İREM ÖZTÜRK MIHCI DIVERSE BELIEF TASK PERFORMANCE Bilkent University 2022

DETAILED INVESTIGATION OF TURKISH CHILDREN'S DIVERSE BELIEF TASK PERFORMANCE

A Master's Thesis

by

İREM ÖZTÜRK MIHCI

The Department of Psychology
İhsan Doğramacı Bilkent
University
Ankara

December 2022

To My Family

DETAILED INVESTIGATION OF TURKISH CHILDREN'S DIVERSE BELIEF TASK PERFORMANCE

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

İREM ÖZTÜRK MIHCI

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN PSYCHOLOGY

THE DEPARTMENT OF
PSYCHOLOGY
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

December 2022

DETAILED INVESTIGATION OF TURKISH CHILDREN'S DIVERSE BELIEF TASK PERFORMANCE

By İrem Öztürk Mıhcı

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.

Hande Ilgaz

Advisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.

Jedediah W. P. Allen

Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.

Sibel Kazak-Berument

Examining Committee Member

Approval of Graduate School of Economics and Social Sciences

Refet Soykan Gürkaynak

Director of the Graduate School

ABSTRACT

DETAILED INVESTIGATION OF TURKISH CHILDREN'S DIVERSE BELIEF TASK PERFORMANCE

Öztürk Mıhcı, İrem

M.A., Department of Psychology

Supervisor: Asst. Prof. Hande Ilgaz

December 2022

This study investigated Turkish children's Diverse Belief (DB) task performance, which is one of the tasks in Theory of Mind (ToM) battery. Previous literature found that Turkish children underperformed on the DB task when compared to children from other cultures. In addition, Turkish children did not show the expected age-related increase in the DB task performance, and Turkish children's DB performance was not related to the other ToM tasks and cognitive variables such as EF. Therefore, the aim was to examine whether the difficulty in the DB task is related to the task structure for Turkish children. The DB task performances and the EF skills of 45 Turkish-speaking children aged between 3 and 5 were assessed. The DB task was manipulated in terms of mental state verbs (e.g., think, guess, and no mental states)

and the number of characters in the story (e.g., single, and double seekers). Results showed that Turkish children were not affected by the differences in mental state verbs, but the number of characters in the story affected performance. Children were more successful in the double-seeker conditions when the task was presented with the verb “think” (düşün- in Turkish). In the light of these results, it is possible to infer that the problems regarding Turkish children’s DB performance may be related to the curse of belief rather than Turkish children’s conceptual deficiency in belief understanding.

Keywords: Theory of Mind, Diverse Belief, Executive Functions, Socio-Cultural Approach

ÖZET

TÜRK ÇOCUKLARIN FARKLI İNANÇLAR GÖREVİNDEKİ PERFORMANSLARININ DETAYLI İNCELEMESİ

Öztürk Mıhcı, İrem

Yüksek Lisans. Psikoloji Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Hande Ilgaz

Aralık 2022

Bu çalışma Türk çocukların zihin kuramı (ZK) bataryasındaki görevlerden birisi olan Farklı İnançlar (Fİn) görevindeki performanslarını incelemektedir. Literatürdeki çalışmalar, Türk çocukların diğer kültürlerdeki çocuklarla kıyaslandığında Fİn görevinde düşük performans sergilediğini saptamıştır. Buna ek olarak, yürütülen diğer çalışmalarda Türk çocukların Fİn görevinde yaşa bağlı beklenen performans artışını gösteremedikleri görülmüştür. Fİn görevinin zihin kuramı bataryasındaki Farklı İstekler (Fİs) görevi haricindeki diğer görevlerle ilişkili olduğunun bulunamaması ve Fİn görevindeki performansın diğer bilişsel becerilerle (örn. yürütücü işlevler) yordayıcı bir ilişkisinin olmaması dolayısıyla, bu çalışma Fİn yapısının uygunluğunu araştırmayı hedeflemektedir. Türkçe konuşan 3 ve 5 yaş aralığındaki 45 çocuğun Fİn görevindeki performansları ve yönetici işlevler becerileri çalışılmıştır. Fİn görevi

zihin durumu fiilleri (örn. düşün-mek, tahmin et-mek ve -ce/-a göre) ve hikayedeki karakter sayısı (örn. bir ve iki) açısından manipüle edilmiştir. Çalışmanın sonuçları, Türk çocuklarının zihin durumu fiillerindeki farklılıklardan etkilenmediklerini fakat hikayedeki karakter sayısının FİN performansı üzerinde değişiklik yarattığını göstermiştir. İki karakterli görevlerde çocukların daha başarılı oldukları görülmüştür. Sonuçlar, Türk çocuklarının FİN görevindeki performans düşüklüğünün kavramsal olarak inanç çeşitliliğini anlama eksikliği yerine, çocuklara görevde kendi fikirleri sorulduğunda inanç taraflılığı (curse of belief) yaşamaları ile açıklanabileceği şeklinde yorumlanmıştır.

Anahtar Kelimeler: Zihin Kuramı, Farklı İnançlar, Yürütücü İşlevler, Sosyo-Kültürel Bakış Açısı

ACKNOWLEDGMENTS

First and foremost, I would like to express my appreciation to my dear supervisor, Asst. Prof. Hande Ilgaz. Since my undergraduate years, she has embraced me with her endless knowledge and beautiful heart. She has supported me in every step I have taken regarding my career. She lit my way whenever I was afraid to be honest with myself and whenever I did not know how to shape my future. Thanks to her, I realized my potential during the six years we spent together. It would have been impossible to complete this thesis without her presence and efforts. I am heartedly grateful for everything she has brought into my life.

Meanwhile, I would like to express my gratitude to Asst. Prof. Jedediah Wilfred Papas Allen. His role in my career has always been invaluable. Thanks to his courses and intellectual knowledge, I became an individual who questions, criticizes, and thinks more deeply. He always encouraged me to realize my potential and paved the way for me to think about what I could do better at every stage of this thesis. I am also grateful to Prof. Dr. Sibel Kazak-Berument for her priceless contributions to my thesis.

It is surely beyond doubt that the support and love of my dear friends İlke Candar and Begüm Pala psychologically motivated me throughout the whole process. Knowing that I have friends that I can trust endlessly was a chance for me. I am so lucky to have them in my life and I promise to always be there for them. I would like to thank

Feride Nur Haskaraca, Elif Bürümlü-Kısa, Setenay Evsen, Ece Tuğlacı, and Onur Sunay for their help and support. Concurrently, I owe special thanks to my dear friends Rana Ahsen Ak and Yağmur Esendemir for their efforts during the material preparation process. I am also thankful to all the participants and their families.

During my graduate years, I again realized what a loving and perfect family I have. I cannot thank enough to the most precious people in my life, my mother Belgin Öztürk and my father Şahin Öztürk who unconditionally support me and believe in me at every step of my life. Especially, I would like to thank my brother Yiğit Arıkan Öztürk who I love beyond words. It would be impossible to accomplish this journey without your sincere love and endless support. I am also indebted to Tülin Mıhcı, Altan Mıhcı, and Atalay Mıhcı for accompanying me throughout this experience.

My soulmate, Atamert Mıhcı. I am not sure how well I can verbalize my appreciation to you. You have always been my pillar of support along this process having ups and downs. I am grateful that the universe brought us together among its endless possibilities, that you have made room for me in your heart, and that we will walk the bumpy road of life hand in hand. I am looking forward to the countless things we will celebrate together. I really love you. I really love loving you.

Finally, I would like to add a little note to myself: “When you read this years later, remind yourself that you cried, you fell down, and you got disappointed, but you never gave up. Keep being who you are!”

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The ability to understand the mental states of others has been the subject of research in psychology since the 1970s (Wellman, 2017) and it is termed as theory of mind (Premack & Woodruff, 1978), mind reading (Whiten, 1994) or social understanding (Carpendale & Lewis, 2004). The early focus was on children's ability to understand other people's false beliefs. Wellman and Liu's (2004) ToM battery that aimed to assess a variety of related mental abilities (e.g., desires, beliefs, knowledge states, false beliefs, and emotions) broadened the scope of research in this area. ToM studies with children from different cultures revealed that there was variability in the developmental trajectories of ToM tasks. For example, American children passed the diverse belief (DB) task before the knowledge access (KA) task (Wellman & Liu, 2004), while Chinese children showed the reverse sequence (Wellman, Fang, Liu, Zhu, & Liu, 2006). Regarding Turkish children's ToM competence, there are three gaps in the literature. First, Turkish children's DB performance seems lower than their counterparts in other cultures (Ilgaz, Allen, & Haskaraca, 2022; Selçuk, Brink, Ekerim, & Wellman, 2018). In addition, recent studies also raised questions about Turkish children's age-related DB success, the correlations of the

individual ToM tasks with other tasks in ToM battery, and cognitive variables such as EF. Children were expected to show an age-related performance increase in each task in ToM battery. However, the DB performance did not meaningfully change with age (Ilgaz et al., 2022). Moreover, the DB task was not related to the other tasks in the battery except the DD task (Ilgaz et al., 2022) and other child-related cognitive variables such as the EF (Doenyas, Yavuz, & Selçuk, 2018).

Previous literature claimed that Turkish children's relatively lower DB task performance may indicate their deficiency in conceptual belief understanding (Selçuk et al., 2018). They attempted to explain the conceptual deficiency depending on the distinction between individualism and collectivism. Since Turkey is a collectivistic culture that values authority and harmony, Turkish children may have difficulty understanding that two persons may have divergent beliefs (Selçuk et al., 2018). However, Ilgaz et al. (2022) found that Turkish children's sequencing of ToM tasks changes with age. While 3-year-old children were likely to pass the DB before the KA task, the sequence changes both for 4-year- and 5-year-olds. The fact that children from the same culture showed different task sequencing depending on age leaves a question mark over minds regarding the connections between sequential variations and individualism-collectivism distinction. In addition, the lack of age-related performance increase, the lack of relations with the other tasks and with the other cognitive variables of the DB task may suggest that Turkish children's problems with the DB task may be due to a combination of structural and cognitive factors.

One of the problems with the Turkish version of the DB task may be due to the literal translation of the task that may be interpreted slightly differently by Turkish children. Since Turkish is a highly grammaticalized language, Turkish speakers are required to

explicitly mark whether they have first-hand knowledge or evidence-based inference for a particular statement through the use of suffixes (e.g., *-DI* and *-MIŞ*) (Aksu-Koç, Ögel-Balaban, & Alp, 2009). Due to its morpho-syntactic structure emphasizing evidentiality, mental state verbs in Turkish may also indicate the speaker's certainty about the accuracy of the statement. For instance, although the verb "*think*" ("düşün-" in Turkish) is a non-factive verb (i.e., non-truth expressing), it is likely to be used when there is a reason or evidence to hold a particular belief as a factive verb (i.e., truth expressing) (Haskaraca, 2019). Therefore, the use of the verb "*think*" (düşün-) used in the task story of the DB task may cause Turkish children to experience difficulties because children are merely asked to guess the location of the object in the target task. In the story, the protagonist aims to find the lost animal (i.e., the cat) and the cat is either behind the bushes or under the car. Since there is no clue about the location of the lost animal, there is no explicit reason for children to hold a particular belief about where the cat is. Yet, the verb "*think*" (düşün-) may encourage children to seek evidence. In consideration of Turkish evidentiality, previous studies suggested that task language may affect Turkish children's performance on the belief-related tasks. Although Turkish adults reported that the verb "*guess"* is preferable to the verb "*think"* in the belief-related tasks (i.e., diverse belief and false belief tasks), children's task performances were not affected by the change in mental state verbs (Haskaraca, 2019). Likewise, Turkish adults thought that there was no difference in terms of certainty levels between the verb "*think*" and the phrase "*in my opinion*" (Balcı, 2021). In the current study, the first aim was to investigate whether previous findings regarding the verb "*think*" and the verb "*guess*" will be replicated and whether the lack of difference in certainty between the verb "*think"* and the phrase "*in my/your*

opinion" for Turkish adults will be manifested in Turkish children's DB task performance.

In addition, this study is the first to investigate that Turkish children's underperformance on the DB task may be related to a cognitive bias similar to the way that the *curse of knowledge* operates in false belief tasks. *Curse of knowledge* is defined as children's tendency to be biased by their current knowledge when predicting a misinformed or ignorant person's perspective. In other words, individuals' own knowledge becomes more salient for them than a less informed person's knowledge. This causes individuals to make predictions about the other person's behavior that attributes more knowledge to that person than what is objectively available. It was found that younger children's failure in false-belief tasks may be related to the *curse of knowledge*. Since children know the true location of the object, they are biased by their knowledge and they have difficulty understanding that the protagonist will look for the object in the previous location by acting according to their false belief (Ghrear, Fung, Haddock, & Birch, 2021). Although there is no knowledge state in the DB task unlike the FB tasks, children are asked to express a particular belief about the location of the animal in the original DB task. After the child produces a belief about the location of the object, it is stated that the protagonist has a belief that contradicts the participant's belief in the story. Children may be biased by their beliefs, and they may have difficulty understanding that the protagonist will act in line with his/her belief. Therefore, in addition to the changes in the verbs and phrases used in the story, the current study created a new DB task in which the child's own belief was not asked. In this version, there are two characters with different beliefs in the story, and the child does not need to produce a belief about the location of the object. Thus, it is intended to eliminate the possibility of

children being hindered by their own beliefs. From this point of view, this study aims to investigate whether Turkish children's difficulties in the DB task can be explained by the *curse of belief* which is we named inspired by the curse of knowledge.

Therefore, the general aim of the current study is to understand why Turkish children have difficulties with the DB task. The research question is whether Turkish children have a conceptual deficiency in belief or whether the original DB task is methodologically challenging for Turkish children.

In this chapter, two relevant theoretical approaches to ToM development (e.g., theory theory and socio-cultural approach) and the creation of Wellman and Liu's (2004) ToM battery will be explained. Following, the aforenamed problems related to culture, age, and relations in ToM tasks will be discussed in detail to provide a better understanding of the gaps in the literature. This study aims to answer these questions through the epistemic differences across the mental state lexicon in Turkish and the curse of belief. Accordingly, relevant information about these subjects will be presented.

1.1. Theory of Mind and Theoretical Backgrounds

1.1.2. Theory Theory

Gopnik and Wellman (1992) claimed that children are born with different domains in their minds. Children form theories in these pre-existing domains through their interactions with their environments. Altering and revising these theories in the folk psychology domain allows children to understand others' minds. According to this view, children are first described as desire psychologists. Later, their understanding of belief and other mental states progresses with age (Bartsch & Wellman, 1995).

Theory Theory has some assumptions about the relations between culture and development. There is an *instrumental* relation between the concept of culture and reasoning about mental states (Ilgaz & Allen, 2021). In other words, TT position claims that children are already equipped with theories about the mind at birth and the interactive experiences help them to revise their pre-existing theories. According to this view, ToM is universal because it is an innate theory about everyday life. Additionally, it is partly cultural because children's unique experiences shape their theories. However, it is claimed that the cultural aspects of ToM construct do not alter the passing order of the tasks because children are firstly defined as desire psychologists and then belief psychologists. That is, children need to pass ToM tasks in an invariant order (Wellman & Woolley, 1990; Wellman, Fang, Liu, Zhu, and Liu, 2006). To sum up, TT approach explains the phenomenon of ToM as the product of innate theories which become more advanced through social interactions.

1.1.3. Socio-Cultural Approach

In general terms, socio-cultural approach can be situated in the action-based principle that emphasizes the active role of children in development (Allen & Bickhard, 2013). Specifically, socio-cultural perspective is different from TT in two main aspects: the emergence of knowledge and the role of culture (Ilgaz & Allen, 2021). Firstly, SC approach claims that interaction and experience underlie the emergence of knowledge. This contrasts with TT's claim of innate theories (Carpendale & Lewis, 2004). Secondly, these two contrasting approaches differently interpret the role of culture in development. SCU claims that understanding of the mind is shaped by culture (Nelson, 1998). Children's social interactions with others determine their ways of understanding others' minds (Tomasello, 2019). Unlike TT, socio-cultural

approach claims that the interaction is *constitutive* for development. This means that children's interactions with other people or their environments lead to the emergence of an understanding of others' minds. To sum up, SC approach explains the phenomenon of ToM as the product of the interactions children experience through outside environment and culture. Therefore, it leaves the door open for cultural variations in the progression of different mental terms.

1.2. Development of the Wellman and Liu's (2004) ToM Battery

There was a particular focus on false belief understanding (FBU) in early ToM literature (Peterson, Wellman, & Liu, 2005). One of the most frequently used versions of the FB measurement was created by Wimmer and Perner (1983). In this task, a character puts his chocolate in the kitchen cupboard and leaves. While the character is not in the kitchen, his mother comes. She takes the chocolate from the cupboard and puts it in the drawer. When the character returns, participants are asked where the character will look for his chocolate. The FBU was regarded as the definitive of ToM because it requires individuals to be able to make a distinction between a belief about the state of the world and the external reality (Wellman et al., 2001).

In addition to the FBU, later research highlighted the significance of the appreciation of the variability in desires, beliefs, and emotions in response to the same object or event as a critical component of ToM competence (Ruffman, Slade, & Crowe, 2002). In line with this objective, Henry M. Wellman and David Liu (2004) created a ToM battery to consolidate the measurement of ToM competence. The benefit of such a battery would be to facilitate the investigation of the developmental progression of ToM understanding. In the first study of their paper in 2004, Wellman and Liu conducted a meta-analysis of various ToM tasks. Specifically, they focused on three

contrasts: diverse belief vs. false belief, diverse desire vs. diverse belief, and knowledge vs. false belief. There was a crucial difference between diverse beliefs and false beliefs. In the false belief tasks, the truth is known by the child. One belief is false, while the other is true. However, in the diverse belief task, the truth is unknown. Two persons have divergent beliefs, one not truer than the other. Results revealed that children appreciate others' diverse beliefs earlier than their false beliefs. Likewise, it was easier for children to appreciate diverse beliefs and knowledge states than to understand others' false beliefs. Even though the previous literature had inconsistent results about the relations between knowledge and false belief, the ability to understand the knowledge states appeared earlier than understanding of false beliefs. As there is a progressive development in understanding different mental states, researchers aimed to create a scalable ToM battery.

Diverse desires, diverse beliefs, knowledge and ignorance, false belief (content false belief and explicit false belief), belief-emotion, and apparent emotion tasks were assessed. To control for complexity, the tasks were matched in terms of the accompanying stimuli and length. Wellman and Liu (2004) aimed to understand which tasks should be included in the ToM battery, and consequently scale the tasks according to their levels of difficulty. The Diverse Desires (DD) attempts to assess whether children are able to understand that two individuals may have different snack preferences. The Diverse Beliefs (DB) task probes whether children are able to understand that two individuals may hold different beliefs about an unknown fact and act accordingly. Specifically, children see the pictures of a gender-matched protagonist and two locations (i.e., bushes, and a garage). They are told that the protagonist wants to find his/her cat. The cat is either in the bushes or in the garage. Then, children are asked about their own beliefs (i.e., "Where do you think the cat

is?"). They are told that the protagonist holds the opposite belief. The target asks: "Where will the protagonist look for the cat, in the bushes or the garage?" The child is considered as successful if he/she answers the target question opposite from his/her own belief. Knowledge Access (KA) measures the understanding that one cannot know the content of a nondescript box without seeing the inside the box. The Content False Belief (CFB) is regarding whether children are able to understand that someone else may hold a false belief about a diacritical box with surprising content. The Explicit False Belief (EFB) tries to assess children's ability to understand that one's mistaken beliefs determine his/her behaviors. Finally, The Hidden Emotion (HE) task probes the ability that one can express a particular emotion while feeling a different one.

1.3. Theory of Mind and Cultural Variations

The battery created by Wellman and Liu (2004) brought about an increase in cross-cultural ToM studies as it became a matter of curiosity whether there are cultural variations. Several studies were conducted with children from cultural and linguistic backgrounds (e.g., Wellman, 2006; Selçuk et al., 2018; Shahaeian, Henry, Razmioee, Teymoori, & Wang, 2015). Cross-cultural scaling studies revealed the differences in the sequencing of ToM tasks. There are two sequences documented in the literature: Western and Eastern sequences. Scaling studies with Turkish children detected the presence of both sequences. This section will explain the documented sequences and the research conducted with Turkish children.

1.3.1. Western Sequence

Studies mainly with English-speaking children of European origins revealed the existence of a certain line-up in which passing the DB is earlier than the KA among

ToM tasks. In their pioneering work, Wellman and Liu (2004, Study 2) studied children living in the United States and Canada. ToM battery consisting of 7 items was administered to 75 American children aged between 3 and 5. Analyses revealed that 5 of the tasks were scalable. Since the belief-emotion task was not scalable and the explicit false belief task (EFB) showed very close levels of success with the content false belief (CFB) task, these two tasks were not included in the analysis. The sequence of children living in the US and Canada was as follows: DD (95%), DB (84%), KA (73%), CFB (59%), and HE (32%). Put it differently, while DD was the easiest task for them to pass, HE was the hardest one.

Peterson and her colleagues (2005) reported that Australian children show the same sequence as their American and Canadian counterparts. They had 4 groups of children with different characteristics (e.g., children with deafness, autism, and typical development). Since the sample of the current study comprises typically developing children, their ToM sequencing will be explained. There were 62 typically developing children aged between 3 and 5. Only the objects in the stories and the names of the protagonists were changed to make the tasks culturally appropriate for Australian children. Typically developing children showed the same sequence as American and English-speaking Canadian children. In short, they passed the DD (95%) before DB, DB (85%) before KA, KA (73%) before CFB, and CFB (32%) before HE (19%). Similarly, German children followed the same sequence as their peers from America, Canada, and Australia (Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006 as cited by Wellman, 2012).

The Western sequence was observed in non-Western cultures as well. Kuntoro, Saraswati, Peterson, and Slaughter (2013) conducted a study with children living in

Indonesia. The aim was to investigate whether there are micro-cultural variations across three groups: children who are trash pickers in Jakarta living in low socio-economic circumstances, middle-class Jakartans, and Australian children coming from middle SES. ToM tasks were presented in their native languages. Although there were dramatic variations in socio-economic backgrounds, children from the quite low SES group showed the same sequence as their American and Australian counterparts.

1.3.2. Eastern Sequence

The other documented sequence is named as Eastern sequence which was found in children from China, and Iran. Wellman and his colleagues (2006) conducted a study on Chinese children. ToM performance of Mandarin-speaking Chinese pre-schoolers was compared with their Australian and American peers. The comparison was made both for children's overall ToM scores and the scaling of ToM tasks. There was no significant difference in children's total ToM performance calculated through aggregating their success in the tasks of the ToM battery. However, sequence differences were documented. Contrary to their American and Australian peers, Chinese children performed better in the KA task. That is, the sequence of success was as follows: DD (89%) > KA (79%) > DB (71%) > CFB (54%) > EFB (49%) > HE (37%).

Shahaeian et al. (2011) investigated ToM sequencing in a comparative study with 77 English-speaking Australian and 58 Farsi-speaking Iranian children. The wording of narratives and questions in ToM tasks (e.g., target question) remained the same even though the tasks were organized according to the different cultural characteristics of the two groups (e.g., character names, and objects). Both groups of children had similar levels of overall ToM mastery. However, unlike Australian and American

children, Iranian children found the KA (88%) task easier than the DB (47%) task. That is, Iranian children showed the same sequence as Chinese children. Emphasizing that the two samples match each other in terms of demographics (e.g., preschool attendance, parental education, socioeconomic status), researchers suggested that this variation could only be due to a cultural difference.

1.3.3. Scaling Studies with Turkish Children

Studies with Turkish children demonstrated the existence of both Western and Eastern sequences that change depending on age and EF abilities. However, the reason for the multiplicity of sequences is not clear. Etel and Yagmurlu (2015) conducted a study with children who do not live with their biological families and stay in institutions. This study had two main goals. First, researchers aimed to measure institution-reared children's ToM understanding. Secondly, they attempted to examine Turkish children's ToM scaling. This study was conducted with 107 children aged between 3 and 5. Children who are under government guardianship found the DB task easier than the KA task. The sequence was as follows: DD (91%) was easier than DB, DB (71%) was easier than KA, KA (44%) was easier than EFB (20%), and CFB (12%). In other words, although these underprivileged children faced parental death, abuse, or poverty, they had the same ToM sequence as children who live in Western populations such as the United States, Canada (Wellman & Liu, 2004), and Australia (Peterson et al., 2005).

In contrast to Etel and Yagmurlu (2015), Selçuk et al. (2018) conducted another study with a Turkish sample from typical backgrounds. This study was conducted with 260 children from 5 different cities in Turkey. The clear majority of children followed the Eastern sequence. They passed the KA task (79%) before the DB task (64%) like their

Chinese and Iranian peers. In further analysis that aimed to investigate existing sequences within this Turkish sample, 161 children (62%) were eliminated from the data who matched the Eastern sequence. While 21 children (8%) showed the Western pattern, the remaining 15 children (6%) showed neither Eastern nor Western patterns. The sequential differences among Turkish children were connected to the household population. Further investigations indicated that children living with more adults at home were likely to pursue the Western sequence.

Ilgaz et al. (2022) implemented another study to reveal Turkish children's ToM scaling. Turkish children performed lower than their peers in other cultures (e.g., Australia, China, and the United States) especially in the belief-related ToM tasks (e.g., diverse belief and false belief tasks). This study yielded different sequences in terms of age and EF performance. With regards to EF abilities, children with lower-than-average EF performance followed the Western pattern. That is, children who were relatively unsuccessful in the EF task firstly passed the DD, then the DB, and lastly the KA task. Children with average EF performance showed the Eastern sequence previously described in the literature. That is, children found the DD task (82%) easier than the KA task (73%), and the KA task easier than the DB task (57%). Interestingly, children who had higher-than-average EF performance showed a third sequence that fit neither the Western nor the Eastern sequence. These children first passed the KA task before the DD, and the DD before the DB. The sequences observed according to children’s EF performance also matched the age profiles. In that, 3-year-olds followed the Western pattern while 4-year-olds had Eastern pattern. As for 5-year-olds, they first passed the KA (89%) task followed by the DD (77%) and the DB (63%), respectively.

To conclude, previous literature indicated that the difficulty levels of the tasks varied intra-and inter-culturally. The sequence of children who are from the United States and Australia was termed as Western sequence and these children passed the DB before the KA. Chinese and Iranian children's sequence was named as Eastern sequence. Regarding Turkish children, they have lower DB performance than their peers from other cultures, and the sequence changes depending on age and EF performance.

1.3.4. Explanations for Cross-Cultural Variations

The underlying reasons for sequential differences in ToM tasks both within and between cultures need to be answered. In this section, two relevant views proposing that these variations can be explained by cultural attitudes and environmental factors will be presented and the points of criticism regarding both arguments will be discussed.

1.3.4.1. Can sequential differences be explained by cultural attitudes?

Previous research emphasized differences in cultural attitudes as the reason for cross-cultural variations. Of these cultural variations, the most likely aspect was conceptualized to be the dichotomy between individualism and collectivism (Liu, Wellman, Tardif, & Sabbagh, 2008; Shahaeian et al., 2011; Liu et al., 2008). These concepts describe the principles and practices individuals have in order to fit into their social environment (Hofstede, 2011). People in individualistic cultures are prone to prioritize their own values and opinions rather than societal norms. Since the expression of personal values is appreciated in these types of cultures, people are likely to share their beliefs with others (Kabha & Berger, 2020). In contrast,

individuals from collectivistic societies look out for cultural values before their personal opinions. Sharing knowledge is more important than expressing one's own opinions in these kinds of populations. According to this view, children from Western cultures such as the US find the DB task easier than the KA task because they are culturally encouraged to express their thoughts regardless of whether their opinions are in line with cultural context. As for children in collectivist cultures such as China and Iran, young children may not be able to express their divergent beliefs so as to preserve harmony within the family, and therefore, they may find the DB task more difficult than the KA task when compared to their peers in other cultures (Shahaeian et al., 2011).

However, literature shows that collectivism and individualism do not explain cross-cultural variations in ToM performance assessed via the first-order, second-order false belief performance and the emotion inference from false beliefs (Hughes, Devine, Ensor, Koyasu, Mizokawa, & Lecce, 2014). Likewise, research that used the ToM scale did not find relations between cultural attitudes and ToM sequence on the scale. Kuntoro, Peterson, and Slaughter (2017) obtained parents' self-reported individualism and collectivism through their responses to several items (e.g., "Competition is important for my children." or "My children must attend every family gathering."). Though Indonesian parents were high in collectivism as compared to individualism in the continuum as expected due to the collectivistic attitudes in Indonesia, no meaningful influences of attitude dimension on sequencing were not found. In line with Indonesian children, Kabha and Berger (2020) studied Israeli-Arab Muslim children who were raised in a patriarchal and hierarchical cultural system that emphasizes collectivistic attitudes. However, they did not find any relation between parents' individualistic/collectivistic values and children's ToM sequencing as well.

Likewise, the acquisition of intra-cultural differences in studies with Turkish children (e.g., Ilgaz et al., 2022) may emphasize that this explanation remains insufficient to unclose the intriguing cultural variations.

Another criticism regarding individualism/collectivism is that each culture does not have to fit one of these categories. There may be some countries that have both individualistic and collectivistic structures such as Turkey (Selçuk et al., 2018). Turkish children who are raised in institutions showed the Western sequence whereas Turkish children living with their biological parents showed the Eastern sequence (Etel & Yagmurlu, 2015; Selçuk et al., 2018). Researchers thought that institution-reared children's ToM scaling might be due to the specific environment they live in. More precisely, some of these children had no interaction with their biological caregivers while others had limited interaction. However, they have more frequent than usual interactions with their peers in these institutions. Peer interactions might encourage children to appreciate the various beliefs and opinions. Therefore, they might find the DB task easier than the KA task. In light of Turkish children's sequential differences depending on age, EF, and environmental factors, it is not possible to support the claim that children who live in individualistic cultures show the one sequence while children from collectivist cultures demonstrate the other one.

1.3.4.2. Is it possible that family factors instead of broad cultural contrast account for cross-cultural variations?

Researchers also suggested that the cross-cultural variations in the sequencing might be related to the environmental and familial dynamics rather than the broad cultural context (e.g., Shahaeian, Peterson, Slaughter, & Wellman, 2011). Although a particular culture may promote highly collectivistic attitudes, families may or may not

conform to the same level of tight-knit unity that is encouraged by collectivism (Kuntoro et al., 2017). There are a limited number of studies investigating the effects of familial and environmental factors on ToM sequence. In this subsection, the possible familial predictors of ToM sequencing such as SES, the number of siblings, and the household number will be discussed.

Selçuk et al. (2018) examined the demographic factors that might be associated with the typical Turkish children's sequence of ToM acquisition. They assessed the socio-economic status and the number of adults at home in light of the previous findings showing that higher family size was associated with higher ToM competence (e.g., Cutting & Dunn, 1999). Higher socio-economic conditions were related to fewer adults at home. Further, children whose households had higher adult populations followed the Western pattern whereas the Eastern pattern was observed in smaller families. Leblanc, Bernier, and Howe (2017) revealed that having an older sibling predicted higher ToM abilities as assessed by non-verbal perspective taking and diverse desires tasks (Flavell, Everett, Croft, & Flavell, 1981; Wellman & Woolley, 1990) whereas having a younger sibling did not support ToM development. However, Selçuk et al. (2018) found no correlation between the ToM score obtained on the ToM scale and neither the number of younger or older siblings nor SES when age was controlled. With the intent of investigating whether there is a sibling effect in the sequencing, Ilgaz et al. (2022) included SES and the number of siblings to investigate whether environmental variables affect ToM scaling. Interestingly, the DB task performance was positively correlated with SES, but it was negatively correlated with the number of siblings. Similarly, Doenyas et al. (2018) found no significant relation between the siblings, number of adults, and SES on individual ToM tasks.

It was argued that the existence of sequential differences in cross-cultural research might be related to environmental and familial structures. However, studies investigating the predictive role of family interactions and experiences on the cultural differences in ToM sequence were very limited. There is a wide range of research conducted to examine the familial factors and their effects on ToM development, but they produced inconsistent findings. Therefore, it is possible to claim that environmental factors may not create solid foundations to explain cultural variances as the hypothesis regarding cultural attitudes. To conclude, two different hypotheses were discussed in the literature to explain cross-cultural variations in ToM. However, there are points of criticism in both hypotheses which make them insufficient to account for differences in ToM sequencing.

1.4. Age-Related Transitions in the DB Task Performance

Previous studies found that ToM is universal in some respects. For instance, children are expected to show increasing competence for each ToM task by age. A meta-analysis showed that older children from both Western and Eastern cultures become more successful than younger children in the FB tasks (Wellman et al., 2001). Looking at Turkish children, ToM mastery was related to age in two scaling studies (Etel & Yagmurlu, 2015; Selçuk et al., 2018). The increase in performance was attributed to the development of children's other cognitive abilities related to theory of mind understanding such as executive functioning (EF) and language skills. However, recent studies yielded unexpected results on Turkish children's age-related individual task performance contrary to previous findings.

Ilgaz et al. (2022) conducted a study to investigate the factors affecting Turkish children's ToM performance. Age-related variations in ToM task performances were

analyzed. Children's performance demonstrated expected developmental trajectories in the KA, CFB, and HE tasks. In other words, their performance on these tasks increased with age. However, their performance on the DD, DB, and EFB tasks did not meaningfully change with age. While 40% of children who were at the age of 4 passed the DB task, 35% of both 3-year- and 5-year-olds passed the DB task. In a recent study, Evsen (2022) studied 74 Turkish-speaking children aged between 3 and 5. In parallel with Haskaraca (2019), the DB performance did not significantly increase by age (68% for 3-year-olds, 63% for 4-year-olds, and 63.6% for 5-year-olds).

It was argued in the literature that one of the problems that need to be answered is cross-cultural variations in the developmental progression of the ToM tasks. One of the most striking findings about the cross-cultural variations was the lower performance of Turkish children on the DB task than their counterparts in other cultures. It is not surprising that they showed lower performance than their Western peers due to a sequential difference. However, it is noteworthy that although Turkish children followed the same sequence as Chinese children, they showed lower performance than their Chinese peers (71% in Wellman et al., 2006) as well (71% in Etel & Yagmurlu, 2015, 57% in Haskaraca, 2019, and 64% in Selçuk et al., 2018). The fact remains that Turkish children did not show an age-related performance increase on the DB task unlike the other ToM tasks in the battery. These two problems essentially suggest that the DB task may be problematic for Turkish children. Put differently, it could be valuable to question the internal validity of the DB task for Turkish children because it might be possible that the task does not measure what it is intended to measure (i.e., appreciating the diversity in beliefs) for the Turkish sample.

1.5. DB Task, Its Relations with Other ToM Tasks, and Child Variables

Wellman and Liu (2004) created a ToM battery that comprises a package of tasks tapping into different categories of mental states. The attempt was to keep the tasks similar (e.g., in terms of length) because the aim was to ensure internal consistency within ToM tasks (Wellman & Liu, 2004). However, it seems impossible to eliminate task-related cognitive requirements for children to pass the tasks. At least, it seems plausible to think that all tasks require some EF abilities such as children's ability to appreciate a different belief that does not belong to themselves in the DB task (Ilgaz et al., 2022). Therefore, there must be a relation between ToM tasks and EF skills. However, recent studies conducting with Turkish children showed that the DB task is not related to all other tasks in the battery and child cognitive variables such as EF. There are a few studies that investigate the correlations between individual ToM tasks and child variables for Turkish children. Ilgaz and Burumlu-Kisa (2021) attempted to assess whether ToM tasks were related to each other. The DD, DB, KA, two appearance-reality, two first-order, and one second-order false belief task were applied to children. It was found that the diversity-related tasks (DD and DB) were only correlated with each other (r = .18, p < .05). The remaining items were related to each other, but they were not related to the DD and DB tasks. When age controlled, Ilgaz et al. (2022) also yielded the correlations between the DB and the DD tasks (r = .23, p < .01) and Evsen (2022) obtained that the DB task was only related to the DD (r = .33, p = .004) and the EFB (r = .33, p = .005) tasks. With regards to the relations between ToM tasks and child variables, previous literature emphasized the correlations between ToM growth and EF abilities beyond age, gender (Müller et al., 2012), and culture (e.g., Carlson, Moses, & Breton, 2002; Shahaeian et al., 2015).

Yet, Doenyas et al. (2018) found that the DB task individually is not correlated with EF abilities even though the FB tasks were significantly related to EF.

In short, previous literature showed that Turkish children have problems in the DB task in terms of cross-cultural variations, and age-related changes in performance. In addition to these problems, the DB task was not related to the other tasks in the battery and the other variables such as EF. Turkish children's relatively lower success rates in the DB task compared to their peers in other cultures were explained by cultural or familial factors (Etel & Yagmurlu, 2015; Selçuk et al., 2018) and it was suggested that Turkish children may be delayed in conceptual belief understanding that make them pass the DB task after the KA task. However, neither of these explanations established a solid foundation for cross-cultural varieties. In recent studies, the unexpected age-related change on the DB performance and its lack of relation with the other tasks and EF raised the question of whether Turkish children's underperformance on the DB task may be related to the structure of the task rather than Turkish children's conceptual deficiency. To that end, the current study aimed to investigate the DB task in detail by making task modifications depending on the possible epistemic differences in Turkish mental state repertoire and a bias similar to the 'curse of knowledge' that is typically observed in the FB tasks (e.g., Ghrear, 2021).

1.6. Possible Effects of the Epistemic Differences in MS Verbs for DB Performance

It is one of the most consistent findings in the literature that there is a relationship between language and theory of mind (Carpendale & Lewis, 2004). All languages contain the components of semantics, syntax, and pragmatics. Studies indicated that these are individually related to ToM competence (Milligan, 2007). However, each

language may have unique characteristics that can influence social understanding (i.e., theory of mind) in different ways as well. Some linguistic structures in Turkish express certainty at the morphosyntactic level, unlike in English. This emphasis is made by syntactic (e.g., *-DI* and *-mIş)* or by semantic (e.g., mental state words) nuances. In this section, the syntactic and semantic specifics of the Turkish language and its relations with Turkish children's ToM development will be presented.

One of the culture-specific linguistic features that are closely related to ToM understanding is the concept of *evidentiality* (Papafragou, Li, Choi, & Han, 2007). Evidentiality explicitly indicates whether an expression is obtained first-hand or by evidence (Aksu-Koç, 1988). There are two commonly used modes of knowledge acquisition in Turkish evidentiality: direct experience and indirect experience. If one has direct access to knowledge through perception, the suffix *-DI* is used: "Uçağı kaçır-dı." ("[I saw/know that] she missed the plane"). If the knowledge access occurs through inference instead of perception, this means that one has an indirect experience. In these circumstances, - *mIş* marker is used: "Uçağı kaçır-mış." ([I heard/inferred that] she missed the plane"). These evidential markers are used to highlight the source of knowledge and consequently the speaker's level of certainty about the truth value of the statement. Therefore, the source of knowledge and how confident the person is of the truth value of the statement are common and inescapable factors in consideration for Turkish speakers when they relay information (Aksu-Koç et al., 2009). Regarding the relations between ToM and evidentiality, Özoran (2009) found that Turkish children performed better when ToM stories were presented with *-DI* (e.g., "Ayşe kedinin çalılıkların altında olduğunu düşün-üyor-du" / "Ayşe was thinking that the cat is behind the bushes") than the neutral version (e.g.,

“Ayşe kedinin çalılıkların arkasında olduğunu düşün-üyor” / “Ayşe thinks that the cat is behind the bushes”).

Along with syntactic emphasis, certainty can be explicitly indicated by semantics (e.g., mental state verbs) as well. In Wellman and Liu’s (2004) study, the verb "*think*" was used in stories of the belief-related tasks (e.g., "Where do you think the cat is?" in the DB task). Subsequently, researchers conducting studies with English-speaking children kept the original wording of the tasks. In an early study, Moore (1989) studied English-speaking children’s understanding of relative certainty in mental terms (age range: 3 to 8). Although children started to appreciate the “*guess-know”* and “*think-know”* pairs at the age of 4, the conceptual nuance between “*think”* and “*guess”* was not yet developed. Statements with the verb “*know*” indicates the accuracy of the expression while they are likely to interpret the statements with the verbs “*think*” and “*guess*” as untrue. In consideration of Moore’s findings about English-speaking children in the 1989 paper, it is a mystery how Turkish children conceptualize mental state terms and how this conceptualization shapes their ToM understanding.

With the exception of Haskaraca (2019), Turkish children were presented with the direct translation of the verb “*think*” (düşün- in Turkish) in other studies (e.g., Etel & Yagmurlu, 2015; Selcuk et al., 2018). However, it was thought that Turkish children’s conceptualization of belief-related mental state verbs may affect their performance on the DB task. Due to the highly grammaticalized structure of Turkish around the explicit demonstration of the source of knowledge, Turkish children may be more sensitive to the different mental state verbs that carry different levels of certainty. Therefore, Turkish children’s relatively lower performances on the belief-related tasks

led researchers to think that the use of direct translations of English verbs in ToM tasks might be the reason for Turkish children's difficulty in belief-related DB task. However, young children's cultural conceptualization of the epistemic differences among mental state verbs is an important topic that was studied limited.

Haskaraca (2019) studied whether the morphosyntactic structure of Turkish language affects Turkish children's performances on ToM tasks. For this purpose, she investigated the direction of children's success in ToM tasks when belief-related sentences or questions were expressed with different levels of certainty. Firstly, Haskaraca (2019) conducted a study with Turkish-speaking adults to discover their judgments about the verbs "*think*" and "*guess*". Turkish adults found the verb "*guess*" more appropriate to use than "*think*" in case the accuracy of the belief is unknown (i.e., diverse belief task) because Turkish speakers may be likely to use "*think*" when there is an explicit reason to hold a particular belief about an object. However, when the DB task was both presented to children with "*think*" and "*guess*" (e.g., "Dilek guesses that the cat is under the car."), no meaningful performance difference was found between DB-Original (36.7%) and DB-*Guess* (36.7%). In other words, the task performance of Turkish children did not change depending on which verb the task was presented with. However, when evidentiality was provided in another way (paw prints were added to both locations (i.e., car and bushes)) shown in the original DB task, children performed significantly better (56.7%) than the original version which is without paw prints as evidence (36.7%). Along with mental states, Turkish language has room for expressing mental activities with certain types of phrases that specify perspectives such as "*-a göre*" ("in my/your opinion" in English) and "*-ce*" ("according to" in English). Balcı (2021) conducted a study with Turkish adults to see if there was a difference for participants in certainty judgments between "*ben-ce*" and

"*düşün*-" ("think" in English). There was no difference between the verb "*düşün*-" and the phrase "*ben-ce*" in terms of certainty judgments. In this study, DB tasks were presented with "*think*" (düşün-), "*guess*" (tahmin et-), and "*in my opinion*" ("ben-ce"). The difference was not found between the verb "*think*" and "*guess*", but the replicability of the findings is important. Besides, the lack of difference in certainty judgments of adults between "*think*" and "*in my opinion*" should be examined for children as well.

1.7. Possible Effects of the Curse of Belief on Turkish Children's Underperformance on the DB Task

Previous research demonstrated that children's failures in the ToM tasks were because of the lack of a conceptual understanding (Wellman et al., 2001), or the unadvanced cognitive mechanisms such as memory and executive functioning abilities in preschool years (Bloom & German, 2000). From a cognitive constraint view, Birch and Bloom (2003) presented the phenomenon of the *curse of knowledge*. The curse of knowledge or *reality bias* is defined as a person's claim to have known the truth from the very beginning (Mitchell & Taylor, 1999). Previous studies revealed that the curse of knowledge may lead children to experience difficulties passing the FB tasks (Birch & Bloom, 2007). The current study aims to investigate whether Turkish children's underperformance on the DB task may be explained by a similar mechanism we dubbed the *curse of belief*.

People are susceptible to having biases while inferring other agents' knowledge or beliefs (Birch & Bloom, 2007). For instance, when people have knowledge or a belief about an event, they tend to overestimate the other party's knowledge or belief about the event. Previous research asserted that younger children's failures in the false

belief tasks might be explained by the curse of knowledge. Children should respond to the target question according to the character's false belief (e.g., basket), not to the current location (e.g., box) of the object (e.g., marble) (Wellman et al., 2001). However, children's answers to the target question in line with the current location of the object might be because they were under the influence of their own knowledge (Birch, 2005). To test this claim, children were presented with 12 toys. It was told that either the toy was familiar or unfamiliar to the experimenter's puppet friend. Children were instructed that there was an object in each toy. They were either shown the content of the toy or not. In line with researchers' expectations, results revealed that children were prone to be biased about the puppet's knowledge when they knew the toy's content. In other words, they tended to overestimate the puppet's state of knowledge when they were shown the object in the toy.

There are three underlying explanations leading to the curse of knowledge for children (Ghrear, 2021). The *inhibitory control approach* claims that people's inability to suppress or inhibit their own perspectives when reasoning about a less knowledgeable person leads to the occurrence of this bias (Bayen, Pohl, Erdfelder, & Auer, 2007). Inhibitory control as a component of EF abilities develops around preschool years and therefore, the influence of the curse of knowledge is expected in young children more (Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011). The other mechanism proposed by researchers is the *fluency misattribution account*. It is suggested that when the information is easier to process or recall, this means that the fluency of that information is high (Bernstein, Kumar, Masson, & Levitin, 2018). Individuals are prone to assume that the information with high fluency is also easy to process or recall for others. More clearly, if one knows that the marble is in the box instead of the basket, he/she should in a sense overlook the fluency of their own

knowledge to be able to make correct inferences about others' minds. Researchers found evidence that supports the fluency misattribution account on the curse of knowledge although the inhibitory control account does not solitarily explain this phenomenon (Ghrear, 2021). It is because it was revealed that the curse of knowledge did not show up every time but was affected by the familiarity of information. These findings may imply that along with the well-known role of the EF (e.g., inhibitory control) on social cognition in the literature (e.g., Carlson et al., 2002), the impact of the fluency misattribution is noteworthy (Geurten, Lloyd, & Willems, 2017) as well. More to the point, researchers discussed the sense-making mechanism that can also be evaluated under the fluency misattribution account. According to this mechanism, children are able to make sense of particular knowledge, they are likely to think that knowledge is highly apparent to another one (Bernstein et al., 2018).

In the light of the underlying mechanisms, there are certain reasons to discuss the curse of knowledge in this thesis investigating Turkish children's DB task performance. Unlike the false belief tasks, there is no knowledge state in the DB task. However, just as younger children are susceptible to being biased by their knowledge states in the FB tasks, it might be plausible to assume that a child's own belief may be more salient than the protagonist's belief in the DB task. Therefore, the curse of knowledge in essence turns into the *curse of belief* that indicates children's commitment to their own beliefs. It is predicted that children who overestimate their own knowledge states may overestimate their own beliefs and ignore someone else's perspective as well. Therefore, when children are first asked to voice their beliefs in the DB task ("What do you think where the cat is?"), they may overvalue their own beliefs (maybe because they make sense of the belief that was produced by themselves) and answer the target question in line with their own beliefs, not with the

protagonist's belief. To conclude, in the original DB task, the child's own perspective is taken as one. The failure in the DB task may be either due to children's inability to consider two perspectives regardless of who they belong to or children's primacy of their own perspectives. It is suggested that eliminating children's own perspectives may untangle the reason for Turkish children's underperformance on the DB task.

1.8. Current Study

Wellman and Liu (2004) developed a battery to measure different mental state understandings. Studies conducted with children from different cultures raised many questions in the literature that need to be investigated. First of all, there are cultural differences in ToM scaling. Children living in Western cultures pass the KA task before the DB task while children in Eastern cultures show the reverse pattern. More importantly, scaling studies with Turkish children partially revealed that Turkish children underperform on the DB task compared to their peers in other cultures.

In addition to the gap in the literature on cross-cultural variations, Turkish children showed unexpected age-related changes in ToM task performances. Despite children in other cultures' increasing ToM performance by age, recent studies (e.g., Ilgaz et al., 2022) revealed that Turkish children did not show a consistent improvement in the DB task performance as they get older. Together with these two issues, another motivation to focus on Turkish children's DB task performance is the DB task's relation with other tasks in ToM battery and with other cognitive variables such as EF.

Therefore, the current study concentrates on a detailed investigation of the Turkish children's DB task performance. The research question is whether Turkish children

have difficulty passing the DB task as it is framed in Wellman and Liu's (2004) original task although they are able to understand the concept of belief diversity. To answer this question, the DB task was manipulated in two different aspects. First, the DB task was presented with the verb “guess” and the phrase “in my opinion/according to”. Second, the aim was to investigate whether the original version of the DB task encourages children to be biased by their beliefs. To that end, a different DB task was created in which participant children were not asked to state their own beliefs. Instead, another protagonist was included in the story whose belief contradicts the other protagonist in the story.

Regarding the differences in mental state lexicon, it is expected that no mental state verb condition would be easier for Turkish children than the verb think and guess. For the number of characters, it is predicted that children would be more successful when a character’s opinion (rather than the child’s own) is pitted against another story character’s belief.

The current study intends to contribute to the literature on cultural ToM understanding. If the DB task modifications lead to meaningful changes on performance, the underlying reason for Turkish children’s underperformance on the DB task may be attributed to the internal validity of the task instead of Turkish children’s deficient understanding of belief.

CHAPTER II

METHODS

2.1. Participants

This study recruited 45 Turkish-speaking pre-schoolers. Since this study was conducted online, participants were living in different cities in Turkey. 25 participants were living in Ankara. Other participants were coming from Balıkesir (5), İzmir (3), Bursa (3), Adana (2), Eskişehir (2), Amasya (1) Erzurum (1), İstanbul (1), Kırklareli (1), and Trabzon (1). Children's age ranged from 37 to 67 months (M = 54.13, SD = 8.66), 15 for 3-year-olds, 14 for 4-year-olds, and 15 for 5-year-olds. Their mothers' age ranged from 23 to 47 years (M = 34.91, SD = 4.19) and their fathers were between 30 and 53 years (M = 37.84, SD = 5.05). Mothers' educational status ranged from high school (15.6%) to PhD (6.7%) while fathers' education ranged from elementary school (2.2%) to PhD (6.7%). The majority of the mothers and fathers graduated from university (57.4% and 46.8%, respectively). The percentage of parents who held a full-time job was 48.9% for mothers and 91.1% for fathers. In addition, 11.1% of mothers reported working from home. Parents were not asked how much they earn monthly, but we asked about their perceptions of their income levels. Two parents

reported that they perceived their income levels as having a very good level, while 23 parents stated that they perceived their income levels at a good level and 20 parents at an average level. Based on the parents' reports of their occupational status and their salary perception, it would be possible to claim that the participants recruited in this study were coming from middle or upper-middle socioeconomic backgrounds.

All of the participants were recruited from social media platforms and through the snowball sampling technique. It is a technique in which parents whose children already participated in the study recommend the study to their relatives, friends, or colleagues who meet the criteria for participation (Johnson, 2014). However, the tester had no previous contact with any of the children participating in the study. Children were tested online through the Zoom platform after their parents signed the consent form. Consent forms and demographic forms were again collected online through Google Forms. To protect the privacy of the participants, no one other than the lead director of the study had access to Google Forms. Verbal consent was obtained from each parent and the child at the beginning of the testing session. All materials and procedures were approved by the Bilkent University Ethics Committee and the Ministry of Education.

2.2. Materials

2.2.1. Demographic Information Questionnaire

This form consists of questions about parents' demographic information (i.e., age and education), child's birthdate, number of children in the household, and family income.

2.2.2. Diverse Belief (DB) Tasks

The current study focused on the DB task created by Wellman and Liu (2004) as a task in ToM battery. Five different versions were produced that included variations on the mental state verbs and the number of characters in the task. Each task was presented with pictures and characters were matched with the participant children's gender. Below are the descriptions of the original DB task created by Wellman and Liu (2004) and the modified versions produced for this study (see Appendix for more details).

2.2.2.1. Single Seeker – Think (Original Version)

This task was presented to measure children's ability to understand that two persons (e.g., the participating child and the character in the story) may have different beliefs about the location of the same object when these two persons have no clue about the placement of the target object. Firstly, the experimenter showed the picture of the protagonist with two different pictures of locations (e.g., an armchair and a bed) to the participant child. Then, the experimenter told to the participant child: *"This is Dilek/Mert. Dilek/Mert wants to find the cat. The cat is either behind the armchair or under the bed. Where do you think the cat is, behind the seat or under the bed?"* This question was asked to learn the participant child's own belief about where the cat is. When the child made a guess about the location of the cat (e.g., behind the armchair), the experimenter stated that the character in the story thinks that the cat is in the other place which was not chosen by the child: *"It is a good idea, but Dilek/Mert thinks the cat is under the bed*." Afterward, the child was asked to answer the target question: *"Where will Dilek/Mert look for the cat, behind the seat or under the bed?".* Children

received 1 point if they answered the target question in accordance with the story character's belief as opposed to their own belief.

2.2.2.2. Single Seeker - Guess

The narratives of the original task remained the same but the character's name, the lost animal, and the locations were changed. The important difference between the original task and this version is the verb used in the story. As different from the original task, child's own belief was asked with *guess* (i.e., *"Where do you guess the dog is, in the closet or under the table?*") and the target question was the same as the original version (i.e., *"Where will Zehra/Tamer look for the dog, in the closet or under the table?*").

2.2.2.3. Single Seeker - No Mental State

This task was the same as the two tasks described above. The character's name, the lost animal, and the locations are again different. In addition, unlike the other two tasks, no mental state verbs such as "*think*", or "*guess*" were used in this task. Instead of mental state verbs, child's own belief question was directed by using the adverbs (i.e., *"In your opinion, where is the chick?")* and the character's belief was expressed by using the adverbials (i.e., *"According to Oya/Ali, the chick is in the tub/behind the towels."*). target question was the same with the other tasks (i.e., *"Where will Oya/Ali look for the chick, in the tub or behind the towels?*").

2.2.2.4. Double Seeker - Think

The original DB task presented to children was first modified according to the type of mental state verbs as explained above. The other modification was related to the number of characters in the story, which somehow alters the content of the story. In

the 3 tasks above, the child was expected to first come up with an answer and then understand the character's own position (e.g., belief, guess, or opinion) that contradicts the child's own answer. In the "number of seekers" modification, the part where children are asked about his/her position at the beginning of the story was eliminated. Instead, another character that holds a conflicting belief with the target character was included in the story. First of all, participants were shown the target character: *"Look, this is Burcu/Burak. Burcu/Burak wants to find the caterpillar.* Then, the other story character was shown: *"Look, this is Neşe/Caner. Neşe/Caner also wants to find the caterpillar".* Then, the experimenter presented the first location: *"Neşe/Caner thinks the caterpillar is behind the flowers.* Lastly, the experimenter showed the second location to the child: *"However, Burcu/Burak thinks the caterpillar is behind the stones."* The target question is: *"Where will Burcu/Burak look to find a caterpillar, behind the flowers or behind the stones?"* Children who said that Neşe/Caner would search for the caterpillar behind the stones were considered successful and received 1 point.

2.2.2.5. Double Seeker - Guess

This modified task was the same as the previous task with one exception which was the type of mental state verb. The animal, characters' names, and locations were changed. In this instance, the beliefs of both characters and the target question were conveyed with the verb *guess* instead of *think* (i.e., *"Ayça/Özlem guesses that the bird is in the tree/behind the roof*."). The target question asks about the action of one of the characters depending on his/her belief (i.e., "*Where will Özlem look for the bird, in the tree or behind the roof?*").

2.2.2.6. Double Seeker - No Mental State

In this task, the content of the story was the same as in the previous double-seeker tasks. The animal, characters' names, and locations were altered. The only difference was that the characters' beliefs were presented with phrases including no mental states (i.e., "*According to Sevgi/Leyla, the turtle is inside the wood block/behind the fence.*"). The target question asks about the action of one of the characters depending on his/her belief (i.e., "*Where will Leyla look for the turtle, inside the wood block or behind the fence?*").

2.2.3. Day-Night Task

This task was designed to measure children's interference control abilities (Gerstadt et al., 1994). Originally, there are two cards in this task: the picture of the sun and the moon with stars. The task started with the two training trials. The training trials aim to teach the child about the rules of the task. Children are expected to say "*day*" when they are presented with a picture of the sun whereas they need to say "*night*" when they show a picture of the moon. Since the current study was conducted online, we used PowerPoint for testing sessions. Therefore, we used slides instead of cards as different from the original version of the task. We used the fixed order of the original version of the task and the experimenter passed to the next cards by pressing the button on the computer. After the child gives his/her answer verbally, the experimenter shows the next card. Since the task was applied on the computer, a white blank page was placed between each trial card to prevent the children from being confused when they see the same card over and over. Based on the progress and implementation of the original task, children's "sun" and "moon" answers were also evaluated as correct. The word "day" in Turkish has two exact equivalents: "sabah"

and "gündüz". Both answers were qualified as correct. Likewise, the word "night" can be translated into Turkish in two ways: "akşam" and "gece". These answers were also regarded as correct. Each correct answer is worth 1 point if the child gives a correct response for the target item whereas each wrong answer is worth 0 points. Each child can get a maximum of 16 points (two practice trials and fourteen test trials).

2.3. Procedure

Families were contacted via WhatsApp or e-mail and two forms were sent to them. One of these forms was the demographic form and the other was the consent form. After explaining why these forms were sent to the families, they were asked to fill out. A suitable day and time were determined with the families who filled out the forms. One day before the testing session, families were reminded of the session and the Zoom link was sent. At the beginning of the testing session, the parents were asked to tell their own name and the name of their child in order to verbally reaffirm that their child participated in this study. After this verbal consent was obtained, the child's verbal assent was taken, and the tasks were applied. Each child was tested once, and the testing session lasted an average of 30 minutes.

The testing session started with a warm-up game. In this game, children needed to know the animal sounds in which children try to know the sounds of animals, while children saw the pictures of animals on the screen. A filler game that may potentially make a buffer between two tasks was played after each task, as the applied diverse belief tasks had similar characteristics in essence. These filler games were given to each child in the same order and the order was as follows: day-night task (EF measurement), “zoom-in”, “find the differences”, “what is missing”, and “painting”. The day-night task, which is an executive functioning measurement, was also

implemented as a filler task. The day-night task was positioned right after the warm-up game and then the first diverse belief task presented to make children respond with as clear a mind as possible. A video from YouTube was played for the "zoom-in" game. In this video, the pictures were first shown very close up, and then slowly they zoomed out. In the "finding the differences" game, children tried to find 5 differences between two pictures. In the 'what is the missing?' game, children were first shown 6 objects and the objects are as follows: a book, a pencil, an apple, a balloon, a hamburger, and a car. After confirming that the children recognized all of these objects, they were asked to count to 3, and then one of the objects disappeared. The child was asked which object was lost and this process was repeated three times. The last filler task was the painting game. A website was opened, and the child was asked which picture he/she wanted to paint among 12 pictures. Then, children were asked to direct the tester to complete the painting of the picture on the computer.

The distribution of the tasks was counterbalanced before the data collection. Three tasks in one seeker condition (think, guess, and no mental state) were handled as the first group and three tasks in double seeker condition (think, guess, and no mental state) were handled as the second group. These groups were counterbalanced within themselves. For example, if children were presented with SS-Guess, SS-NoMS, and SS-Think, they then were presented with DS-Guess, DS-NoMS, and DS-Think. The first group (i.e., single seeker tasks) always came before the second group (i.e., double seeker tasks). The ordering of the single-seeker tasks was also applied to the double-seeker tasks. In the double seeker condition, one more counterbalancing was applied in addition to the shifting in the task orders. Children were always asked where the one character from two characters would look for the animal in the story. Namely, children were always expected to respond according to the one character in the task.

The place of the correct response on the page was counterbalanced to eliminate the possibility of children taking the direction as a cue to reach the correct response. Therefore, both left and right conditions were created in double-seeker tasks. While the correct response is on the left in the left condition, it is on the right in the other condition (see Table 1 for detailed descriptions of the counterbalancing). It should be noted that all combinations in the counterbalancing were not applied because of the sample size. At the end of the testing session, children were asked to applaud himself/herself for completing the games. Since most of the participants were collected with the snowball sampling method, debriefing forms were sent to the parents via e-mail after the whole data collection process was completed.

Table 1. Counterbalancing Scheme

	Group 1 (Single Seeker)			Group 2 (Double Seeker)		
1	Think	Guess	No MS	Think (R)	Guess (L)	No MS (R)
2	Think	Guess	NoMS	Think (L)	Guess (R)	No MS (L)
3	Guess	Think	NoMS	Guess (R)	Think (L)	No MS (R)
4	Guess	Think	NoMS	Guess (L)	Think (R)	No MS (L)
5	Guess	NoMS	Think	Guess (R)	No MS (L)	Think (R)

Table 1 (cont'd)

6	Guess	NoMS	Think	Guess (L)	No MS (R)	Think (L)
7	Think	NoMS	Guess	Think (R)	No MS (L)	Guess (R)
8	Think	NoMS	Guess	Think (L)	No MS (R)	Guess (L)

CHAPTER III

RESULTS

The main research question is whether Turkish children have difficulty passing DB task as it is framed in Wellman and Liu's (2004) original task although they are able to understand the concept of diversity in belief. There were four hypotheses regarding this question. First, children will be more successful in double seeker conditions than in one seeker conditions, regardless of the mental state verb used in the story. Second, task success will not significantly change according to the mental state verbs used. Third, regardless of the task, older age groups will show better task performance than younger age groups cross-sectionally. Fourth, task performance in each task will be positively correlated with the EF performance.

Results will be explained in three sections. Preliminary findings and McNemar's analyses will be presented in order to see whether children benefited from task modifications. Guttman analyses will be reported to examine the sequencing of the tasks according to age and in total. Finally, point biserial correlations between day-night and individual DB task performance and one-way ANOVA to assess whether

there is an age-related development in EF task were performed.

3.1. The Effects of Modifications on Task Performance

The percentages of success of children in each task depending on age and in total were investigated (see Table 2). An increasing developmental trajectory was expected in which each age group performed better than the younger age groups. However, it is shown that this is not the case for any task including the original version (i.e., SS-*Think*). In the original task of Wellman and Liu (2004), while the highest performance belongs to 3-year-old children, 5-year-old children show the lowest performance. While ever-growing performance was expected in other modified tasks, younger age groups seem to outperfom older ones. Children performed better in the DS-*Think* condition than the SS-*Think* condition both for 3- and 5-year-old children. Indeed, Figure 1 describes the percentages children who pass both SS-*Think* and DS-*Think*, pass SS-*Think* but fail DS-*Think*, fail SS-*Think* but pass DS-*Think*, fail both SS-*Think* and DS-*Think*. 3-year-old (40%) and 5-year-old (50%) children were successful in the DS-*Think* task although they failed in the SS-*Think* task, but similar comparisons were not found in the SS-*Guess* vs. DS-*Guess* and the SS-*NoMS* and DS-*NoMS*.

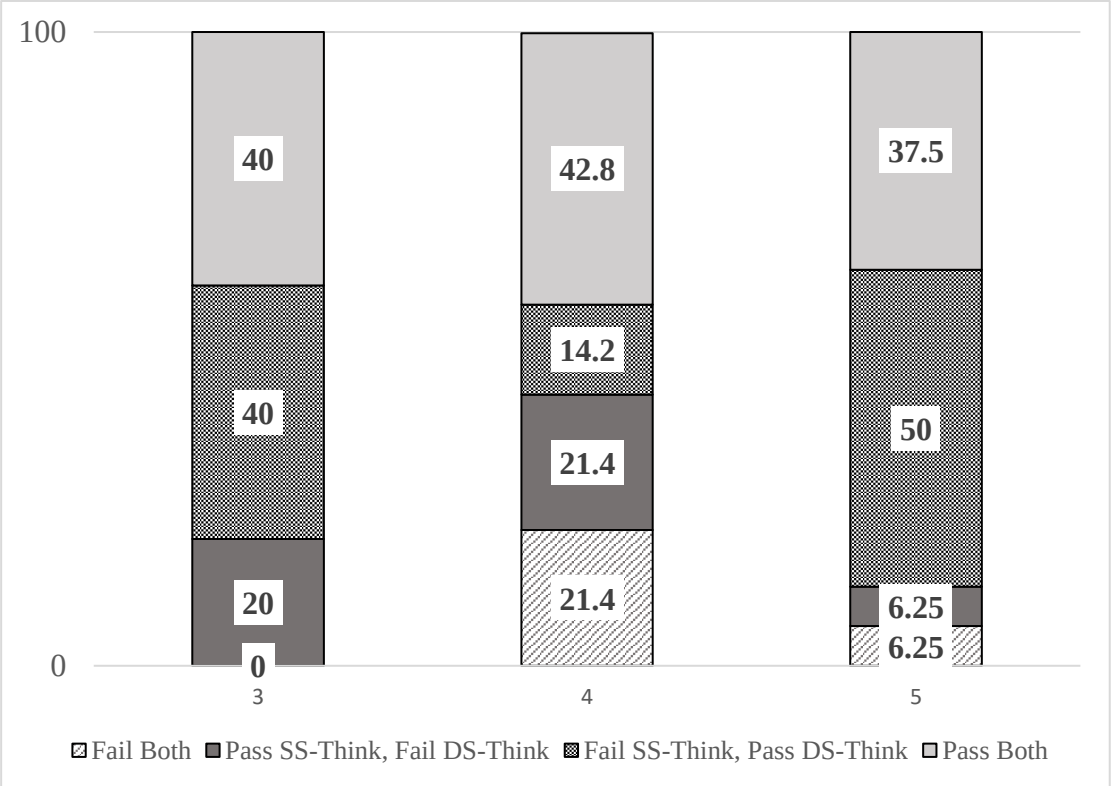
Table 2. Children's Task Performances by Age and in Total

	3 years (n = 15)	4 years (n = 14)	5 years (n = 16)	Total
SS-*Think*	73%	64%	44%	60%
SS-*Guess*	53%	36%	50%	47%

Table 2 (cont'd)

SS-*NoMS*	47%	71%	44%	53%
DS-*Think*	80%	57%	88%	76%
DS-*Guess*	60%	64%	25%	49%
DS-*NoMS*	73%	57%	63%	64%

Figure 1. Percentages of SS-*Think* and DS-*Think* Performance by Age



McNemar test analysis was conducted to investigate if children show performance improvement on modified tasks when compared to the original version. The McNemar test is a kind of chi-square method used to analyze categorical variables

collected from repeated or matched samples (Adedokun & Burgess, 2012). In order to assess whether children's performance significantly differed in the modified tasks as compared to the original task, five McNemar tests were conducted (i.e., SS-*Think* vs. SS-*Guess*, SS-*Think* vs. SS-*NoMS*, SS-*Think* vs. DS-*Think*, SS-*Think* vs. DS-*Guess*, and SS-*Think* vs. DS-*NoMS*). None of these tests yielded significant results. In order to examine whether the double versus the single seeker conditions were easier for children, two additional McNemar tests were implemented (SS-*Guess* vs. DS-*Guess*, and SS-*NoMS* vs. DS-*NoMS*). There were no significant results in these tests either (see Table 3).

Table 3. Significance Levels of the Task Pairs in McNemar Analyses

McNemar Pairs	*p*-values
SS-*Think* vs. SS-*Guess*	.146
SS-*Think* vs. SS-*NoMS*	.607
SS-*Think* vs. DS-*Think*	.189
SS-*Think* vs. DS-*Guess*	.405
SS-*Think* vs. DS-*NoMS*	.824
SS-*Guess* vs. DS-*Guess*	1.000
SS-*NoMS* vs. DS-*NoMS*	.405

Children's answers were coded as 0 (incorrect response) and 1 (correct response) in all tasks. Therefore, we investigated whether children's performance was different than chance in each task. The binomial test is used to see whether the observations are different from expectations if the expected value is known beforehand (50% for the current study because children received either 0 or 1) (Singh et al. 2013). Based on this, binomial analyses were performed separately for each task and age group. Children's performance was at chance level for all tasks, for all ages, as well as the age collapsed values (see Table 4). The only exception was the DS-*Think* task. Both 3-year-olds (p = .035) and 5-year-olds (p = .004) were more successful than chance. Interestingly 4-year-olds performance remained at chance level. The age collapsed values for this task also revealed a significantly better than chance performance (p = .001).

Table 4. Significance Levels of the Binomial Tests

	3 years	4 years	5 years	Total
SS-*Think*	.118	.424	.804	.233
SS-*Guess*	1.000	.424	1.000	.766
SS-*NoMS*	1.000	.180	.804	.766
DS-*Think*	.035*	.791	.004**	.001*
DS-*Guess*	.607	.424	.077	1.000
DS-*NoMS*	.118	.791	.454	.072†

†p < .10, * p < .05, ** p < .01, ***p < .001

3.2. Sequencing Analysis: Six-Item Guttman Scale

This study also aimed to understand whether the six DB task presented to children create a meaningful pattern from easiest to hardest. For this purpose, Guttman's scaling method was performed to obtain the sequencing of these tasks for Turkish children. The reason for conducting the Guttman analysis is that it has been commonly used in studies that target to create a sequence for ToM tasks (Etel & Yagmurlu, 2015; Shahaeian et al., 2011; Wellman & Liu, 2004) and it enables to demonstrate the progressive difficulty across tasks.

Green (1956) described Guttman's scalogram analysis. According to this scalogram analysis, if the child passed a task (i.e., received a positive score), he/she should also receive positive scores from the easier tasks. Based on the logic behind this analysis, children's scores for each DB task were ranked from the hardest to the easiest. The most difficult task can be explained as the one in which children fail the most. Similarly, the easiest task is the one in which children give the most correct answers. Ideally, if an individual's performance on one task is known, their performance on other tasks should also be known. This analysis produces two values. First one is the *coefficient of reproducibility*. The score of reproducibility is defined as "the extent to which item responses fit the ideal patterns" (Wellman & Liu, 2004, p. 532). If this score is over .90, items can be interpreted as scalable. The second value is the *index of consistency* which tests "whether the observed coefficient of reproducibility was greater than what could be achieved by chance alone" (Wellman & Liu, 2004, p. 532). If this value is above .50, it indicates significance. However, it should be noted that the index of consistency is an optional measure (Green, 1956).

Guttman analyses were performed separately depending on age and in total. Analysis of total samples revealed that children had the highest success rate on the DS-Think task, while the SS-*Guess* task had the lowest performance. Looking at the total sample, the easiest task is the DS-Think while the most difficult task is the SS-*Guess*. Although the *coefficient of reproducibility* is not significant (.84), it is closer to the significance level and may converge to .90 with larger sample size. 3-year-old children were the most successful in the DS-*Think* and the least successful in the SS-*NoMS*, however, the score of reproducibility was not significant. The highest performance for 4-year-old children was shown in the SS-*NoMS*, while the lowest performance was shown on the SS-*Guess*. The value of reproducibility was very close to the significance level (*the coefficient of reproducibility* = .892). Guttman's analysis showed significant results when performed with 5-year-old children (*the coefficient of reproducibility* = .91). The task sequence for 5-year-olds is as follows: DS-*Think*, DS-*NoMS*, SS-*Guess*, SS-*NoMS*, SS-*Think*, and DS-*Guess* (see Table 5).

Table 5. Guttman Scalogram Analysis for DB Tasks

	3 years	4 years	5 years	Total
	DS-*Think*	SS-*NoMS*	DS-*Think*	DS-*Think*
	SS-*Think*	SS-*Think*	DS-*NoMS*	DS-*NoMS*
Easiest	DS-*NoMS*	DS-*Guess*	SS-*Guess*	SS-*Think*
⇓	DS-*Guess*	DS-*NoMS*	SS-*NoMS*	SS-*NoMS*
Hardest	SS-*Guess*	DS-*Think*	SS-*Think*	DS-*Guess*
	SS-*NoMS*	SS-*Guess*	DS-*Guess*	SS-*Guess*

Table 5 (cont'd)

The coefficient of reproducibility	.82	.89	.91	.84

3.3. Day-Night Task, Its Relations with DB Tasks, and Age Group Differences

Point biserial correlation was performed to investigate the relations between one categorical (ToM performance) and one continuous variable (Day-Night task total scores). Children's performance in the DB tasks individually was not significantly related to performance on the EF task (see Table 6). Children's EF performance by age was also analyzed. There was no difference across the ages of 3 (M = 12.80, SD = 4.21), 4 (M = 11.93, SD = 5.61), and 5 (M = 13.63, SD = 3.98) in terms of day-night total scores, F (2, 42) = .504, p = .607. Then, children were divided into 4 groups. The first group included children who failed both the SS-*Think* and the DS-*Think* tasks (n = 4, M = 12.25, SD = 6.85). The second group represented the ones who failed the SS-*Think* but passed the DS-*Think* (n = 16, M = 12.88, SD = 4.10). The third group consisted of the children who failed the DS-*Think* but passed the SS-*Think* (n = 7, M = 10.71, SD = 5.22). Lastly, children who passed both the SS-*Think* and the DS-*Think* were termed as the fourth group (n =18, M = 13.72, SD = 4.30). The aim was to investigate whether children who were advantaged in double-seeker conditions as compared to single-seeker conditions performed better in the EF task. However, one-way ANOVA analysis showed that there was no significant performance difference across four groups, F (3,41) = 739, p = .535.

Table 6. The Correlations between Total EF Scores and Individual ToM Tasks

	Total EF Score	SS-*Think*	SS-*Guess*	SS-*NoMS*	DS-*Think*	DS-*Guess*	DS-*NoMS*
Total EF Score	1	.051	-.082	-.046	.199	-.179	-.060

†$p < .10$, * $p < .05$, ** $p < .01$, ***$p < .001$

CHAPTER IV

DISCUSSION

The ability to appreciate the mental states of others and predict their behaviors based on these mental states is an essential part of human social cognition. Wellman and Liu's (2004) ToM battery which includes a range of tasks tapping into different mental states created as a comprehensive measurement of ToM abilities. Since Wellman's theoretical stance (i.e., Theory Theory) claims that children are equipped with folk psychology domains at birth that helps them understand the mental states of others in preschool years, the progression of ToM concepts for children should be universal. Although TT approach also discusses the impact of culture on ToM concepts, Wellman's early conceptualization of ToM development that begins with the mastery of desires (i.e. *noncognitive state*) and progresses onto beliefs (i.e. *representational state*) is a defining feature of ToM scale (Wellman et al., 2006, p. 1076). However, cross-cultural ToM studies found that the degree of difficulty for each task is not universal for children.

While children from America, Australia, and Germany passed the diverse beliefs task before the knowledge access task, the ordering was the opposite for children from China and Iran. When examining Turkish children, it was possible to talk about the presence of both sequences and their underperformance on the diverse beliefs task compared to their peers from other cultures. It is worth noting that Turkish children did not show age-related improvement in the DB task, and their DB performance were not related to the performance on other ToM battery tasks and the child cognitive variables (i.e., EF). From the point of sociocultural perspective, which assigns a constitutive role to culture (i.e., social interactions and language) in the formation of mental state concepts, the current study endeavored to investigate the cultural relevance of the DB task for Turkish children. Concordantly, the mental state verbs in the story were changed considering the morpho-syntactic structure of the Turkish language. More importantly, the number of characters in the task was manipulated in order to assess whether asking about children's own beliefs creates a bias similar to the curse of knowledge experiencing in the FB tasks.

This study is of importance in the sense that it attempted to explore whether Turkish children have difficulty in the DB task because of methodological reasons although they have a conceptual understanding of belief diversity. Overall, there are important points to be discussed. It was observed that there was a change in performance on the double-seeker condition presented with the verb *think* compared to the single-seeker condition presented with the verb *think*. In other words, children were more successful in the DS-*Think* task than the SS-*Think* task which was created by Wellman and Liu in 2004 as the original version. However, the change in performance between single

versus double seeker conditions was not encountered on guess and no mental state versions. In addition, convergent with previous findings (e.g., Haskaraca, 2019), an expected developmental trajectory was not observed in the DB task regardless of the types of mental state verbs and the number of characters. In this section, these results will be discussed.

4.1. Task Success by Age

When taking task success percentages by age into consideration, it could be seen that children's task performance did not increase depending on age. More 4-year-old children were more successful than 5-year-old children in some of the tasks (i.e., SS-Think, SS-NoMS, and DS-Guess). It does not seem possible to explain these performances based on demographics. When considering both maternal and paternal educational and occupational status, it could be possibly interpreted that the sample in the current study consists of children who have middle or upper-middle class families. Therefore, additional analyses investigating the relations between task performance and demographic backgrounds were not included in this study. Since the current research has a limited sample size, it is anticipated that task success percentages which did not increase depending on child age cannot be explained by demographics. Future studies should consider studying with a larger sample size to see whether age-related changes in task performance may account for demographic differences across age groups.

4.2. Modifications Regarding the Mental State Verbs

In this study, the DB task was presented with different mental state verbs: "*think"* (düşün- in Turkish, "*guess"* (tahmin et-in Turkish) and *no mental state verb* condition

with the adverbs "*in my/your opinion*" (bence/sence in Turkish) used while asking own-belief question to children and adverbials "*according to me/you*" (bana/sana göre in Turkish) used while stating the protagonists' beliefs. It was found that none of these three manipulations produced an advantage for children to pass the task over the others. In line with Haskaraca (2019), there was no difference in terms of task performance between the verb *"think"* and "*guess"*. Further, Balcı (2021) revealed that Turkish adults interpreted the verb "*think*" and the phrase "*in my/your opinion"* as similar in terms of certainty levels. In this study, creating the no mental state verb (MSV) condition by using adverbs and adverbials aimed to observe whether Turkish children might interpret the *think* vs. the no-MSV versions of the same epistemic reality differently. However, it did not make any difference on Turkish children's DB task performance in line with the lack of difference between these two in terms of Turkish adults' certainty judgments.

Neither Haskaraca (2019) nor the current study found statistically meaningful differences for children's performance on the *think* vs. the *guess* version of the task. As Moore (1989) suggested, the distinction in certainty judgments between *think* and *guess* was not well understood by American children who are between 3 and 8 years of age. Since the current study was conducted with a small sample size and with

children aged 3 and 5, it cannot provide a solid ground for asserting whether Turkish children in this age range were able to make sense of the epistemological differences between mental state verbs. Future research should investigate how the use of the verbs *think* and *guess* creates a pragmatic difference for Turkish children through a larger sample size and wider age groups.

4.3. Modifications Regarding the Number of Characters

The other fundamental point that needs to be discussed is the performance differences between the single seeker and the double seeker conditions. The current study hypothesized that children might be biased by their own beliefs, and therefore, they might have difficulty in the original DB task in which children are asked own-belief question although they can conceptually appreciate diversity of beliefs. Findings partially supported the hypothesis. In addition, DS-Think was the easiest task for 3- and 5-year-old children in the Guttman analyses. It should be evaluated why the double seeker condition created an advantage only with the verb think and why this advantage was especially for children ages 3 and 5.

4.3.1. Why can double-seeker be easier than single-seeker?

First, the performance difference between single and double seeker conditions highlights the importance of the starting point of the current study. This study investigated whether the verbatim-translated versions of the original DB task created by Wellman and Liu (2004) might be problematic for Turkish children. It was due to cultural diversity (e.g., Ilgaz et al., 2022), age-related changes on the DB performance (e.g., Haskaraca, 2019), and the lack of relation of the DB task performance with the performance on the other tasks in the ToM battery for Turkish children (Ilgaz et al., 2022) and other cognitive variables such as EF (Doenyas et al., 2018; Evsen, 2022). When children's own perspectives were removed, and the story was reconstructed based on the conflicting beliefs of the two characters about the unknown location of an animal, a performance increase was observed. This result demonstrated that children might have difficulty passing the DB task due to the task structure rather than the lack of conceptual understanding of belief diversity.

Therefore, it seems plausible to infer that the phenomenon of the curse of belief may underlie Turkish children's underperformance on the DB task. Previous research stated that children were affected by the *curse of knowledge* at younger ages, leading them to fail in FB tasks. According to the inhibitory control account, since children may be affected by this bias because of their EF skills (e.g., inhibitory control) that are not yet developed at younger ages (Ghrear, 2021), the impact of the *curse of knowledge* should decrease with increasing age. To investigate whether *curse of belief* is a factor that affects children's DB performance, children's EF skills were measured. There was no age group difference in terms of the EF task performance. Likewise, further analyses found no difference on EF performance between children who failed both the SS-*Think* and the DS-*Think* and those who failed the SS-*Think* but succeeded in DS-*Think*. There may be a number of reasons for these findings. The other cognitive mechanism behind the curse of knowledge was the fluency misattribution account including the sense-making mechanism as well. It seems still plausible to claim that the phenomenon of the curse of belief may underlie Turkish children's underperformance on the DB task because of the sense-making mechanism described in the literature. More precisely, it is possible to speculate that in the original version Turkish children have difficulty answering according to the protagonist's belief because they already make sense of their own beliefs (when they make sense of their own beliefs, these beliefs become more fluent for them) and there is no explicit reason to prove the accuracy of the protagonist's belief that can potentially make children dissuade from their own beliefs.

One of the primary reasons concerns the validity of our EF task. Since the data collection process of the current study started during the pandemic, the aim was to

include a computer-friendly EF measurement which was the day-night task. However, implementing this task online with preschoolers brought some problems. In the original version created by Gerstadt, Hong, & Diamond (1994), children were provided with a certain amount of time to respond to each card. It was found that response time was positively correlated with accuracy. In other words, the longer children took to answer, the more likely the answer was correct. The original article studied with children from 3 to 7 years old. They found that older children gave faster and more correct responses than younger children in the day-night task. In the current study, response time to each card could not be received for consideration because of possible internet disconnections and situations that may develop suddenly about children (e.g., the need to use the bathroom). Therefore, only children's total score in testing trials was used in EF analyses. The underlying reason for no difference across age groups regarding EF performance can be interpreted as the problem with the response time for the abovementioned concerns. Future studies should consider using another measurement to assess EF skills because of these reasons.

It should also be noted that children in the current study were always presented with single-seeker conditions followed by double-seeker conditions. This study hypothesized that children would find the double-seeker tasks easier than the single-seeker tasks. Therefore, the aim was to prevent the order effect in which performance in the double seeker conditions enable children to perform better in the single-seeker conditions. Since children are generally more successful in the double-seeker tasks as compared to single-seeker tasks regardless of the mental state manipulations, if there would be other counterbalancing conditions which children are firstly presented with double-seeker tasks, children would be more successful in the single-seeker

conditions as well. Future studies may take the possible order effects into consideration while counterbalancing the tasks.

4.3.2. Why did 3- and 5-year-olds benefit from double-seeker equally?

Although EF measurement did not work as expected, the fact that both 3-year-olds and 5-year-olds benefited almost equally in DS-*Think* condition as compared to SS-*Think* condition (40% and 50%, respectively) could be explained by U-shaped development. In other words, children firstly showed a decremental performance followed by an incremental performance instead of a monotonic development (Pauls et al., 2013). It would be plausible to infer that children could pass the task despite their lack of knowledge and then fail because of incomplete knowledge. In the current study, 3-year-olds' impeccable performance on the SS-*Think* condition could be because they formed simple associations in their minds (e.g., "me → table and character → armchair"). Therefore, they all passed SS-*Think,* although they have not yet developed an understanding of belief diversity. Considering that 3-year-old children were successful in SS-*Think* for the wrong reasons, but they were also just as successful in DS-*Think* as 5-year-olds, the impacts of double seeker conditions on children's performance could be striking. When it comes to the benefit of the DS-Think task compared to SS-Think for 5-year-olds, it is plausible to see the effects of the curse of belief. Future studies should consider conducting a longitudinal study in order to investigate how the U-shaped developmental pattern will be ingendered. For instance, if 3-year-old children's SS-*Think* performance decreases or remains the same at Time 2 while DS-*Think* performance increases, it can be more clearly demonstrated that the double seeker condition becomes more suitable for Turkish children than the original DB task.

4.3.3. Why does the advantage of double-seeker only appear with the verb *think*?

The difference in performance between single-seeker and double-seeker conditions that is specific to the *think* version of the task is thought-provoking. Considering that the SS-*Guess* and the SS-*NoMS* conditions did not make a performance difference from SS-*Think*, it cannot be plausible to discuss this finding as a matter of debate. However, it is still possible for double-seeker conditions to create an advantage with *guess* and no mental state conditions with a larger sample as well. During testing sessions, it was observed that children did not pay attention to the verbs used. The change in verbs was just used when the experimenter asked the child's own belief and/or expresses the character's belief. Studying with a larger sample and using the different mental state verbs also in the target question might attract children's attention to the target verb. Apart from the larger sample, there may also be another task characteristic that caused this result to be achieved. This study strived to keep the double-seeker condition parallel to the original version in terms of stimuli, narratives, and questions. Just as in the original version, children were not asked a control question about whether they could distinguish between their self-produced beliefs and the characters' beliefs in double-seeker conditions. More precisely, when children were instructed that "Neşe thinks that the caterpillar is behind the flowers, but Burcu thinks that the caterpillar is behind the stones.", children may be confused to understand which character holds which belief that may make them fail in the task. Asking a memory question that tests children's mastery of both characters' divergent beliefs, it would be possible for children to show greater improvement in double seeker conditions than in single-seeker conditions regardless of the mental state

manipulation applied. Future studies should consider the idea of including the control question in both single-seeker (i.e., original) and double-seeker conditions.

4.4. Conclusion

To conclude, the current study attempted to investigate whether the underlying reason behind Turkish children's difficulty in the DB task is about task-specifics (i.e., mental state verbs used and the number of characters) instead of Turkish children's conceptual deficiency in belief understanding. The findings did not support the first hypothesis which states that children will be more successful in the No MS condition both in the single-seeker and the double-seeker versions. Changes in mental state verbs did not create an improvement on the DB task performance. However, the second hypothesis regarding the number of characters was partially supported. The presence of double characters in the task enhanced children's performance for the tasks presented with the verb *think*, however, this was not the case for the tasks with *guess* and No MS. It may be interpreted that since Turkish children do not see any evidence for the story character to hold that conflicting particular belief, they are likely to give an answer to the target question according to their own beliefs. Despite the low sample size and the EF measurement that did not work as expected, the current study has the features that can alert future researchers to the cross-cultural validity of the DB task with reference to the sociocultural approach.

REFERENCES

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation, 8*(17), 125-131.

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

Aksu-Koç, A., Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development, 2009*(125), 13-28.

Aksu-Koç, A., Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development, 2009*(125), 13-28.

Allen, J. W., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development, 28*(2), 96-133.

Balcı, A. S. (2021). *Turkish as a grammaticalized language: the effect of factivity alternation and certainty on theory of mind understanding* (Doctoral dissertation, Bilkent University).

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46.

Bartsch, K., & Wellman, H. M. (1995). Children talk about the mind. Oxford university press.

Bayen, U. J., Pohl, R. F., Erdfelder, E., & Auer, T. (2007). Hindsight bias across the life span. *Social Cognition, 25* (1), 83–97.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 378–391.

Bernstein, D. M., Kumar, R., Masson, M. E., & Levitin, D. J. (2018). Fluency misattribution and auditory hindsight bias. *Memory & Cognition, 46*, 1331–1343.

Birch, S. A. (2005). When knowledge is a curse: Children's and adults' reasoning about mental states. *Current directions in psychological science*, *14*(1), 25-29.

Birch, S. A., & Bernstein, D. M. (2007). What can children tell us about hindsight bias: A fundamental constraint on perspective–taking?. *Social cognition*, *25*(1), 98-113.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25–B31.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *The American Psychologist, 32*, 513–531.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and brain sciences*, *27*(1), 79-96.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70,* 853–865.

Doenyas, C., Yavuz, H. M., & Selcuk, B. (2018). Not just a sum of its parts: How tasks of the theory of mind scale relate to executive function across time. *Journal of Experimental Child Psychology*, *166*, 485-501.

Etel, E., & Yagmurlu, B. (2015). Social competence, theory of mind, and executive function in institution-reared Turkish children. *International Journal of Behavioral Development*, *39*(6), 519-529.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129-153.

Geurten, M., Lloyd, M., & Willems, S. (2017). Hearing "quack" and remembering a duck: Evidence for fluency attribution in young children. Child Development, 88, 514–522.

Ghrear, S., Fung, K., Haddock, T., & Birch, S. A. (2021). Only familiar information is a "curse": children's ability to predict what their peers know. *Child Development*, 92(1), 54-75.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory.

Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrika*, *21*(1), 79-88.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161.

Haskaraca, F. N. (2019). *Turkish speakers' conceptualization of belief-related words and its implications for theory of mind development* (Doctoral dissertation, Bilkent University).

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture*, *2*(1), 2307-0919.

Hughes, C., Devine, R. T., Ensor, R., Koyasu, M., Mizokawa, A., & Lecce. S. (2014). Lost in translation? Comparing British, Japanese, and Italian children's theory of mind performance. *Child Development Research, 2014*, 1-10.

Hughes, C., Devine, R. T., Ensor, R., Koyasu, M., Mizokawa, A., & Lecce, S. (2014). Lost in translation? Comparing British, Japanese, and Italian children's theory of mind performance. *Child Development Research, 2014*, 1–10.

Ilgaz, H., & Allen, J. W. (2021). (Co-) Constructing a theory of mind: From language or through language? *Synthese, 198*(9), 8463-8484.

Ilgaz, H., Allen, J. W. P., & Haskaraca, F. N. (2022). Is cultural variation the norm? A closer look at sequencing of the theory of mind scale. *Cognitive Development*, *63*, 101216.

Ilgaz, H., Burumlu-Kisa, E. (2021, April 7). The Relation Between Change in Mothers' Discourse Across Repeated Storytellings and Theory of Mind, SRCD 2021, Online.

Kabha, L., & Berger, A. (2020). The sequence of acquisition for theory of mind concepts: The combined effect of both cultural and environmental factors. *Cognitive Development, 54*, 100852.

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Validation of the" Theory of Mind" Scale. *Zeitschrift Fur Entwicklungspsychologie Und Padagogische Psychologie, 38*(4), 190-199.

Kuntoro, I. A., Saraswati, L., Peterson, C., & Slaughter, V. (2013). Micro-cultural influences on theory of mind development: A comparative study of middle-class and pemulung children in Jakarta, Indonesia. *International Journal of Behavioral Development*, *37*(3), 266-273.

Kuntoro, I., Peterson, C. C., & Slaughter, V. (2017). Culture, parenting, and children's theory of mind development in Indonesia. *Journal of Cross-cultural Psychology*, *48*(9), 1389–1409.

Leblanc, É., Bernier, A., & Howe, N. (2017). The more the merrier? Sibling composition and early manifestations of theory of mind in toddlers. *Journal of cognition and development*, *18*(3), 375-391.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false belief understanding across cultures and languages. *Developmental Psychology, 44*(2), 523– 531.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Mitchell, P., & Taylor, L., M. (1999). Shape constancy and theory of mind: Is there a link? *Cognition, 70,* 167-190.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. Child Development, 60(1), 167–171.

Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I. M., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology, 111*, 331–348.

Nelson, K. (1998). *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press.

Oberle, E. (2009). The development of theory of mind reasoning in Micronesian children. *Journal of Cognition and Culture*, *9*, 39–56.

Özoran, D. (2009). *Cognitive development of Turkish children on the relation of evidentiality and Theory of Mind* (Master's thesis, Enformatik Enstitüsü).

Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253-299.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child development*, *76*(2), 502-517.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Royzman, E. B., Cassidy, K. W., & Baron, J. (2003). “I know, you know”: Epistemic egocentrism in children and adults. *Review of General Psychology, 7*(1), 38-65.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers’ mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734-751.

Selcuk, B., Brink, K. A., Ekerim, M., & Wellman, H. M. (2018). Sequence of theory-of-mind acquisition in Turkish children from diverse social backgrounds. *Infant and Child Development*, *27*(4), e2098.

Shahaeian, A., Henry, J. D., Razmjoee, M., Teymoori, A., & Wang, C. (2015). Towards a better understanding of the relationship between executive control and theory of mind: An intra-cultural comparison of three diverse samples. *Developmental Science, 18,* 671–685.

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental psychology*, *47*(5), 1239.

Singh, N. U., Roy, A., Tripathi, A. K., Test, C. S., Test, B., Test, O. S. K. S., ... & Kendall’s, W. (2013). Non Parametric Tests: Hands on SPSS. *ICAR Research Complex for NEH Region, Umiam, Meghalaya*.

Tomasello, M. (2019). Becoming human. In *Becoming Human*. Harvard University Press.

Wellman, H. M. (2012). Theory of mind: Better methods, clearer findings, more development. *European Journal of Developmental Psychology*, *9*(3), 313-330.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives*, *11*(3), 207-214.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, *75*(2), 523-541.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development*, *72*(3), 655-684.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological science*, *17*(12), 1075-1081.

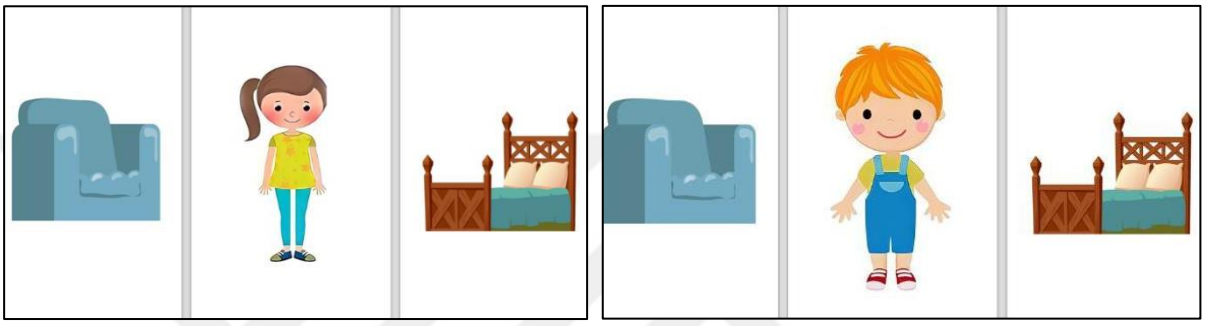
Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological science*, *17*(12), 1075-1081.

Whiten, A., Lewis, C., & Mitchell, P. (1994). Grades of mindreading. *Children's early understanding of mind: Origins and development*, *47*.

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

APPENDIX

1. Single Seeker - Think



Bak bu Dilek/Mert. Dilek/Mert kediyi bulmak istiyor. Kedi ya koltuğun arkasında ya da yatağın altında. Sen kedinin nerede olduğunu düşünüyorsun, koltuğun arkasında mı yoksa yatağın altında mı?

- Eğer çocuk “koltuk” derse: Pekâlâ, bu çok güzel bir fikir ama Dilek/Mert kedinin yatağın altında olduğunu düşünüyor.
- Eğer çocuk “yatak” derse: Pekâlâ, bu çok güzel bir fikir ama Dilek/Mert kedinin koltuğun arkasında olduğunu düşünüyor.

Peki Dilek/Mert kediyi bulmak için nereye bakacak, koltuğun arkasına mı yoksa yatağın altına mı?

(ENG): Look, this is Dilek/Mert. Dilek/Mert wants to find the cat. The cat is either behind the armchair or under the bed. Where do you think the cat is?

- If the child says "armchair": Well, this is a nice idea, but Dilek/Mert thinks that the cat is under the bed.
- If the child says "bed": Well, this is a nice idea, but Dilek/Mert thinks that the cat is behind the armchair.

Where will Dilek/Mert look for the cat, behind the armchair or under the bed?

2. **Single Seeker - Guess**

Bak bu Zehra/Tamer. Zehra/Tamer köpeği bulmak istiyor. Köpek ya dolabın içinde ya da masanın altında. Sen köpeğin nerede olduğunu tahmin ediyorsun, dolabın içinde mi yoksa masanın altında mı?

- Eğer çocuk “dolap” derse: Pekâlâ, bu çok güzel bir fikir ama Zehra/Tamer köpeğin masanın altında olduğunu tahmin ediyor.
- Eğer çocuk “masa” derse: Pekâlâ, bu çok güzel bir fikir ama Zehra/Tamer köpeğin dolabın içinde olduğunu tahmin ediyor.

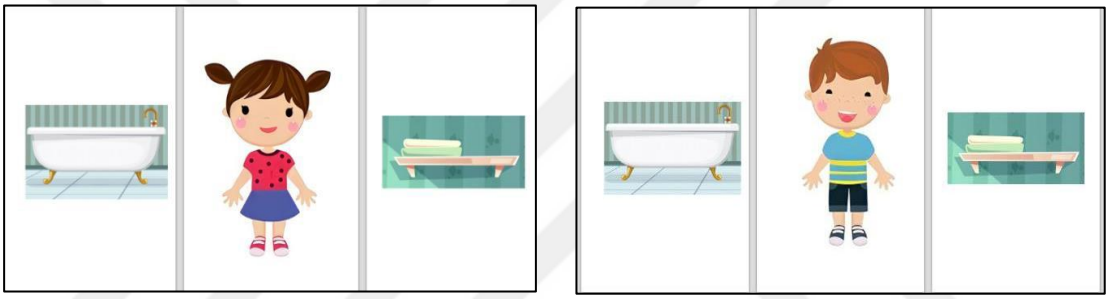
Peki Zehra/Tamer köpeği bulmak için nereye bakacak, dolabın içine mi yoksa masanın altına mı?

(ENG): Look, this is Zehra/Tamer. Zehra/Tamer wants to find the dog. The dog is either in the closet or under the table. Where do you guess the dog is?

- If the child says “closet”: Well, this is a nice idea, but Zehra/Tamer guesses that the dog is under the table.
- If the child says "table": Well, this is a nice idea, but Zehra/Tamer guesses that the dog is in the closet.

Where will Zehra/Tamer look for the dog, in the closet or under the table?

3. **Single Seeker – No Mental State**



Bak bu Oya/Ali. Oya/Ali civcivi bulmak istiyor. Civciv ya küvetin içinde ya da havluların arkasında. Sence civciv nerede, küvetin içinde mi yoksa havluların arkasında mı?

- Eğer çocuk “küvet” derse: Pekâlâ, bu çok güzel bir fikir ama Oya’ya/Ali’ye göre civciv havluların arkasında.
- Eğer çocuk “havlu” derse: Pekâlâ, bu çok güzel bir fikir ama Oya’ya/Ali’ye göre civciv küvetin içinde.

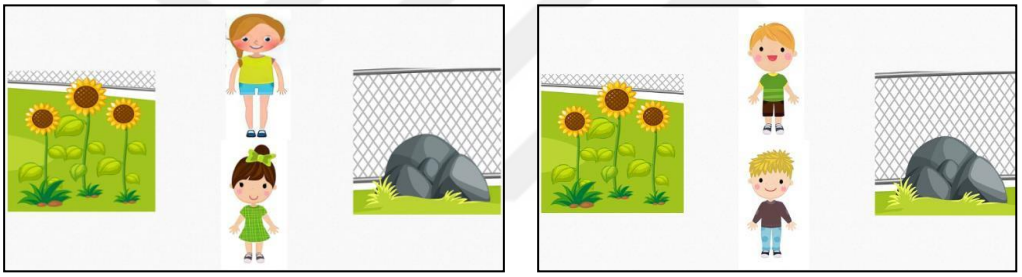
Peki Oya/Ali civcivi bulmak için nereye bakacak, küvetin içine mi yoksa havluların arkasına mı?

(ENG): Look, this is Oya/Ali. Oya/Ali wants to find the chick. The chick is either in the tub or behind the towels. Where is the chick, in your opinion? In the tub or behind the towels?

- If the child says “tub”: Well, this is a nice idea, but according to Oya/Ali, the chick is behind the towels.
- If the child says "towels": Well, this is a nice idea, but according to Oya/Ali, the chick is in the tub.

Where will Oya/Ali look for the chick? In the tub or behind the towels?

4a. Double Seeker – Think (Right)



Bak bu Burcu/Burak. Burcu/Burak tırtılı bulmak istiyor. Bak bu Neşe/Caner. Neşe/Caner de tırtılı bulmak istiyor.

Neşe/Caner, tırtılın çiçeklerin arkasında olduğunu düşünüyor “ama” Burcu/Burak tırtılın, taşların arkasında olduğunu düşünüyor.

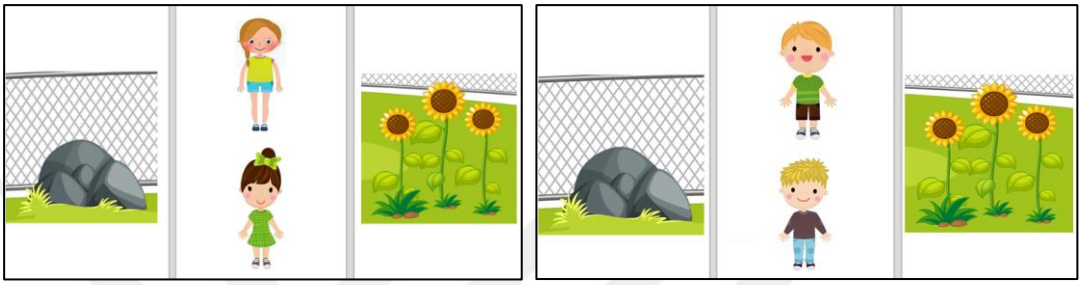
Peki Burcu/Burak tırtılı bulmak için nereye bakacak, çiçeklerin arkasında mı yoksa taşların arkasında mı?

(ENG): Look, this is Burcu/Burak. Burcu/Burak wants to find the caterpillar. Look, this is Neşe/Caner. Neşe/Caner wants to find the caterpillar as well.

Neşe/Caner thinks that the caterpillar is behind the flowers, “but” Burcu/Burak thinks that the caterpillar is behind the stones.

Where will Burcu/Burak look for the caterpillar, behind the flowers or behind the stones?

4b. Double Seeker – Think (Left)



Bak bu Burcu/Burak. Burcu/Burak tırtılı bulmak istiyor. Bak bu Neşe/Caner. Neşe/Caner de tırtılı bulmak istiyor.

Neşe/Caner, tırtılın çiçeklerin arkasında olduğunu düşünüyor “ama” Burcu/Burak tırtılın, taşların arkasında olduğunu düşünüyor.

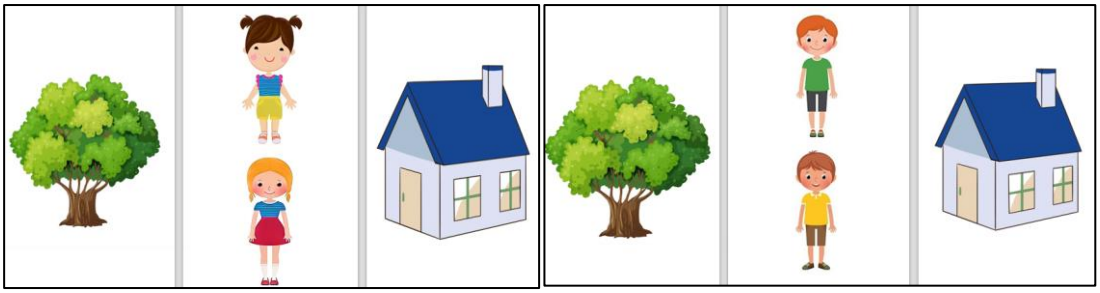
Peki Burcu/Burak tırtılı bulmak için nereye bakacak, taşların arkasına mı yoksa çiçeklerin arkasına mı?

(ENG): Look, this is Burcu/Burak. Burcu/Burak wants to find the caterpillar. Look, this is Neşe/Caner. Neşe/Caner wants to find the caterpillar as well.

Neşe/Caner thinks that the caterpillar is behind the flowers “but” Burcu/Burak thinks that the caterpillar is behind the stones.

Where will Burcu/Burak look for the caterpillar, Behind the stones or behind the flowers?

5a. Double Seeker – Guess (Right)



Bak bu Özlem/Ömer. Özlem/Ömer kuşu bulmak istiyor. Bak bu Ayça/Ahmet. Ayça da/Ahmet de kuşu bulmak istiyor.

Ayça/Ahmet, kuşun ağacın içinde olduğunu tahmin ediyor “ama” Özlem/Ömer, kuşun çatının arkasında olduğunu tahmin ediyor.

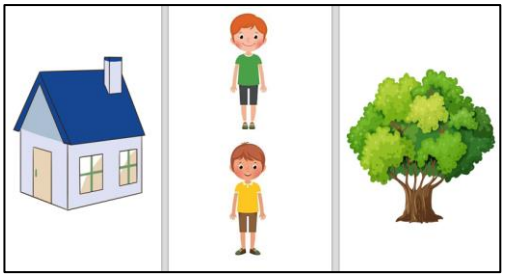
Peki Özlem/Ömer kuşu bulmak için nereye bakacak, ağacın içine mi yoksa çatının arkasına mı?

(ENG): Look, this is Özlem/Ömer. Özlem/Ömer wants to find the bird. Look, this is Ayça/Ahmet. Ayşe/Ahmet wants to find the bird as well.

Ayça/Ahmet guesses that the bird is in the tree “but” Özlem/Ömer guesses that the bird is behind the roof.

Where will Özlem/Ömer look for the bird, in the tree or behind the roof?

5b. Double Seeker – Guess (Left)



Bak bu Özlem/Ömer. Özlem/Ömer kuşu bulmak istiyor. Bak bu Ayça/Ahmet. Ayça da/Ahmet de kuşu bulmak istiyor.

Ayça/Ahmet, kuşun ağacın içinde olduğunu tahmin ediyor “ama” Özlem/Ömer, kuşun çatının arkasında olduğunu tahmin ediyor.

Peki Özlem/Ömer (mouse ile göster) kuşu bulmak için nereye bakacak? Çatının arkasına mı yoksa ağacın içine mi?

(ENG): Look, this is Özlem/Ömer. Özlem/Ömer wants to find the bird. Look, this is Ayça/Ahmet. Ayşe/Ahmet wants to find the bird as well.

Ayça/Ahmet guesses that the bird is in the tree “but” Özlem/Ömer guesses that the bird is behind the roof.

Where will Özlem/Ömer look for the bird, behind the roof or in the tree?

6a. Double Seeker – No Mental State (Right)

Bak bu Leyla/Osman. Leyla/Osman kaplumbağayı bulmak istiyor. Bak bu Sevgi/Selim. Sevgi/Selim de kaplumbağayı bulmak istiyor.

Sevgi'ye/Selim'e göre, kaplumbağa çitlerin arkasında "ama" Leyla'ya/Osman'a göre, kaplumbağa kütüğün içinde .

Peki Leyla/Osman kaplumbağayı bulmak için nereye bakacak, çitlerin arkasına mı yoksa kütüğün içine mi?

(ENG): Look, this is Leyla/Osman. Leyla/Osman wants to find the turtle. Look, this is Sevgi/Selim. Sevgi/Selim wants to find the turtle as well.

According to Sevgi/Selim, the turtle is behind the fence "but" according to Leyla/Osman, the turtle is in the wood block.

Where will Leyla/Osman look for the turtle, behind the fence or in the wood block?

6b. Double Seeker – No Mental State (Left)

Bak bu Leyla/Osman. Leyla/Osman kaplumbağayı bulmak istiyor. Bak bu Sevgi/Selim. Sevgi/Selim de kaplumbağayı bulmak istiyor.

Sevgi'ye/Selim'e göre, kaplumbağa çitlerin arkasında "ama" Leyla'ya/Osman'a göre, kaplumbağa kütüğün içinde.

Peki Leyla/Osman kaplumbağayı bulmak için nereye bakacak, kütüğün içine mi yoksa çitlerin arkasına mı?

(ENG): Look, this is Leyla/Osman. Leyla/Osman wants to find the turtle. Look, this is Sevgi/Selim. Sevgi/Selim wants to find the turtle as well.

According to Sevgi/Selim, the turtle is behind the fence "but" according to Leyla/Osman, the turtle is in the wood block.

Where will Leyla/Osman look for the turtle, in the wood block or behind the fence?