades ($\chi^2(2) = 24.67$, $p = .000$). In order to see which grades differ significantly from each other, Mann Whitney Tests were used. As explained in section 5.1 above the alpha level for the Bonferroni correction was set to .008. According to the results, there is no difference between the first, third and fifth grades, while there is a significant difference between kindergarten and grade one (Z= -3.06, p= .002), kindergarten and grade three (Z= -4.14, p= .000), and kindergarten and grade five (Z= -3.59, p= .000).

Table 8: Number, mean rank and median of subjects for the WST

Grades	N	Mean Rank	Median
Kindergarten	21	18.76	4
First	17	35.32	5
Third	15	47.53	5
Fifth	15	42.57	5
Total	68		

5.4 PTT

For the PTT, the number of subjects, the mean value, standard deviation and median are shown in Table 9. Since there were two close-ended questions in the story, the maximum score for PTT is 2. Figure 9 shows the mean values of the PTT score according to the grades.

Table 9: Descriptive statistics for the PTT

	N	Mean	Std. Deviation	Median
PTT	68	1.28	0.59	1

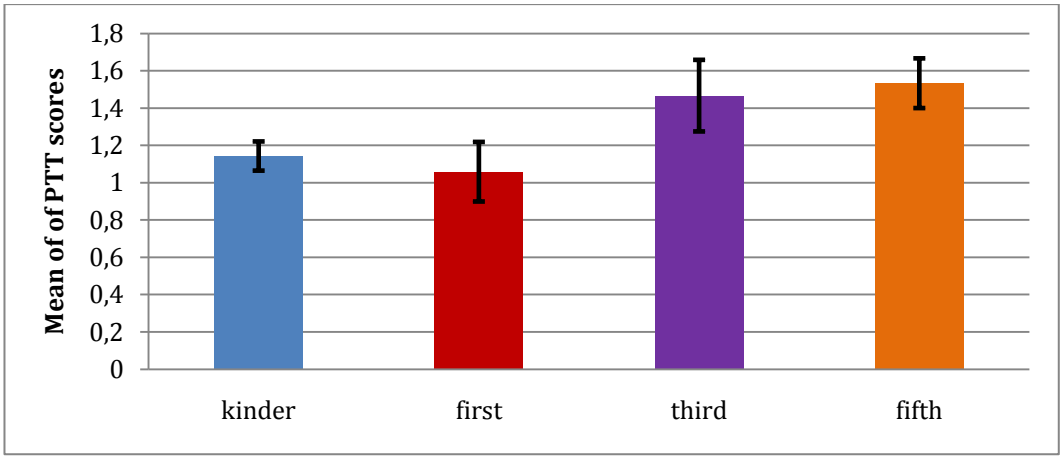


Figure 9: Mean values of conditions of the PTT scores (Error bars represent SEs)

In order to analyze the developmental trend in PTT, the data was divided into four groups according to the participants' grades. Table 10 shows the number, the mean rank and the median of the subjects according to their grades. Kindergarten children and first graders have scores around 1 which is the score expected by chance. The non-parametric Kruskal-Wallis Test showed that there is a significant difference between the grades ($\chi^2(2) = 8.53$, $p = .036$). In order to see which grades differ significantly from each other, Mann Whitney Tests were used. According to the results, there is no difference between the kindergarten and grade one, grade one and three, grade one and three, grade three and five, while there is a significant difference between the kindergarten and grade five (Z= -2.473, p= .006, one-tailed).

Table 10: Number, mean rank and median of subjects for the PTT

Grades	N	Mean Rank	Median
Kindergarten	21	29.50	1
First	17	28.53	1
Third	15	40.97	2
Fifth	15	41.80	2
Total	68		

5.5 REL_2

For the REL_2, the number of subjects, the mean value, standard deviation and median are shown in Table 11. Since there were 6 different questions in the task, the maximum score of total REL_2 is 6. The total score of REL_2 was taken into consideration for the rest of the analysis. Figure 10 shows the mean values of the REL_2 score according to the grades.

Table 11: Descriptive statistics for the REL_2

	N	Mean	Std. Deviation	Median
Total REL_2	68	2.28	2.02	2

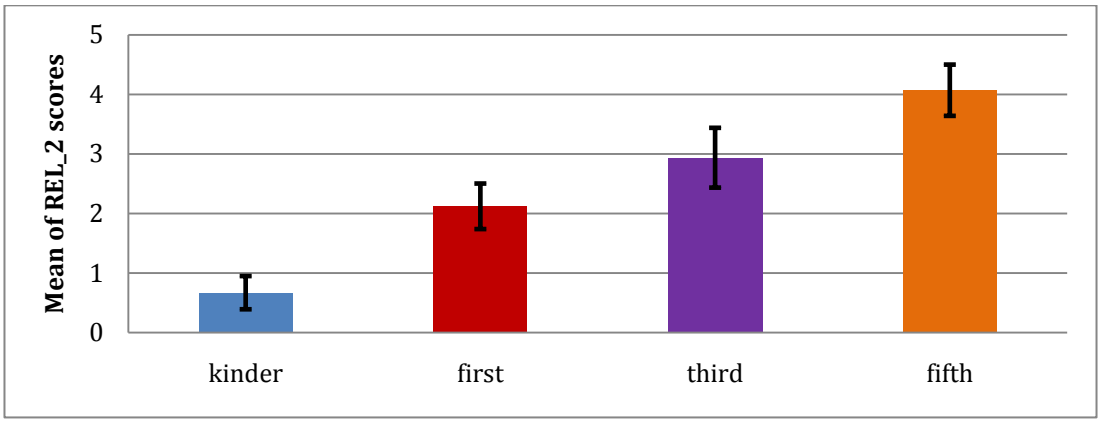


Figure 10: Mean values of conditions of the REL_2 scores (Error bars represent SEs)

In order to analyze the developmental trend in understanding REL_2, the data was divided into four groups according to the participants' grades. Table 12 shows the number, the mean rank and median of the subjects according to their grades. The non-parametric Kruskal-Wallis Test showed that there is a significant difference

between the grades ($\chi^2(2) = 27.37$, $p = .000$). In order to see which grades differ significantly from each other, Mann Whitney Tests were used. According to the results, there is no difference between the first and third grade and between the third and fifth grade, while there is a significant difference between kindergarten and grade one (Z= -2.94, p= .003), kindergarten and grade three (Z= -3.58, p= .000), kindergarten and grade five (Z= -4.65, p= .000), and grade one and grade five (Z= -2.90, p= .004).

Table 12: Number, mean rank and median of subjects for the REL_2

Grades	N	Mean Rank	Median
Kindergarten	21	18.62	0
First	17	33.68	2
Third	15	40.77	3
Fifth	15	51.40	4
Total	68		

5.6 LST

For the LST, the number of subjects, the mean value, standard deviation and median are shown in Table 13. The maximum score of LST is 6. Figure 11 shows the mean values of the LST score according to the grades.

Table 13: Descriptive statistics for the LST

	N	Mean	Std. Deviation	Median
LST	68	1.24	1.31	1

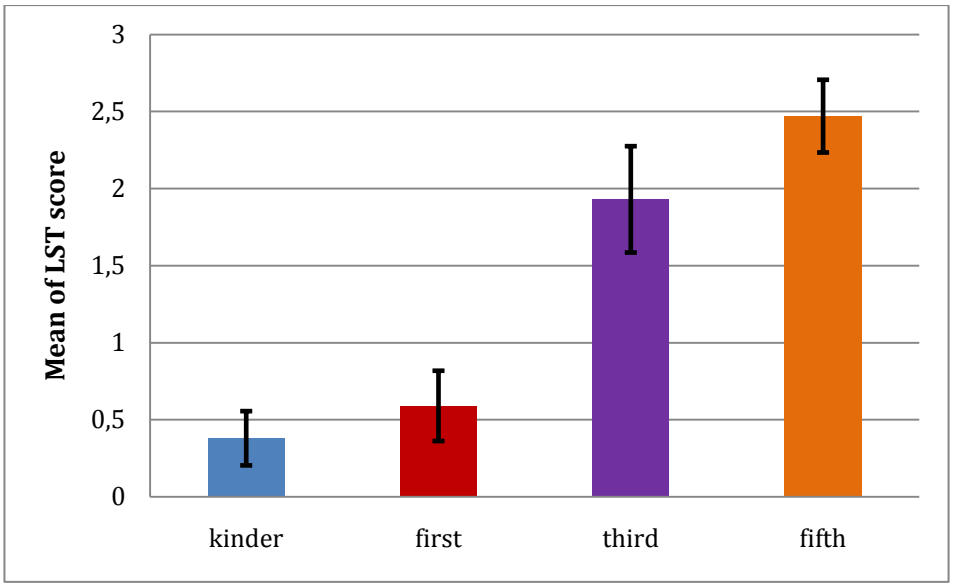


Figure 11: Mean values of conditions of total LST scores (Error bars represent SEs)

In order to analyze the developmental trend in LST performance, the data was divided into four groups according to the participants' grades. Table 14 shows the number, the mean rank and median of the subjects according to their grades. The non-parametric Kruskal-Wallis Test showed that there is a highly significant difference between the grades ($\chi^2(2) = 30.87$, $p = .000$). In order to see which grades differ significantly from each other, Mann Whitney Tests were used. Again, the alpha level for the Bonferroni correction was set to .008. According to the results, there is no difference between the kindergarten and first grade and third and fifth grades, while there is a significant difference between the kindergarten and grade three (Z=-3.53, p= .000), kindergarten and grade five (Z= -4.64, p = .000), grades one and three (Z= -2.92, p = .003), and grades one and five (Z=-4.08, p = .000).

Table 14: Number, mean rank and median of subjects for the LST

Grades	N	Mean Rank	Median
Kindergarten	21	22.64	0
First	17	25.44	0
Third	15	44.17	2
Fifth	15	51.70	3
Total	68		

5.7 Sentence Comprehension Predicting Second-order False Belief

Since the data violates normality, the non-parametric Spearman's Rank Order Correlation was used to test the relationship between total FBT_2 and PTT scores. This analysis showed that there is no significant relationship between total FBT_2 and PTT (r_s = .19, *p*= .126).

Partial correlation was also used in order to control the other variables, when investigating the relationship between FBT_2 and PTT scores. Table 15 shows the control variables, correlation coefficients and p values of partial correlation results for FBT_2 and PTT.

Table 15: Control variables, correlation coefficients and p values of partial correlation results for FBT_2 and PTT

Control Variable	Partial Correlation	p
Age	- .095	.444
WST	.12	.922
REL_2	.036	.772
LST	- .22	.860

The table reads as follows: when age is controlled for, the previous correlation of r_s= .19 between PTT and FBT_2 drops to -.095; when WST is controlled for, the correlation drops to .12, and so on for the other variables.

5.8 Double-embedded Relative Clauses Predicting Second-order False Belief

Since the data violates normality, the non-parametric Spearman's Rank Order Correlation was used to test the relationship between total FBT_2 and REL_2. This analysis showed that there is a significant relationship between total FBT_2 and REL_2 scores (r_s = .54, *p*= .000).

Bivariate regression was also used in order to predict the model of REL_2 score predicting FBT_2 score. Using the enter method, the FBT_2 score could be predicted from REL_2 score by the following formula: 0,24 X REL_2 + 0.673 ($F_{66,1}$= 26.196, p= .000, r= .533, R^2= .284)

Partial correlation was also used in order to control the other variables, when investigating the relationship between FBT_2 and REL_2 scores. Table 16 shows the control variables, correlation coefficients and p values of partial correlation results for FBT_2 and REL_2.

Table 16: Control variables, correlation coefficients and p values of partial correlation results for FBT_2 and REL_2

Control Variable	Partial Correlation	p
Age	.10	.421
WST	.39	.001**
PTT	.52	.000**
LST	.25	.041*

In the light of the partial analyses, multiple regression was used by using age and REL_2 scores as independent variables and FBT_2 as dependent variable. Using the enter method, the FBT_2 score could be predicted from age and REL_2 score by the following formula: FBT_2= 0.039 X REL_2+ 0.25 X age – 0.751 ($F_{65, 2}$= 42.091, p= .000, r= .751, R^2= .564). However, only the contribution of age is significant (β= .692, t= 6.47, p = .000).

5.9 Multiple Regression for FBT_2

Two models were constructed by using multiple regression to predict FBT_2 scorefirst just with the contribution of age and second with age and all tasks. Table 17 shows the correlations of all tasks for FBT_2. Using the enter method, the FBT_2 score could be predictedby age by the following formula: 0.27 X age – 0.814 ($F_{66, 1}$= 83.965, p= .000) and could be predicted by age and all tasks by the following formula: 0.236 X age + 0.145 X WST + 0.045 REL_2 – 0.034 X LST – 0.130 X PTT – 1.098 ($F_{62, 5}$= 17.519, p= .000, r= .765, R^2= .586). However, only the contribution of age is significant (β= .655, t= 5.45, p = .000).

Table 17: Correlations of all tasks and age for FBT_2

Variable	Correlation	p
Age	.748	.000**
WST	.518	.000**
PTT	.160	.096
REL_2	.533	.000**
LST	.503	.000**

The below collinearity table represents the data on an age and the other tasks in terms of their linear relatedness. In the table, age (94%) and WST (90%) load highly on a different single dimension. This means that both age and the WST can explain only one independent measure separately. On the other hand, the PTT, the REL_2 and the LST share some proportions with the other tasks. Still they mainly load on their own distinctive dimension. This is because they are also related to different abilities. Moreover, the LST (60%) and the REL_2 (75%) load highest on the same dimension which shows that both tasks tap into the same cognitive ability.

Table 18: Collinearity Dignostics

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	age	LST	PTT	REL_2	WST
1	1	1,949	1,000	,03	,03				
	2	,051	6,172	,97	,97				
2	1	5,216	1,000	,00	,00	,00	,00	,00	,00
	2	,516	3,179	,01	,00	,13	,02	,07	,00
	3	,133	6,264	,02	,02	,14	,66	,14	,01
	4	,082	7,988	,01	,03	,60	,30	,75	,01
	5	,034	12,336	,20	,94	,11	,02	,02	,07
	6	,020	16,346	,77	,01	,02	,00	,01	,90

a. Dependent Variable: fbt-total

5.10 Serial Processing Bottleneck

Since our aim was to test the hypothesis that developmental latencies between first- and second-order social reasoning was due to the children's need to overcome serial processing rather than simple working memory capacity, the relationship between

LST and FBT_2, WST and FBT_2, LST and REL_2, and WST and REL_2 was investigated by using non-parametric parametric Spearman's Rank Order Correlations. The number of subjects, the **ρ** and *p* values of Spearman's Rank Order Correlation are shown in Table 19.

Table 19: Spearman's Rank Order Correlations

	N	r_s	*p*
LST & FBT_2	68	.496	.000**
WST & FBT_2	68	.500	.000**
LST & REL_2	68	.804	.000**
WST & REL_2	68	.467	.000**

5.10.1 LST and FBT_2

Bivariate regression was also used in order to predict the model of LST score predicting FBT_2 score. Using the enter method, the FBT_2 score could be predicted from LST score by the following formula: 0,348 X LST + 0.790 ($F_{66,1}$= 22.356, p= .000, r= .503, R^2= .253)

Partial correlation was also used in order to control the other variables, when investigating the relationship between FBT_2 and LST scores. Table 20 shows the control variables, correlation coefficients and p values of partial correlation results for FBT_2 and LST.

Table 20: Control variables, correlation coefficients and p values of partial correlations results for FBT_2 and LST

Control Variable	Partial Correlation	p
Age	.02	.862
WST	.35	.004**
PTT	.48	.000**
REL_2	.15	.234

In the light of the partial analyses, multiple regression was used by using age and LST scores as independent variables and FBT_2 as dependent variable. Using the enter method, the FBT_2 score could be predicted from age and LST score by the

following formula: FBT_2= 0.13 X LST+ 0.266 X age – 0.796 ($F_{65, 2}$= 41.381, p= .000, r= .748, R^2= .560). However, only the contribution of age is significant (β= .736, t= 6.736, p = .000).

5.10.2 WST and FBT_2

Bivariate regression was also used in order to predict the model of WST score predicting FBT_2 score. Using the enter method, the FBT_2 score could be predicted from WST score by the following formula: 0.47 X WST – 0.920 ($F_{66,1}$= 24.263, p= .000, r= .518, R^2= .269)

Partial correlation was also used in order to control the other variables, when investigating the relationship between FBT_2 and WST scores. Table 21 shows the control variables, correlation coefficients and p values of partial correlation results for FBT_2 and WST. For example, the partial correlation is equal to .19, when we controlled the age variable when looking at the correlation between the FBT_2 and the WST.

Table 21: Control variables, correlation coefficients and p values of partial correlation results for the FBT_2 and theWST

Control Variable	Partial Correlation	p
Age	.19	.131
LST	.38	.002**
PTT	.50	.000**
REL_2	.36	.002**

In the light of the partial analyses, multiple regression was used by using age and WST scores as independent variables and FBT_2 as dependent variable. Using the enter method, the FBT_2 score could be predicted from age and WST score by the following formula: FBT_2= 0.138 X WST+ 0.24 X age – 1.204 ($F_{65, 2}$= 44.004, p= .000, r= .758, R^2= .575). However, only the contribution of age is significant (β= .666, t= 6.87, p = .000).

5.10.3 LST and REL_2

Bivariate regression was also used in order to predict the model of LST score predicting REL_2 score. Using the enter method, the REL_2 score could be predicted by LST score by the following formula: 1.235 X LST – 0.754 ($F_{66,1}$= 121.268, p= .000, r= .805, R^2= .648)

Partial correlation was also used in order to control the other variables, when investigating the relationship between REL_2 and LST scores. Table 22 shows the control variables, correlation coefficients and p values of partial correlation results for REL_2 and LST.

Table 22: Control variables, correlation coefficients and p values of partial correlation results for REL_2 and LST

Control Variable	Partial Correlation	p
Age	.66	.000**
WST	.75	.000**
PTT	.79	.000**
FBT_2	.73	.000**

In the light of the partial analyses, multiple regression was used by using age and LST scores as independent variables and REL_2 as dependent variable. Using the enter method, the REL_2 score could be predicted by age and LST score by the following formula: REL_2= 1.03 X LST+ 0.163 X age – 0.218 ($F_{65,\ 2}$= 66.286, p= .000, r= .819, R^2= .671) and both the contributions of LST (β= .671, t= 7.10, p = .000) and age (β= .203, t= 2.15, p = .035) are significant.

5.10.4 WST and REL_2

Bivariate regression was also used in order to predict the REL_2 score by WST score predicting REL_2 score. Using the enter method, the REL_2 score could be predicted by WST score by the following formula: 0.946 X WST – 1.934 ($F_{66,1}$= 17.748, p= .000, r= .460, R^2= .212)

Partial correlation was also used in order to control the other variables, when investigating the relationship between REL_2 and WST scores. Table 23 shows the control variables, correlation coefficients and p values of partial correlation results for REL_2 and WST.

Table 23: Control variables, correlation coefficients and p values of partial correlation results for the REL_2 and the WST

Control Variable	Partial Correlation	p
Age	.16	.193
LST	.18	.157
PTT	.42	.000**
FBT_2	.25	.038*

In the light of the partial analyses, multiple regression was used by using age and WST scores as independent variables and REL_2 as dependent variable. Using the enter method, the REL_2 score could be predicted by age and WST score by the following formula: REL_2= 0.304 X WST+ 0.45 X age – 2.465 ($F_{65, 2}$= 24.585, p= .000, r= .656, R^2= .431). However, only the contribution of age is significant (β= .562, t= 4.997, p = .000).

Finally, multiple regression was used by using age, WST and LST scores as independent variables and REL_2 as dependent variable. Using the enter method, the REL_2 score could be predicted by age, WST and LST by the following formula: REL_2 = 1.015 X LST+ 0.123 X WST + 0.141 X age – 0.568 ($F_{64, 3}$= 43.995, p= .000, r= .821, R^2= .673). However, only the contribution of LST is significant (β= .661, t= 6.898, p = .000).

5.11 Results for the Adult Control Group

In this subsection, adults' performance in all tasks used in the study will be presented and compared with children's performance. The comparison was only made with the fifth-graders, as they were the oldest children's group.

5.11.1 FBT_2

The Kruskal-Wallis Test was used in order to analyze the difference between the adults and the fifth graders in FBT_2 performance. Table 24 shows the descriptive statistics and Table 25 shows the number, the mean rank, median of the subjects. The test showed no significant difference between the adults' and children's FBT_2 performance ($\chi^2(2) = 0.00$, $p = 1.00$). Since all of the adults and all of the fifth grade children answered all of the FBT tasks correctly, the differences between the evidentiality conditions are not significant.

Table 24: Descriptive statistics for the FBT_2

	N	Minimum	Maximum	Mean	Std. Deviation
Adults	10	2	2	2	0
Fifth Grades	15	2	2	2	0

Table 25: Number and mean ranks of subjects for the FBT_2

	N	Mean Rank	Median
Adults	10	13.00	2
Fifth Grades	15	13.00	2

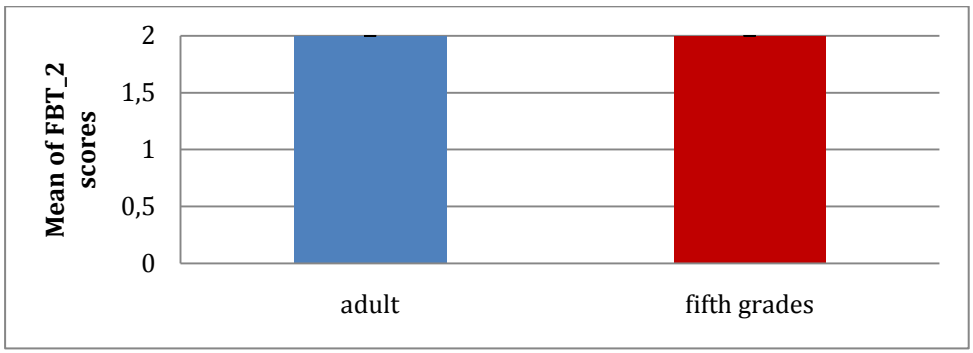


Figure 12: Mean values of the FBT_2 scores

5.11.2 WST

The Kruskal-Wallis Test was used in order to analyze the difference between the adults and the fifth grades in WST performance. Table 26 shows the descriptive statistics and Table 27 shows the number, the mean rank, median of the subjects. The

test showed that there is a significant difference between the adults' and children's WST performance ($\chi^2(2) = 8.925$, $p = .003$).

Table 26: Descriptive statistics for the WST

	N	Minimum	Maximum	Mean	Std. Deviation
Adults	10	4	7	5.90	0.738
Fifth Grades	15	3	6	4.87	0.834

Table 27: Number and mean ranks of subjects for WST

	N	Mean Rank	Median
Adults	10	18.05	6
Fifth Grades	15	9.63	5

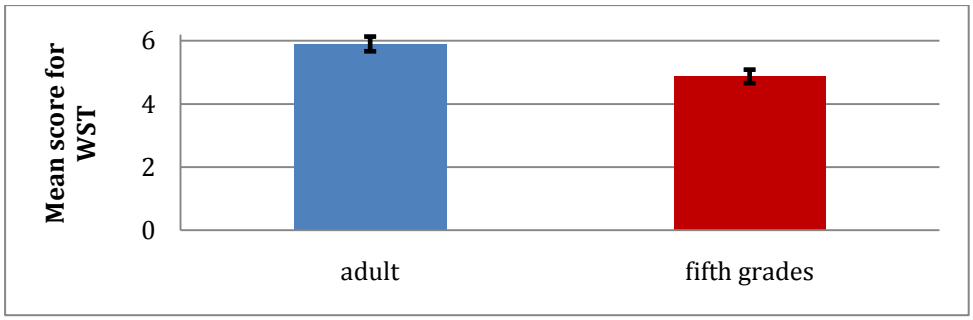


Figure 13: Mean values of WST scores (Error bars represent SEs)

5.11.3 PTT

The Kruskal-Wallis Test was used in order to analyze the difference between the adults and the fifth grades in PTT performance. Table 28 shows the descriptive statistics and Table 29 shows the number, the mean rank, median of the subjects. The test showed that there is no significant difference between the adults' and children's PTT performance ($\chi^2(2) = 1.778$, $p = .182$).

Table 28: Descriptive statistics for the PTT

	N	Minimum	Maximum	Mean	Std. Deviation
Adults	10	1	2	1.80	0.422
Fifth Grades	15	1	2	1.53	0.516

Table 29: Number and mean ranks of subjects for PTT

	N	Mean Rank	Median
Adults	10	15.00	2
Fifth Grades	15	11.67	2

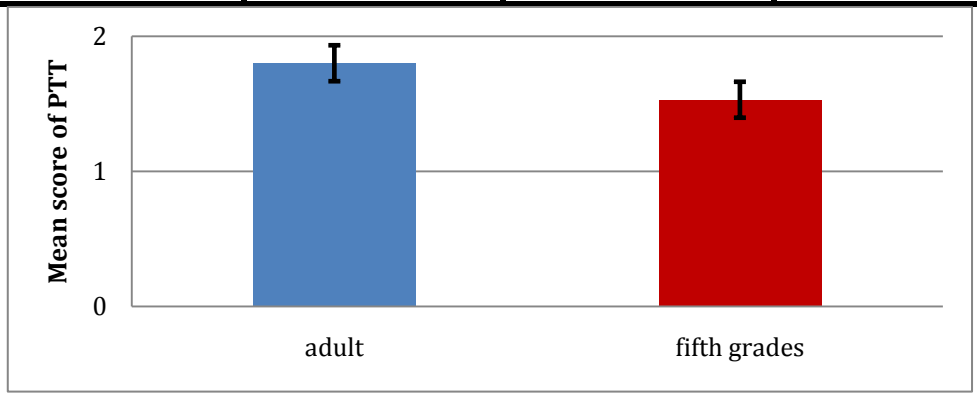


Figure 14: Mean values of PTT scores (Error bars represent SEs)

5.11.4 REL_2

The Kruskal-Wallis Test was used in order to analyze the difference between the adults and the children in REL_2 performance. Table 30 shows the descriptive statistics and Table 31 shows the number, the mean rank, median of the subjects. The test showed that there is a significant difference between the adults' and children's REL_2 performance ($\chi^2(2) = 6.096$, $p = .014$).

Table 30: Descriptive statistics for the REL_2

	N	Minimum	Maximum	Mean	Std. Deviation
Adults	10	4	6	5.60	0.843
Fifth Grades	15	1	6	4.07	1.668

Table 31: Number and mean ranks of subjects for the REL_2

	N	Mean Rank	Median
Adults	10	17.10	6
Fifth Grades	15	10.27	4

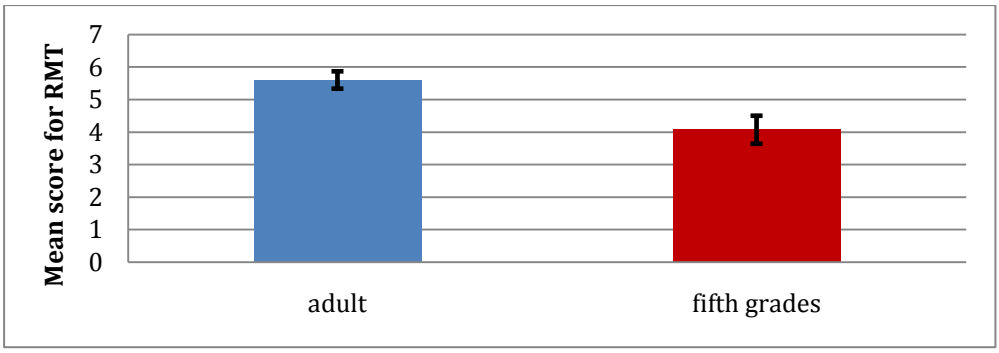


Figure 15: Mean values of the REL_2 scores (Error bars represent SEs)

5.11.5 LST

The Kruskal-Wallis Test was used in order to analyze the difference between the adults and the fifth grades in LST performance. Table 32 shows the descriptive statistics and Table 33 shows the number, the mean rank, median of the subjects. The test showed that there is a significant difference between the adults' and children's LST performance ($\chi^2(2) = 4.729$, $p = .030$).

Table 32: Descriptive statistics for the LST

	N	Minimum	Maximum	Mean	Std. Deviation
Adults	10	2	4	3.30	0.823
Fifth Grades	15	0	4	2.47	0.915

Table 33: Number and mean ranks of subjects for the LST

	N	Mean Rank	Median
Adults	10	16.70	3.5
Fifth Grades	15	10.53	3.0

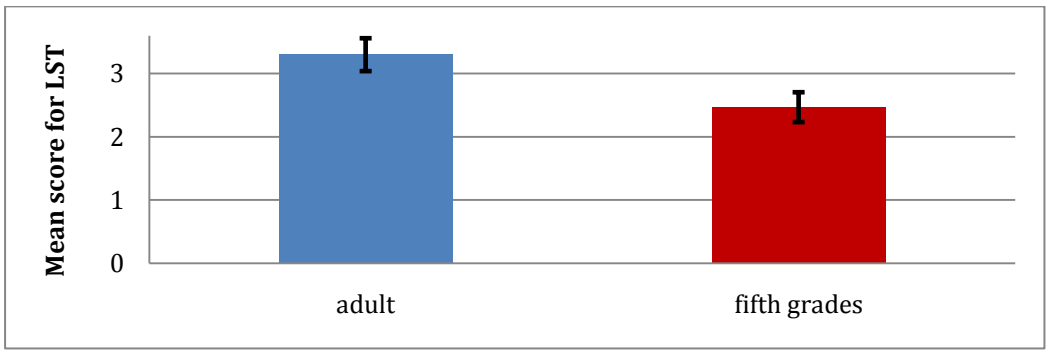


Figure 16: Mean values of the LST scores (Error bars represent SEs)

Summarizing, the results of the adult sample revealed no difference in FB_2 understanding between the oldest age group (grade 5) and adults. However, adults outperformed the fifth graders in all of the other tasks except the PTT. We can conclude from these results that second-order false belief understanding, as tested by our two stories, is fully achieved at around age 10-11, whereas simple and complex working memory and double-embedded relative clause understandings still develops after the age of 10.

CHAPTER 6

GENERAL DISCUSSION

In this section, first of all the developmental trend in the performance of kindergarten children, first, third and fifth graders in all studied tasks is discussed: in the second-order false belief task (including the effect of the acquisition of Turkish evidentials), word span task (WST), perspective taking test (PTT), double-embedded relative clause task (REL_2) and listening span task (LST), respectively. After presenting the developmental trends for all tasks, the predictions of the second-order false belief task (FBT_2) from the remaining tasks are discussed with respect to the literature. Finally, the serial bottleneck hypothesis is examined in terms of the relation between the second-order false belief task and working memory tasks (WST and LST) and also the relation between the REL_2 and working memory tasks.

6.1 Development of second-order false belief reasoning

As can be seen clearly from Figure 4, a linear developmental trend was found for the FBT_2 score for the grade one (6- 7 years) to grade five (10- 12 years). However, there is a jump between kindergarten children (4- 5 years) and first graders ($Z= -3.73$, $p= .000$). While there is a significant difference between grade one and grade five ($Z= -2.99$, $p = .003$), there is no significant difference between grade one and grade three (9, 9;5). All of the fifth graders and adults answered both second-order false belief questions correctly. More explicitly, we can say that second-order false belief reasoning starts to develops around the age of 6, and reaches the adult-like understanding at around the age of 9;5. These findings are compatible with Perner

and Wimmer's (1985) study, which states that second-order false belief understanding occurs after the age of 6 and with our first hypothesis, which states the expected developmental trend. Although kindergarten children failed in FBT_2 on average, there were three of them who succeeded in the Birthday Puppy Story and one of them succeeded in both the Birthday Puppy Story and Chocolate Bar Story. These cases deserve special attention. Possible explanations for this early second-order false belief understanding are given in the last part of this section in terms of serial processing bottleneck.

Flobbe et al.'s (2008) results showed that children's performance in the Chocolate Bar Story was better than in the Birthday Puppy Story. However, in our study the performance of the children between the ages 8 to 10 was better in the Birthday Puppy Story (14 children) than in the Chocolate Bar Story (12 children). Moreover, the Birthday Puppy Story is the one story that the four kindergarten children succeeded in. Since the previous studies of strategic games showed that more concrete presentation of the games increased the children's performance (cf. Flobbe et al., 2008; Meijering et al., 2010), a possible explanation for this diverse finding might be the fact that we added a 'grandmother' character to Flobbe's drawing in order to make the story more explicit.

6.2 The effect of the acquisition of evidentiality on the development of second-order false belief reasoning

We did not find any difference between the three evidential conditions of the two FB_2 tasks: -DI, -mIş, and neutral condition. Our findings about the effect of the acquisition of evidentiality on the development of false belief reasoning are therefore against our second and third hypotheses and against the hypotheses in the literature that there should be such an effect (cf. Aksu-Koç, 1988). However, the result is compatible with Papafragou's (2007) source monitoring study where she did not find any effect of evidentials. Özoran (2009) studied the effect of evidentials on first-order theory of mind with 4- to 7- year-old children and found that children's performance of the ToM tasks was significantly better in the condition –DI than the conditions neutral and –mIş. Our results for 4- to 5- year-old kindergarten children

show the same pattern with Özoran's (2009) study. The mean value of the FBT_2 score for the condition –DI (Mean= 0.57; SD = .079) is higher than the condition –mIş (Mean= 0.14; SD = 0.38) and the condition neutral (Mean= 0.00; SD= 0.00). Figure 6 displays the mean scores of the FBT_2 in terms of the three evidential markers. However, since the FBT_2 scores of the kindergartens were too low, the effect of the evidentials on FBT_2 score is not statistically significant.

How can we interpret this similarity in pattern with Özoran's (2009) finding for young children but not older ones? A possible interpretation of these findings may become availableby considering Gözenman's (2011) very recent study. She conducted an experiment with preschool children aged 4 to 6. She told five different stories not involving theory of mind in 3 different conditions (-DI, -mIş, neutral). The same significant facilitatory effect of –DI over the neutral version and –mIş was found. In the light of these results, it appears that the influence of the evidentials is rather indirect via facilitation of understanding of stories/narratives. Since false belief tasks are presented in story form, facilitation of understanding these stories would automatically – but mistakenly – result in higher scores for false belief understanding. This facilitatory effect is only seen between the ages 4 to 6/7, which is in line with Aksu-Koç's (1988) study, which states that fully understanding of –DI and –mIş does not occur before the age of 6. After the acquisition of evidentials is sufficiently stable, they do not make a difference in understanding of stories/narratives. The findings of the present study are consistent with this explanation. Why evidentials fail to exert any influence on false belief understanding – although the information they provide is relevant – is yet to be discussed.

6.3 Development of the Word Span Task

The results show a significant and clear developmental trend from kindergarten to 3rd grade (Figure 8). We adapted this task from Ünal (2008). She also found a similar developmental trend from 1st to 5th graders. Özoran (2009) had also used the WST, however, he did not find any difference between the younger (3;6 to 5;6 years) and older (5;7 to 7;5 years) group. However, in our study 5th graders' WST score was somewhat lower than that of the 3rd graders, which is not compatible with Ünal's

(2008) study. We did not prepare any questionnaire to measure teacher ratings in order to test a child's social skills or IQ score. However, since the other complex working memory task (LST) also shows a developmental trend even for the 5th graders, we cannot relate this contrary finding simply to the participants' social skills or IQ. A possible but rather ad hoc explanation for this finding might be 5th graders' temporary lack of attention during the task. Adults' performance on WST is significantly higher than that of the 5th graders ($\chi^2(2) = 8.925$, $p = .003$).

6.4 Development of the Perspective Taking Test

When we were constructing this test, we were inspired by Flobbe et al.'s (2008) sentence comprehension test which was used to examine children's ability to reason about the speaker's linguistic alternatives in describing an event. In their test, referential reading of indefinite subjects in canonical sentences was compared to non-referential reading in existential sentences. Since Turkish has scrambled word order, it was hard to find a test considering canonical versus existential sentences. That is why we preferred to use case- marking with two alternatives which made a difference in meaning. The results showed that kindergarten children and first graders had scores around 1 which is the score expected by chance. The salient development occurs between 1st and 3rd grade. Making pragmatic inference by picking up morpho-syntactic clues like case- marking is a very advanced meta-linguistic skill. Giving correct answers to the questions needs a comparison between the two case forms and a decision which of them is better suited for the given context. Even adults' performance was not perfect and did not significantly differ from that of 5th graders. However, unlike children some of adults changed their first wrong answer and gave a correct answer after hearing the second question. This shows that some of the adults took the hearer's perspective and/or the experimenter's intention of asking those questions into account. Still, this task might be ameliorated as a production test. For example, after the questions the participants might be asked to give an answer just by using one or two words before the word '… al-dı-m' (I bought …) Thus, the context for using the correct case form would be more natural.

6.5 Development of the Double-embedded Relative Clause Task

We adapted Özge's (2010) single-embedded relative clause task and constructed a double-embedded relative task in order to predict second-order false belief reasoning. To the best of our knowledge this is the first time that a REL_2 task has been devised in a Turkish developmental study. Generally, complement clauses are studied in the literature (cf. de Villiers et al., 2005; Hollebrandse et al., 2011) in order to investigate the relationship between the syntactical component of language and ToM. Unlike complement clauses, relative clauses do not necessarily involve mental state predicates. Using relative clauses instead of complement clauses allows us to specifically focus on the structural format of 2-way embedding. This is a purely structural parallel between 2^{nd} order embedding in the thought domain and 2^{nd} order embedding in the language domain. Our result revealed a very strong developmental trend (Figure 10). Also, adults outperformed 5^{th} graders in this task ($\chi^2(2) = 6.096$, $p = .014$). Whether, however, both LST_2 and FB_2 (partly) tap into the same ability will be discussed in the paragraph on the serial bottleneck hypothesis below.

6.6 Development of the Listening Span Task

This task was mainly used to test the hypothesis that second-order theory of mind reasoning was related to serial processing efficiency, rather than simple working memory capacity. Participants were expected to judge the semantic truth of the sentences, to report it, to remember the last word of that sentence, then repeat the same steps again for the next sentence by also reporting the last word of the previous sentence, and so on. Since in Turkish the present form of the verb takes the suffixes –er, -ar, -ir, -ür, -ur for positive sentences and takes the suffixes –maz, -mez for the negative ones, the most challenging part of the task for children and even for some adults was to repeat the last word of the sentence when its semantic truth was false (e.g. 'Muzlar bisiklete biner' (Bananas ride bicycle). That means that for the example of 'Muzlar bisiklete biner', participants are expected to say "Hayır, biner" instead of "Hayır, binmez". So, they must inhibit the regular way of reporting, and have to report in the instructed from. This inhibition in the LST is thought to be related to false belief reasoning. The results showed a strong developmental trend (Figure 11),

again particularly between 1st and 3rd graders. This finding is compatible with earlier studies on the development of complex working memory, as tested by the LST (Gathercole, 1999; Ünal, 2008). It is the only WM task that does not level off in middle childhood but continues to develop further, probably due to the development of the prefrontal cortex which presumably supports complex working memory. In line with the previous studies reported in the literature was our finding that, again, adults' performance was significantly better than that of 5th graders ($\chi^2(2) = 4.729$, $p = .030$).

6.7 Predictions of Second-order False Belief Task from the other Tasks

If we just look at the correlation between FBT_2 and PTT, the correlation between them is $r_s = .19$ ($p = .126$). That is, PTT and FBT_2 do not share significant amounts of common variance. However, when age is controlled for, the previous correlation of $r_s = .19$ between PPT and FBT_2 even drops further to r = -.095; when WST is controlled for, the correlation drops to .12; when REL_2 is controlled for the correlation drops to .036 and when LST is controlled for the correlation drops to -.022 (Table 15). Among the controlled factors, the effect of age is most prominent.

When we just look at the relationship between FBT_2 and REL_2, we can say that there is a significant correlation between them ($r_s = .54$, $p = .000$) and FBT_2 can be predicted by the following formula: 0,24 X REL_2 + 0.673 ($F_{66,1} = 26.196$, p= .000, r= .533, $R^2 = .284$). However, when this relation is controlled for age, the correlation dramatically decreases (r = .10, p = .42) and becomes insignificant (Table 16). If it is controlled for other tasks the correlation between FBT_2 and REL_2 remains still significant which means the contribution of age is very high. For this reason, the age factor was included in the previous formula: FBT_2= 0.039 X REL_2+ 0.25 X age – 0.751 ($F_{65, 2} = 42.091$, p= .000, r = .751, $R^2 = .564$). As it turns out, age is the only factor whose contribution is significant (β= .692, t= 6.47, p = .000). This result reveals that children's second-order false belief reasoning cannot be predicted by their understanding of double-embedded structure of REL_2 as such but only by their common developmental trajectory. This finding rejects our fifth hypothesis.

In order to see how FBT_2 is predicted by all of the tasks, a two-step multiple regression was used. The correlation between the tasks and FBT_2 are highly significant except for PTT (r = .16, p = .096). The correlation between age and FBT_2 is .75 (p= .000), between WST and FBT_2 is .52 (p = .000), between REL_2 and FBT_2 is .53, between LST and FBT_2 is .50 (p = .000). Since age is a prominent factor, it was entered first. The first model showed the regression between FBT_2 and age as follows: 0.27 X age – 0.814 ($F_{66, 1}$= 83.965, p= .000). In the second step, we put all of the other tasks into the model, in addition to age. The following formula shows the regression equation of this second model: 0.236 X age + 0.145 X WST + 0.045 REL_2 – 0.034 X LST – 0.130 X PTT – 1.098 ($F_{62, 5}$= 17.519, p= .000, r= .765, R^2= .586). However, the only significant contribution comes from the factor age (β= .655, t= 5.45, p = .000). This means that almost all variation between the predictors and the criterion is developmental variation and the predictors as such do not overlap with the criterion.

6.8 Testing the Serial Processing Bottleneck Hypothesis

The testing of this hypothesis is not as straightforward as the above hypotheses. We aimed to evaluate this hypothesis by looking at the relationsbetween various tasks, among them FBT_2 and WST, FBT_2 and LST, as well as REL_2 and WST, and REL_2 and LST. The correlations of these tasks are highly significant (Table 11). The correlations are around .50 except the t correlation between LST and REL_2 (r = .80, p = .000), which is the highest found among all tasks. However, if these correlations are controlled for the age factor, the correlations between the tasks decrease and become insignificant (Table 20, Table 21, Table 23) except for the correlation between REL_2 and LST (Table 22). When we control for age, there still remains a very significant partial correlation between REL_2 and LST (r = .66, p = .000). The regression model for REL_2 and LST can be stated as follows: REL_2= 1.03 X LST+ 0.163 X age – 0.218 ($F_{65, 2}$= 66.286, p= .000, r= .819, R^2= .671) and both the contributions of LST (β= .671, t= 7.10, p = .000) and age (β= .203, t= 2.15, p = .035) are significant. According to the collinearity dignostics, LST (60%) and REL_2 (75%) load highest on the same dimension which shows that both tasks tap into the same cognitive ability.

What do these results reveal in terms of the hypothesis that children's late development of second-order social cognition is due to the lack of serial processing efficiency? This hypothesis cannot be directly proven by just looking at our experimental results. What we find is very strong co-development of second-order false belief reasoning with the WM skills, namely with simple WM capacity (WST) and complex WM (LST) as well as with the complex language comprehension tasks (REL_2, and PTT). The hypothesis of a serial bottleneck would be disproved had the results indicated that the cognitive skills had increased but FBT_2 had not or vice versa, that the cognitive skills had stagnated but FBT_2 had increased. However, this is not what we found. Thus, our results are compatible with the hypothesis that second order false belief understanding may have to wait for these other cognitive abilities to evolve. The same may be said for the relation between complex language comprehension and second order false belief understanding. Our study's results are compatible with the argument that children before the age of 6 may have second-order social cognition, but they may not be able to apply it because of the insufficient developing cognitive resources. Our findings covered in section 6.1 that even 4/5-year-old children correctly answer one or both of the second-order false belief tasks might be interpreted with this argument, since their LST and REL_2 scores are better than the others. To sum up: since in our study we find overall strong positive relations between the WM and linguistic predictors and second order false belief understanding, we can neither rule out the serial bottleneck hypothesis nor prove it. Our results, however, are compatible with the serial bottleneck hypothesis.

In view of theoretical accounts of ToM, our findings are compatible with Leslie et al.'s (2004) account of ToM. He and his colleagues argue that theory of mind is a separate cognitive faculty as compared to language or memory. It is innate, i.e., in principle in place from early on, however, in order to manifest itself it may need to await the cognitive maturation of the child. Since in our study we found concurrent development in all the cognitive abilities that we tested, that is, no delay between any of them, ToM may at any time have been supported just sufficiently enough to manifest itself at that level. Indeed, it might be impossible to prove the relation between ToM and the other cognitive domains in a cross-sectional study like ours

but only in a longitudinal study where such delays may be observed within rather than across inviduals.

CHAPTER 7

CONCLUSION

The main goal of this study was to investigate the effect of the acquisition of Turkish evidential markers on the development of second-order false belief understanding in Turkish children. In addition, the study also aimed to examine the relations between second-order false belief and complex language and relative clause understanding. Finally, in the same vein as Verbrugge's (2009) hypothesis that the developmental latencies between first- and second-order social reasoning is due to the children's need to overcome serial processing rather than simple working memory capacity was studied by looking at the correlations between the social cognition tasks and simple and complex working memory tasks. In order to investigate these, five tasks were used, namely second-order false belief, perspective taking, double-embedded relative clause, word span and listening span.

Our results revealed that there is no effect of acquisition of evidentials on false belief understanding. Together with the other reviewed studies, there is a facilitatory effect of –DI (direct evidence) in understanding for the children at the age of 4 to 6/7, however this facilitation does not reflect facilitation of false belief understanding as such but rather facilitation of understanding of stories in the form of whichfalse belief tasks are presented. A general developmental trend was found for all tasks. Even if significant correlations and bivariate regression results were found between FBT_2 scoresand the other tasks, the regression analyses showed that only the contribution of age was significant. Apart from age, none of these other tasks still could predict FBT_2. Although none of these domains may be related to second-

order false belief reasoning in terms of representational content, but develop at the same time, our findings are not incompatible with the serial bottleneck hypothesis.

In sum, the findings are in line with the modularity view that ToM is a faculty of the human mind in its own right that does not share intrinsic content with other faculties such as language and working memory (Leslie et al., 2004). However, it develops together with those other faculties and these other faculties may constrain its expression in the child's false belief understanding, especially for higher levels of ToM, that is, second order false belief understanding

CHAPTER 8

LIMITATIONS OF THE STUDY, OUTLOOK, AND FURTHER STUDIES

In our study, two verbal second-order false belief tasks were used. The number of these tasks may not be sufficient to test the concept, so it may be increased by adding other second-order false belief tasks also (cf. Meijering, 2011). Apart from increasing the number of tasks of the same kind – false belief – it might also be beneficial to increase the diversity of testing second order theory of mind. For first-order theory of mind this has been done by Wellman and Liu (2004), who used other ToM relevant tasks such as diverse desire, diverse belief, knowledge ignorance and real-apparent emotion tasks. Similarly, for second order theory of mind, various tasks could be designed and a second-order ToM scale might be developed. It would also nice to study a first-order false belief task in order to compare it with second-order false belief reasoning especially for the kindergarten children. Since strategic games are another way of testing second-order social reasoning non-verbally, it might be a very good idea to include one of these tasks in a further study. Also, Liddle and Nettle's (2006) study showed that teacher's ratings of the child were positively correlated with their social reasoning. It would therefore be useful to include teacher ratings into the study. In addition to that, it would also be worth-while to investigate at what age children first use second-order embedding and whether this structure exists in storybooks.

As far as we know, for the first time second-order embedding subject relative clauses have been investigated in Turkish children in the present study. Since the results

revealed a very strong developmental trend, it also appears worth-while to study them in more detail in their own right, including object relative clauses also.

Because of time constraints, a computational model could not be implemented in this study. If an ACT-R model was constructed and validated with the experiments, we could arrive at more direct and valid conclusions for the serial processing bottleneck hypothesis. Constructing an ACT-R model and testing it against the experimental findings and also possibly comparing it with neuro-physiological data would be worth-while studying in the future.

REFERENCES

Aikhenvald, Alexandra Y. (2004). *Evidentiality*. Oxford: Oxford University Press.

Aksu-Koç, A. (1988).*The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge: Cambridge University Press.

Aksu-Koc¸, A., & Alici, D. (2000).Understanding sources of beliefs and marking of uncertainty. In E. Clark (Ed.), *Proceedings of the 30th Annual Child Language Research Forum*, 123–130. Stanford: CSLI.

Aksu-Koç, A., Ögel-Balaban, H. & Alp, I. E. (2009).Evidentials and source knowledge in Turkish. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development, New Directions for Child and Adolescent Developmen*t, 125, 13–28.San Francisco: Jossey-Bass.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004).An integrated theory of the mind.*Psychological Review*, 111(4), 1036-1060.

Anderson, J.R., and C. Lebiere. (1998). T*he Atomic Components of Thought*. Lawrence Erlbaum Associates, Mahwah (NJ).

Astington, JW., and Baird, J. (2005). *Why Language Matters for Theory of Mind.* Oxford University Press; Oxford.

Baddeley, Alan D. (2003): Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.

Bayramoğlu, M., and Hohenberger, A. (2007): The development of Theory of Mind in Turkish preschoolers. Poster presented at the 13th EuropeanConferenceon Developmental Psychology, Jena, Germany, 21-25 August 2007.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation.*Journal of Semantics*, 17, 189–216.

Choi, S. (1995).The development of epistemic sentence-ending modal forms and functions in Korean children. In J. Bybee & S. Fleischman (Eds.), *Modality in grammar and discourse,*165–204. Amsterdam: Benjamin.

De Hoop, H., and Krämer, I. (2005/2006). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, 13, 103–123.

de Villiers, J. (2007). The interface of language and Theory of Mind.*Lingua,* 117, 1858-1878.

de Villiers, J. G. & de Villiers, P. A. (2005). Language and Theory of Mind: What every SLP should know. Paper to be presented to the American Speech Language and Hearing Association Annual Meeting, San Diego, CA.

Drummey, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory.*Developmental Science*, 5, 502–513.

Fitneva, S.A., and Matsui, T. (Eds.), Evidentiality: A window into language and cognitive development. *New Directions for Child and Adolescent Development*, 125, 79-93. San Francisco: Jossey-Bass.

Flobbe, L., Verbrugge, R., Hendriks, P. and Krämer I. (2008). Children's application of theory of mind in reasoning and language.*Journal of Logic, Language and Information*, 17 (4), 2008, 417-442.Special issue on formal models for real people, edited by M. Counihan.

Garfield, Jay L., Peterson, Candida, C., and Perry, Tricia. (2001). Social cognition, language acquisition and the development of the theory of mind.*Mind and Language*, 16, 5, 494-541.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3, 410-419.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-193.

Gerrans, P. (2002).The theory of mind module in evolutionary psychology.*Biology and Philosophy*, 17, 305-321.

Gleitman, L. and Papafragou, A. (2005). Language and thought. In Keith J. Holyoak and Robert G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*, 633-661. Cambridge: Cambridge University Press.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, 59, 1366–1371.

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind.*Journal of Experimental Child Psychology*, 68,70–83.

Gözenman, F. (2011). Acquisition of Evidentiality in Different Age Groups.*Term Paper for the Cognitive Development course*. Unpublished study.

Hala, S., Hug, S. and Henderson, A. (2009). 'Executive Function and False-Belief Understanding in Preschool Children: Two Tasks Are Harder Than One', *Journal of Cognition and Development*, 4(3),275-298

Hasselhorn, M., Mähler, C. and Grube, D. (2005). Theory of mind, working memory, and verbal ability in preschool children: The proposalof a relay race model of the developmental dependencies. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), *Young children´s cognitive development: Interrelationships among executive functioning, working memory, verbal ability, andtheoryofmind*, 219-237. Mahwah, NJ Erlbaum 00:219-237

Hedden, T. and Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games.*Cognition*, 85,1–36.

Hendriks, P., van Rijn, H., and Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua*, 117(11), 1879–1896.

Hollebrandse, B., Van Hout, A. and Hendriks, P. (2011). First and second-order false-belief reasoning: Does language support reasoning about the beliefs of others?In: J. van Eijck & R. Verbrugge (Eds), *Proceedings of the Workshop on Reasoning About Other Minds: Logical and Cognitive Perspectives* (RAOM-2011), Groningen, The Netherlands, July 11th, 2011. CEUR Workshop Proceedings Vol. 751, CEUR-WS.org, pp. 93-107.

Lebiere, C., Wallach, D. , and R. West. (2000). A memory-based account of the prisoner's dilemma and other 2x2 games. In N.A. Taatgen and J. Aasman. (Eds), *Proceedings of Third International Conference on Cognitive Modelin*g, 185–193, Veenendaal: Universal Press.

Leslie, Alan M., Friedman, Ori, and German, Tim P. (2004). Core mechanisms in 'theory or mind'.*Trends in Cognitive Sciences*, 12, 528-533.

Liddle, B., and Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children.*Journal of Cultural and Evolutionary Psychology*,4(3-4), 231-246.

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

Meijering, B., van Maanen, L., van Rijn, H. and Verbrugge, R. (2010). The facilitative effect of context on second-order social reasoning. In: R. Catrambone and S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Cognitive Science Society*, Austin (TX), 1423-1428.

Ögel, H. (2007).Developments in source monitoring and linguistic encoding of source.*Unpublished master's thesis*, Boğaziçi University,İstanbul, Turkey.

Onishi, K.H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?,*Science*, 308, 255-258

Özge, D. (2010). Mechanisms and Strategies in the Processing and Acquisition of Relative Clauses in Turkish Monolingual and Turkish-English Bilingual Children.*PhD thesis,* Middle East Technical University, Ankara-Turkey.

Özge, D., Theodoros, M., Zeyrek, D (2009).Final version to appear in Essays on Turkish Linguistics.*Proceedings of the 14th International Conference on Turkish Linguistics*.

Özoran, D. (2009).Cognitive Development of Turkish Children on the Relation of Evidentiality and Theory of Mind.*Master Thesis,* Middle East Technical University.

Özturk, O., & Papafragou, A. (2008). Acquisition of evidentiality and source monitoring. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings from the 32nd Annual Boston University Conference on Language Development,* [BUCLD 31], 368-377. Somerville, Mass.: Cascadilla Press.

Papafragou, A. and Li, P., (2001). Evidential morphology and Theory of Mind. In: Skarabela, B., Fish, S., Do, A.H.-J. (Eds.), *Proceedings from the 26th Annual Boston University Conference on Language Development.* Cascadilla Press, Somerville, MA, 510–520.

Papafragou, A., Li, P., Choi, Y., & Han, C. (2006). Evidentiality in language and cognition.*Cognition*, 103, 253-299.

Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In J.W. Astington, P.L. Harris, and D.R. Olson, (Eds.).*Developing Theories of Mind*. Cambridge University Press, Cambridge. 271–294.

Perner, J. & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of second-order beliefs by 5- to 10-year old children. *Journal of Experimental Child Psychology*, 5, 125-137

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515-526.

Repovs, G. & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience,* 139(1), 5-21.

Schacter, D.L., Kautsaal, W., & Norman, K.A. (1997).False memories and aging.*Trends in Cognitive Sciences*, 6, 229-236.

Schick, Brenda, de Villiers, Jill, de Villiers, Peter, and Hoffmeister, Robert (2007). Language and theory of mind: A study of deaf children. *Child Development*, 78, 2, 376-396.

Ünal, G. (2008). Release from Proactive Interference and its relations to Executive Functions: a developmental study on Turkish children. *Master thesis*, Middle East Technical University, Ankara-Turkey.

Van Maanen, L., & Verbrugge, R. (2010). A computational model of second-order social reasoning. *Proceedings of the 10th International Conference on Cognitive Modeling*, Philadelphia, PA, USA.

Verbrugge, R. (2009). Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic*, 38 (6), 649-680.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., and Liu, G. (2006). Scaling of Theory-of-Mind understandings in Chinese children. *Psychological Science*, 17 (12), 1075-1081.

Wellman, H. W., & Phillips, A. T. (2001). Developing intentional understandings. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition*, 125-148. Cambridge, MA: MIT Press.

Wellman, H.M. & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Developmen*t, 75 (2), 523-541.

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

APPENDICES

Appendix A: Word Span Task Stimuli

SETS OF 2

Köşk – Muz
Pil – Üst
Buz – Dört

SETS OF 3

Göl - Saç - Tuz
Sev - Kürk - Bel
Kir - Ut – Pas

SETS OF 4

Kaş - Sos - Göc - Yat
Cam - But - Sal - Köy
Zar - Kuş - Tüm - Can

SETS OF 5

Suc - Kek - Böl - Top - Zam

Bal - Kurt - As - Tat - Cöp
Ot - Son - Türk - Seç – Kol

SETS OF 6

Hak - Sus - Tek - Mum - Dip - Kar
Kes - Bin - Ter - Aşk - Yut - Sel
Tren - Kel - Söz - An - Koy - Tez

SETS OF 7

Ak - Top - Su - Alt - Bey - Bol - Mart
Tel - Poz - At - Bil - Yok - Fes - Tür
Kış - Ver - Han - Bot - Yıl - Post - Kül

SETS OF 8

Tam - Bak - Uç - Göz - Hal - Boş - Ek - Yurt
Üç - Kas - Al - Mülk - Bir - Tut - Dil - Kum
Bul - Pek - On - Fal - Var - El - Ses - Genç

Appendix B: Three versions of Birthday Puppy and Chocolate Stories with their drawings

Birthday Puppy Story (Neutral)

Bugün Mehmet'in doğum günü ve annesi ona yavru bir köpekle sürpriz yapmak istiyor.Mehmet'in annesi yavru köpeği bodruma saklıyor.Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" diyor.Annesinin yavru köpekle Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" diyor.

Reality control question: **Annesi doğum günü için Mehmet'e gerçekten ne aldı?**

Şimdi Mehmet annesine "Dışarıya oynamaya çıkıyorum." diyor. Dışarıya çıkarken patenlerini almak için bodruma iniyor. Bodrumda doğum günü hediyesi yavru köpeği buluyor! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" diyor.Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmüyor.

1st order ignorance: **Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor mu?**

Linguistic control: **Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor mu?**

O sırada zır zır zır zır telefon çalıyor! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için arıyor.Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye soruyor.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyor.Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye soruyor.

2nd order false belief: **Mehmet'in annesi anneanneye ne cevap verir?**

Justification: **Mehmet'in annesi neden böyle bir cevap verir?**

Birthday Puppy Story (-DI)

Dün Mehmetlerdeydim. Mehmet'in doğum günüydü ve annesi ona yavru bir köpekle sürpriz yapmak istedi. Mehmet'in annesi yavru köpeği bodruma sakladı.Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" dedi. Annesinin yavru köpekle Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" dedi.

Reality control question: **Annesi doğum günü için Mehmet'e gerçekten ne aldı?**

Mehmet annesine "Dışarıya oynamaya çıkıyorum." dedi.Dışarıya çıkarken patenlerini almak için bodruma indi. Bodrumda doğum günü hediyesi yavru köpeği buldu! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" dedi.Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmedi.

1st order ignorance: **Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor muydu?**

Linguistic control: **Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor muydu?**

O sırada zır zır zır zır telefon çaldı! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için aradı.Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye sordu.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyordu.Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye sordu.

2nd order false belief: **Mehmet'in annesi anneanneye ne cevap verdi?**
Justification: **Mehmet'in annesi neden böyle bir cevap verdi?**

Birthday Puppy Story (-MIŞ)

Bak Mehmet. Geçen hafta Mehmet'in doğum günüymüş. Annesi ona yavru bir köpekle sürpriz yapmak istemiş.Mehmet'in annesi yavru köpeği bodruma saklamış.Mehmet annesine, "Anneciğim, doğum günüm için bana yavru bir köpek almanı çok istiyorum" demiş. Annesinin yavru köpekle Mehmet'e sürpriz yapmak istediğini unutma! Bu yüzden ona yavru bir köpek aldığını söylemek yerine annesi, "Üzgünüm Mehmetciğim, doğum günün için sana yavru bir köpek almadım. Onun yerine sana çok güzel bir oyuncak aldım" demiş.

Reality control question: **Annesi doğum günü için Mehmet'e gerçekten ne almış?**

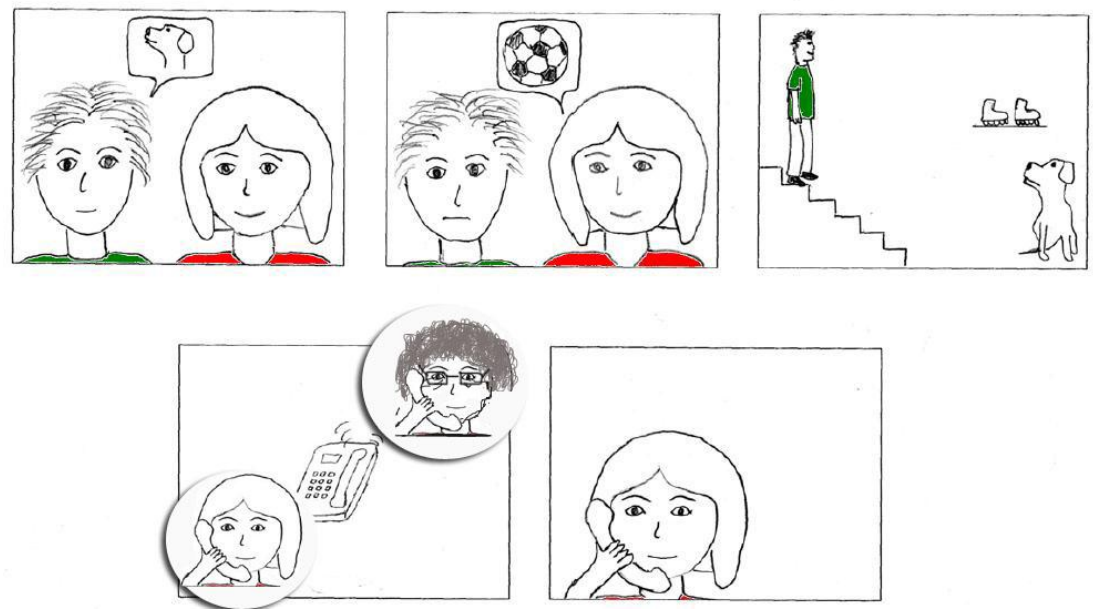
Mehmet annesine "Dışarıya oynamaya çıkıyorum." demiş.Dışarıya çıkarken patenlerini almak için bodruma inmiş. Bodrumda doğum günü hediyesi yavru köpeği bulmuş! Kendi kendine "Vay canına, annem bana oyuncak almamış, gerçekten doğum günüm için bana yavru bir köpek almış" demiş.Annesi Mehmet'in bodruma indiğini ve doğum günü hediyesi yavru köpeği bulduğunu görmemiş.

1st order ignorance**: Mehmet doğum günü için annesinin ona yavru bir köpek aldığını biliyor muymuş?**
Linguistic control: **Annesi Mehmet'in bodrumdaki doğum günü hediyesi yavru köpeği gördüğünü biliyor muymuş?**
O sırada zır zır zır zır telefon çalmış! Mehmet'in anneannesi doğum günü partisinin saat kaçta olduğunu öğrenmek için aramış.Anneannesi telefonda Mehmet'in annesine "Mehmet doğum günü için ona gerçekten ne aldığını biliyor mu?" diye sormuş.

Şimdi hatırlayalım, Mehmet'in annesi, doğum günü için Mehmet'e aldığı şeyi Mehmet'in gördüğünü bilmiyormuş.Daha sonra anneanne Mehmet'in annesine "Mehmet doğum günü için ona ne aldığını düşünüyor?" diye sormuş.

2nd order false belief: **Mehmet'in annesi anneanneye ne cevap vermiş?**

Justification: **Mehmet'in annesi neden böyle bir cevap vermiş?**



The drawings used for birthday Puppy story (Flobbe et al., 2008). The grandmother image was added to the original drawing in order to make the story more explicit.

Chocolate Bar Story (NEUTRAL)

Bak, bunlar Can ile Ece kardeşler.Oturma odasında oynuyorlar.Biraz sonra anneleri alışverişten dönüyor, torbadan bir paket çikolata çıkarıyor.Çikolatayı Can'a veriyor.Ece'ye hiç çikolata vermiyor çünkü yaramazlık yapıyor. Can çikolatanın birazını yiyor ve kalanını çekmeceye koyuyor. Ece'ye hiç çikolata vermiyor.Ece da buna çok sinirleniyor. Can mutfağa annesine yardım etmek için bulaşıkları yıkamaya gidiyor.Ece oturma odasında tek başına oturuyor. Can ise mutfakta. Ece Can'a sinirlendiği için çikolatayı saklıyor.Çikolatayı çekmeceden alıyor ve oyuncak sandığına koyuyor. Can bulaşıkları yıkamakla meşgul. Can, meyve kabuklarını bahçedeki çöp kovasına atmaya giderken pencereden oturma odasını görüyor. Ece'nin çikolatayı çekmeceden alıp oyuncak sandığına koyduğunu görüyor.Ece ise Can'ı görmüyor.

Reality control question:**Çikolata şimdi nerede?**
1st order ignorance: **Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor mu?**
Linguistic control:**Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor mu?**

Can bulaşıkları bitiriyor. Karnı acıkıyor. Çikolatasından biraz yemek istiyor. Can oturma odasına giriyor. "Canım biraz çikolata istiyor." diyor.

2nd order false belief:**Ece çikolata için Can'ın nereye bakacağını düşünüyor?**
Justification:**Ece neden böyle düşünüyor?**

Chocolate Bar Story (-DI)

Bak, bunlar Can ile Ece kardeşler.Geçen gün onların evindeydim.Oturma odasında oynuyorlardı.Biraz sonra anneleri alışverişten döndü, torbadan bir paket çikolata çıkardı. Çikolatayı Can'a verdi. Ece'ye hiç çikolata vermedi çünkü yaramazlık yapıyordu. Can çikolatanın birazını yedi ve kalanını çekmeceye koydu. Ece'ye hiç çikolata vermedi.Ece da buna çok sinirlendi. Can mutfağa annesine yardım etmek için bulaşıkları yıkamaya gitti. Ece oturma odasında tek başına oturuyordu. Can ise mutfaktaydı. Ece Can'a sinirlendiği için çikolatayı sakladı.Çikolatayı çekmeceden aldı ve oyuncak sandığına koydu. Can bulaşıkları yıkamakla meşguldü. Can, meyve kabuklarını bahçedeki çöp kovasına atmaya giderken penceren oturma odasını görüyordu.Ece'nin çikolatayı çekmeceden alıp oyuncak sandığına koyduğunu gördü.Ece ise Can'ı görmedi.
Reality control question: **Çikolata neredeydi?**
1st order ignorance: **Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor muydu?**
Linguistic control: **Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor muydu?**

Can bulaşıkları bitirdi. Karnı acıktı. Çikolatasından biraz yemek istedi. Can oturma odasına gitti. "Canım biraz çikolata istiyor." dedi.

***2nd order false belief*: Ece çikolata için Can'ın nereye bakacağını düşündü?**
***Justification*: Ece neden böyle düşündü?**

Chocolate Bar Story (-MIŞ)

Bak, bunlar Can ile Ece kardeşler.Geçenlerde Can ile Ece oturma odasında oynuyorlarmış.Biraz sonra anneleri alışverişten dönmüş, torbadan bir paket çikolata çıkarmış.Çikolatayı Can'a vermiş.Ece'ye hiç çikolata vermemiş çünkü yaramazlık yapıyormuş. Can çikolatanın birazını yemiş ve kalanını çekmeceye koymuş. Ece'ye hiç çikolata vermemiş.Ece da buna çok sinirlenmiş.Can mutfağa annesine yardım etmek için bulaşıkları yıkamaya gitmiş. Ece oturma odasında tek başına oturuyormuş. Can ise mutfaktaymış. Ece Can'a sinirlendiği için çikolatayı saklamış.Çikolatayı çekmeceden almış ve oyuncak sandığına koymuş. Can bulaşıkları yıkamakla meşgulmüş. Can, meyve kabuklarını bahçedeki çöp kovasına atmaya giderken pencereden oturma odasını görüyormuş. Ece'nin çikolatayı çekmeceden alıp oyuncak sandığına koyduğunu görmüş.Ece ise Can'ı görmemiş.

***Reality control question*: Çikolata neredeymiş?**
***1st order ignorance:*: Can, Ece'nin çikolatayı oyuncak sandığına sakladığını biliyor muymuş?**
***Linguistic control*: Ece çikolatayı saklarken Can'ın onu gördüğünü biliyor muymuş?**

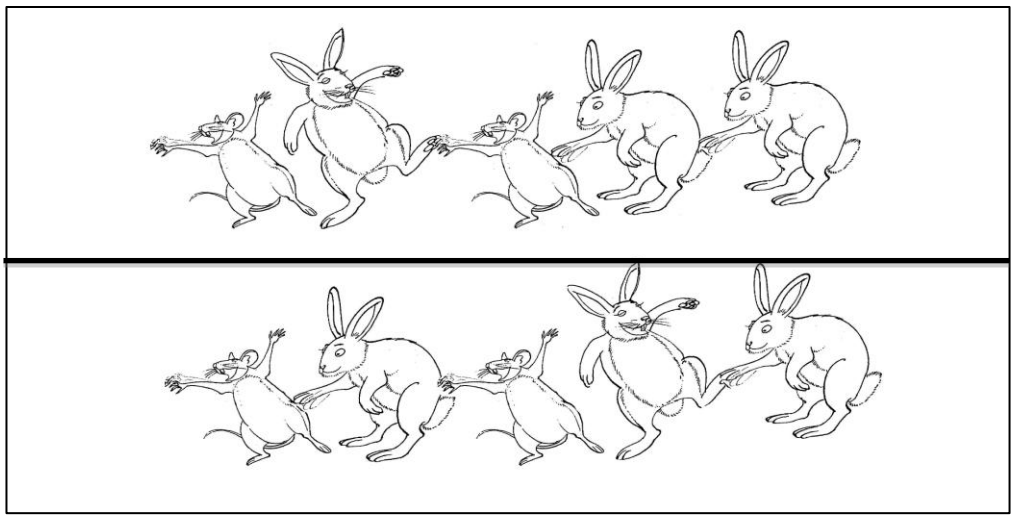
Can bulaşıkları bitirmiş. Karnı acıkmış. Çikolatasından biraz yemek istemiş. Can oturma odasına gitmiş. "Canım biraz çikolata istiyor." demiş .

***2nd order false belief*: Ece çikolata için Can'ın nereye bakacağını düşünmüş?**
***Justification*: Ece neden böyle düşünmüş?**

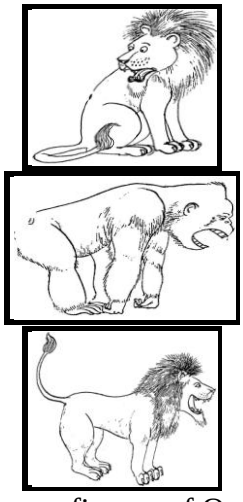
The drawings used for the chocolate bar story (Flobbe et al., 2008)

Appendix C: Second-order Relative Clause Task (REL_2) Questions and Figures

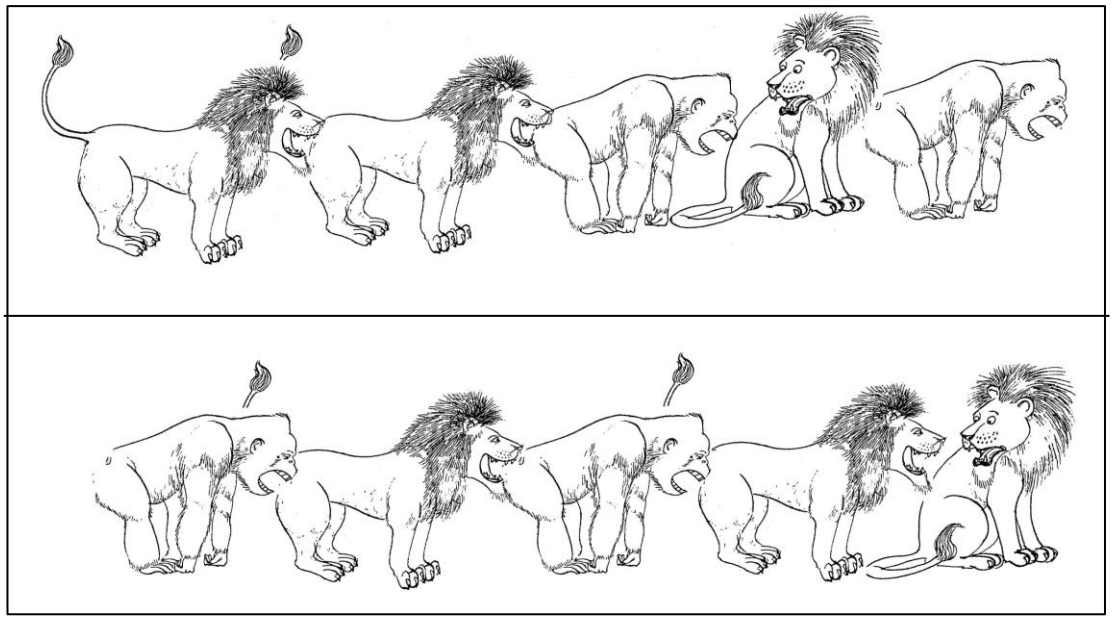
Practice Question: Hangi resimde tavşanı gıdıklayan fareyi gıdıklayan bir tavşan var? ("In which picture there is a rabbit tickling the mouse that is tickling the rabbit?")



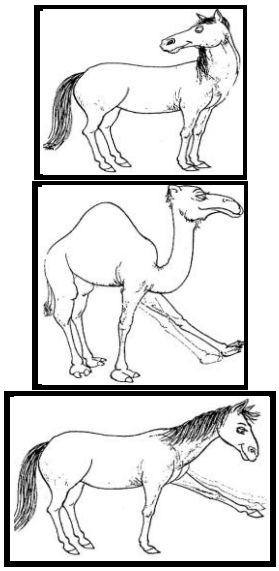
Question 1: Hangi resimde gorili ısıran aslanı ısıran bir aslan var? ("In which picture there is a lion biting the lion that is biting the gorilla?")



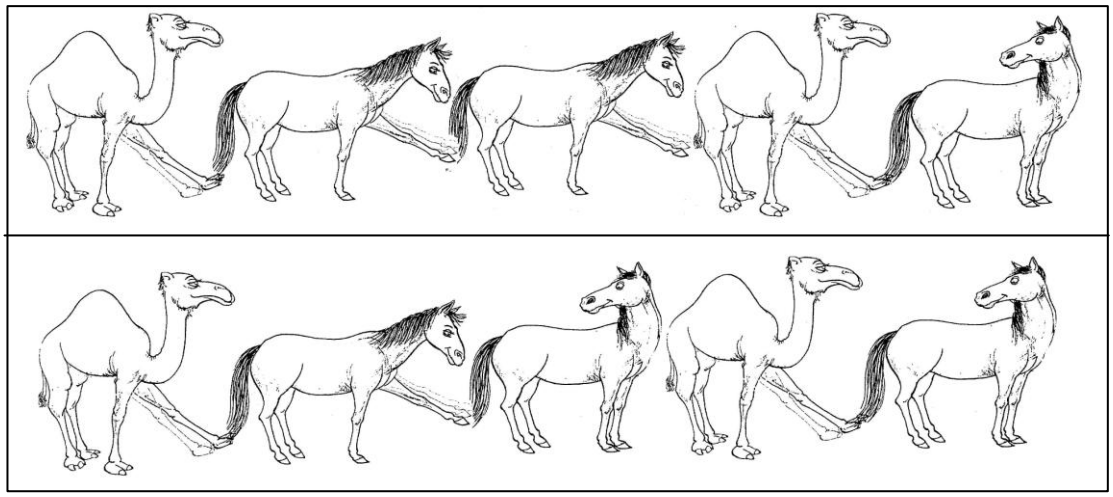
Introductory figures of Question 1



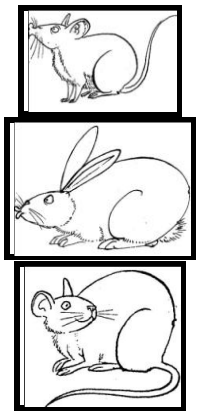
Question 2: Hangi resimde atı okşayan deveyi okşayan bir at var? (“In which picture there is a horse caressing the camel that is caressing the horse?”)



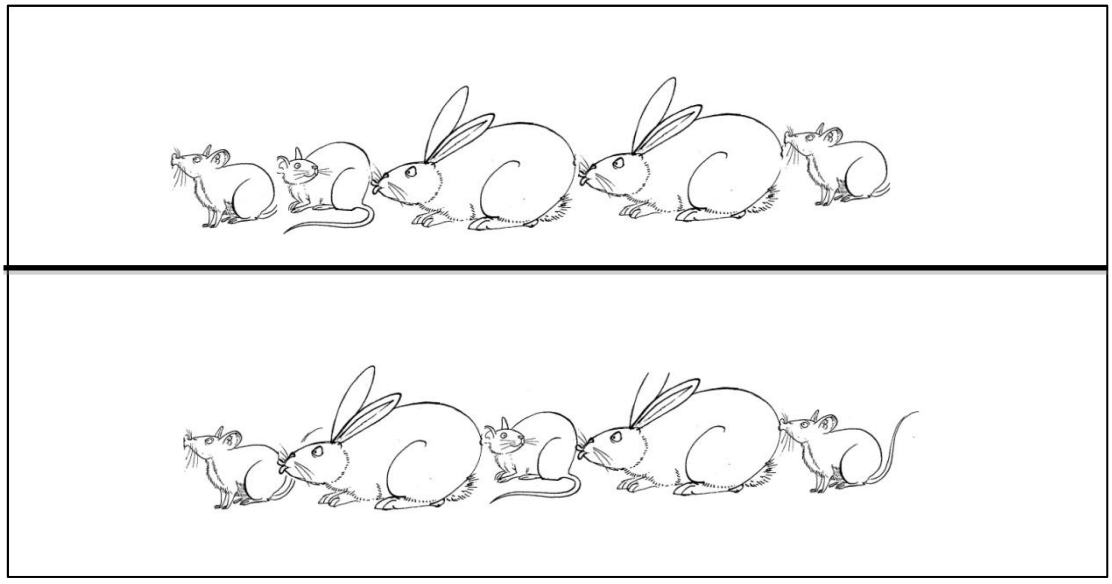
Introductory figures of Question 2



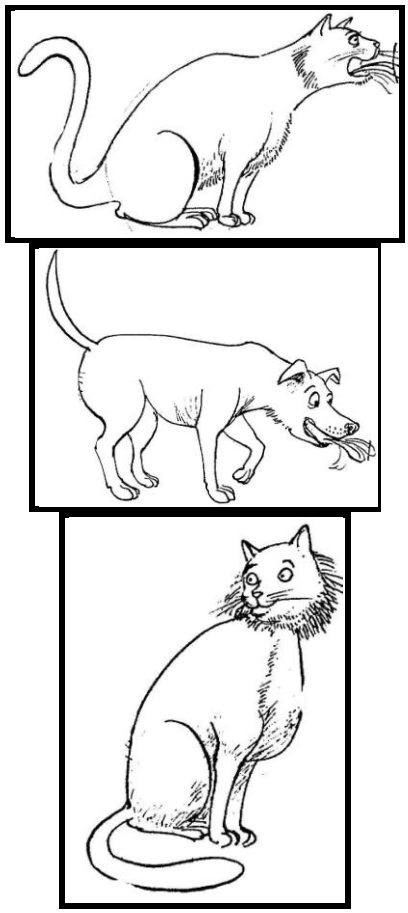
Question 3: Hangi resimde fareyi öpen tavşanı open bir fare var? (“In which picture there is a mouse kissing the rabbit that is kissing the mouse?”)



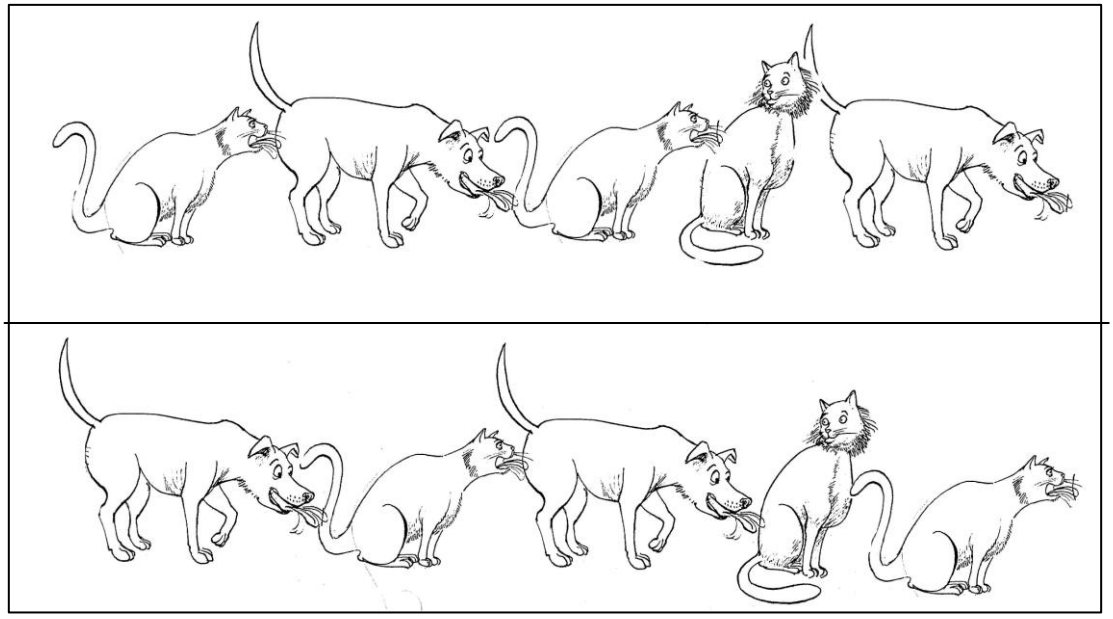
Introductory figures of Question 3



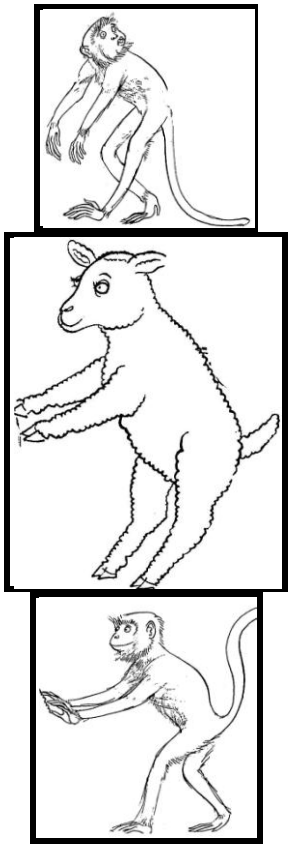
Question 4: Hangi resimde köpeği yalayan kediyi yalan bir kopek var? (“In which picture there is a dog licking the cat that is licking the dog?”)



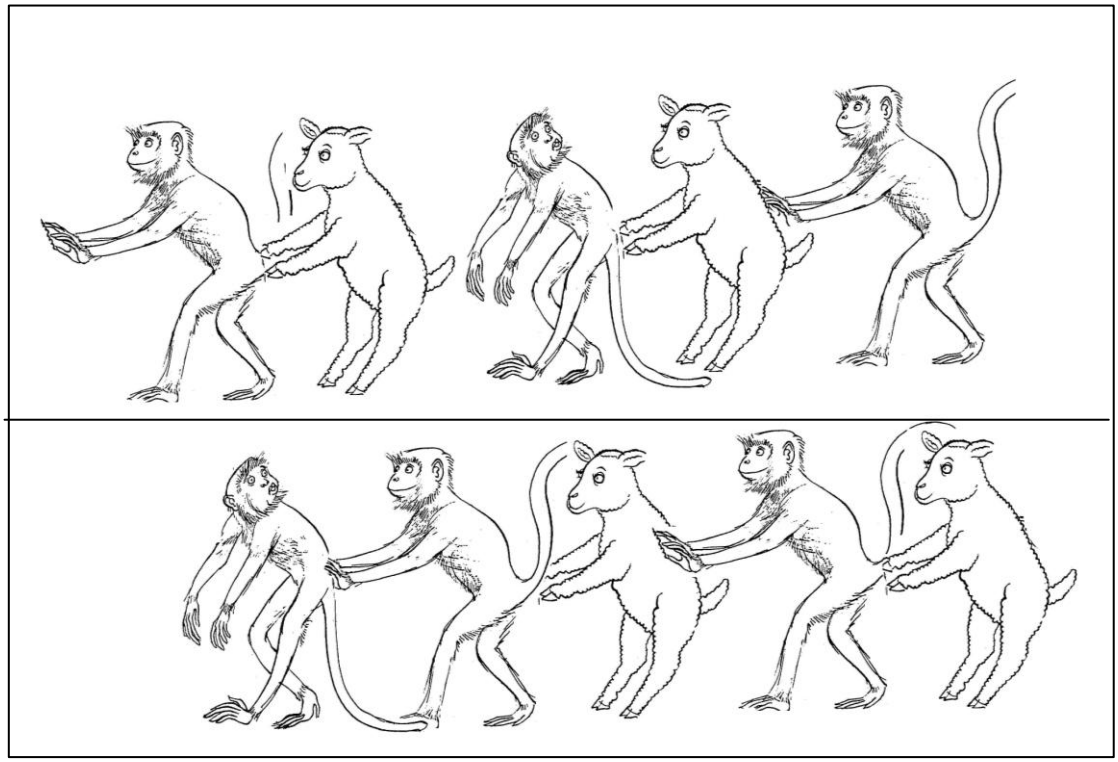
Introductory figures of Question 4



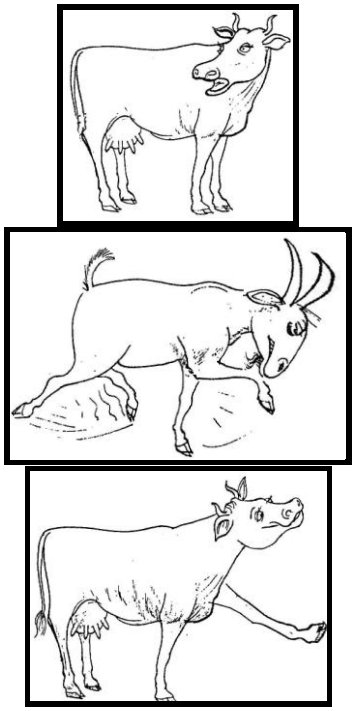
Question 5: Hangi resimde kuzuyu iten maymunu iten bir kuzu var? (“In which picture there is a sheep pushing the monkey that is pushing the sheep?”)



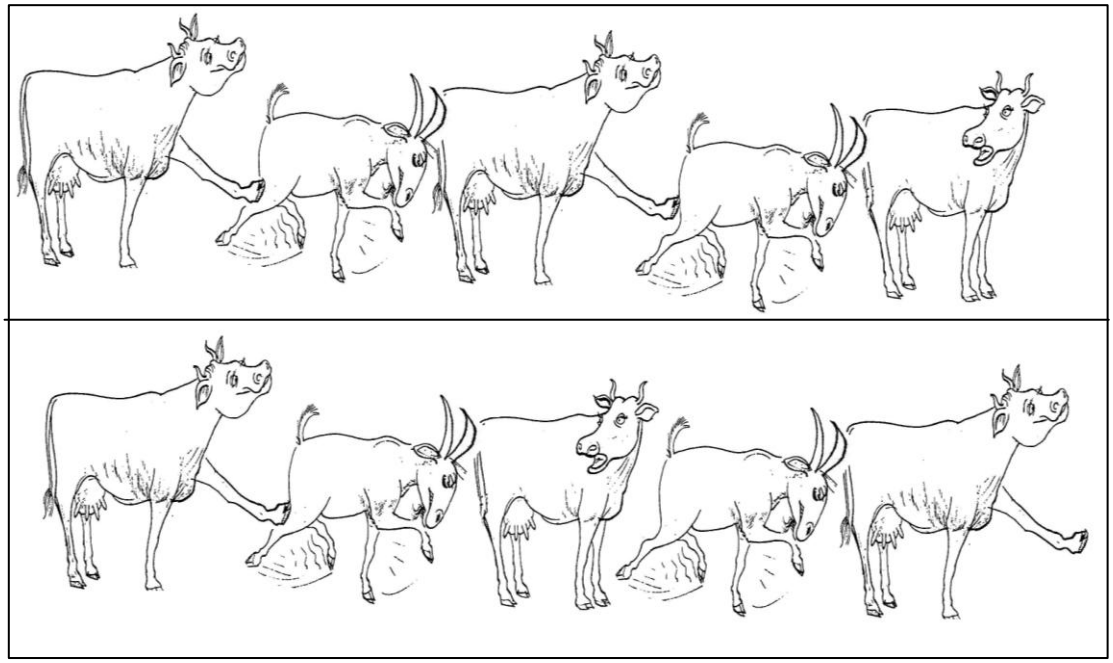
Introductory figures of Question 5



Question 6: Hangi resimde keçiyi okşayan ineği boynuzlayan bir keçi var? (“In which picture there is a goat horning the cow that is caressing the goat?”)



Introductory figures of Question 6



Appendix D: Listening Span Task Stimuli

Used in the trials

1. Çocuklar okula gider.
2. Balıklar havada yaşar.
3. Ağaçlar dans eder.

SETS OF 2

1

1. Biber acıdır.
2. Kediler okulda çalışır.

2

1. Filler çok küçüktür.
2. Ayakkabı ayağa giyilir.

3

1. İnsanlar saçlıdır.
2. Çicekler fare kovalar.

4

1. Ayılar araba sürer.
2. Havuçlar turuncudur.

5

1. Gece karanlıktır.
2. Portakallar suda yaşar.

6

1. Ateş sıcaktır.
2. Balıklar konuşur.

SETS OF 3

1

1. Otobüslerle tatile gideriz.
2. Toplar karedir.
3. Öğretmenler ağaçta yetişir.

2

1. Muzlar bisiklete biner.
2. Elimiz beş parmaklıdır.
3. Soğan acıdır.

3

1. Otobüsler oyuncakla oynar.
2. Kuşlar kanatlıdır.
3. Elmalar ağaçta yetişir.

4

4. Piyanolar müzik çalar.
5. Kardeşlerimiz kuyrukludur.
6. Burnumuzla görürüz.

5

4. Ayağımız çenelidir.
5. Güneş sıcaktır.
6. Taşlar serttir.

6

4. Kaşıklarla yazı yazarız.
5. Limon sarıdır.
6. Köpekler kedileri kovalar.

SETS OF 4

1

1. Zürafalar uzun boyludur.
2. Çiçekler pasta sever.
3. Portakallar kulaklıdır.
4. Öğretmenler okulda çalışır.

2

1. Otobüsler konuşur.
2. Bankalardan para çekeriz.
3. Kışlar sıcaktır.
4. Pastalar tatlıdır.

3

1. Gökyüzü kırmızıdır.
2. Bebekler ağlar.
3. Köpekler konuşur.
4. Muzlar tatlıdır.

4

1. Armutlar mavidir.
2. Şapkalar başa giyilir.
3. Tavşanlar saati gösterir.
4. Filler büyüktür.

5

1. İnsanlar iki ayaklıdır.
2. Portakallar siyahtır.
3. Kediler futbol oynar.
4. Kitapları okuruz.

6

1. Tavşanlar ağaçta yetişir.
2. Biberler yeşildir.
3. Portakallar markette satılır.

4. İnsanlar üç gözlüdür.

SETS OF 5

1

1. Babalar kanatlıdır.
2. Dondurma soğuktur.
3. Portakallar gitar çalar.
4. Arabalar benzinle çalışır.
5. Fareler çok büyüktür.

2

1. Havuçlar mavidir.
2. Kulaklarımızla görürüz.
3. Portakallar turuncudur.
4. Tavuklar yumurta yapar.
5. Bıçak keskindir.

3

1. Elmalar pembedir.
2. Karıncalar yavaştır.
3. Dondurma sıcaktır.
4. Kediler fare kovalar.
5. Bebekler tüylüdür.

4

1. Kuşlar kocamandır.
2. Motorsikletler havlar.
3. Bıçaklar yumuşaktır.
4. Bulutlar beyazdır.
5. Tavuklar yazı yazar.

5

1. Gemiler uçar.
2. Kareler yuvarlaktır.
3. Çorabı ayağımıza giyeriz.
4. Bisikletler süt içer.
5. İnsanlar iki kulaklıdır.

6

1. Uçaklar kanatlıdır.
2. Elmalar şarkı söyler.
3. Dağlar çok küçüktür.
4. Sandalyeler ayaklıdır.
5. Makaslar kağıt keser.

SETS OF 6

1

1. Muzlar dişlidir.
2. Köpekler gitar çalar.
3. Bacağımız parmaklıdır.
4. Mektupları pulla göndeririz.
5. Muzlar sarıdır.
6. Kurbağalar zıplar.

2

1. Oyuncak ayılar yumuşaktır.
2. Ördekler suda yaşar.
3. Çocuklar üç kolludur.
4. Evimiz şarkı söyler.
5. Ördekler beş ayaklıdır.
6. Kar soğuktur.

3

1. Saatler zamanı gösterir.
2. Ayran tatlıdır.
3. Kurbağalar uzun kulaklıdır.
4. Ağaçlar müzik çalar.
5. Toplar yuvarlaktır.
6. Balıklar suda yaşar.

4

1. Arılar sokar.
2. Koyunlar kuyrukludur.
3. İnekler uçar.
4. Köpek balığıkocamandır.
5. Bulutlar siyahtır.
6. Pamuk ağırdır.

5

1. Ağaçlar tüylüdür.
2. Marketler yiyecek satar.
3. Domates kırmızıdır.
4. Kediler çok büyüktür.
5. Tavşanlar uzun kulaklıdır.
6. Tavuklar okula gider.

6

1. Kirazlar mavidir.
2. Ağaçlar yapraklıdır.
3. Demir hafiftir.
4. Yılanlar zıplar.
5. Kekler tatlıdır.
6. Tekerlekler karedir.