ACQUIRING A WORD THROUGH THE SPEAKER'S FALSE-BELIEF

By

GALA STOJNIĆ

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Alan M. Leslie

And approved by

____

____

____

____

____

New Brunswick, New Jersey

January, 2021

ABSTRACT OF THE DISSERTATION

Acquiring a Word Through the Speaker's False-Belief

By GALA STOJNIĆ

Dissertation Director:

Alan M. Leslie

This work bridges two major areas of the cognitive development research–Theory of Mind (ToM) development and word learning. Specifically, I test the idea that ToM is a necessary factor in early word acquisition process, as it helps young learners form a hypothesis about what the most likely referent of a novel label might be.

The idea that ToM is an important factor that young children use to select a candidate referent of a novel label is largely inspired by social approaches to word learning and word meaning (e.g. Baldwin, 1991; Cartmill et al., 2013; Grice, 1969). Critically, we postulate that the nature of early ToM that underlies this process is meta-representational, i.e. it allows for representing genuine propositional attitudes such as beliefs (e.g. Leslie, 1987). We call this *developmental continuity hypothesis.* An opposing view (*dual-systems hypothesis*) postulates that ToM lacks the meta-representational structure prior to the age of four, hence it cannot adequately account for various cases that require flexible mental states reasoning, such as reasoning about identity false-beliefs (e.g. Butterfill & Apperly, 2013). Here, we test these opposing hypotheses by asking whether young children (prior to the age of four) would succeed on a task that required them to map a label onto its referent by correcting for the speaker's identity false-belief. If children could attribute these types of beliefs prior to the age of four and, moreover, use them to learn words this would provide a compelling support in favor of developmental continuity of meta-representational ToM.

To test these hypotheses, I developed a new naming false-belief task, which requires subjects to point to the correct referent of a novel label (a proper name or a common noun), where the only way to infer who the referent is, is through an agent's identity false-belief. I conducted a set of experiments with young 3-year-olds and 2.5-year olds, where I employ this method to show that at least by the age of 3, children are capable of mapping a label and its referent by correcting for the speaker's identity false-belief. This provides compelling evidence in support of developmental continuity of meta-representational ToM. Moreover, it demonstrates that young word learners attend to the speaker's epistemic states to narrow down the space of a word's possible meanings, which suggests strong mentalism of the word learning process.

Acknowledgments

I thank my advisor, Dr. Alan M. Leslie, for his remarkable support and guidance throughout this journey. Alan has been incredibly inspiring and supportive, and has fundamentally shaped my intellectual, professional and personal growth. He truly taught me how to think as a developmental scientist, through his insightful advice, attention to scientific rigor and genuine excitement about the field. This work would not be possible without him and I am tremendously grateful for having him as my advisor.

I want to thank all those who have provided mentorship throughout my graduate program. I owe tremendous gratitude to my committee members, Dr. Judith Hudson, Dr. Paul Pietroski, Dr. Renée Baillargeon and Dr. Lila Gleitman. Their invaluable guidance and insightful feedback have greatly influenced and improved this work. I also want to extend special thanks to Dr. Ernie Lepore, whose feedback on my work played an integral role in my professional development.

I am tremendously grateful to the preschools and families who participated in this research. They truly contributed to the core of this dissertation, and without them this work would be impossible. I thank all the current and former members of the Cognitive Development Lab (CDL) for their inspiring enthusiasm for the field and scintillating intellectual discussions. I also thank many of our Research Assistants for spending countless hours on recruiting subjects and assisting in stimuli creation and for making this work possible.

I thank my family, especially my parents Jasna Stojnic and Danko Stojnic, my sisters, Una Stojnic and Aneta Stojnic, and my husband, Andreas Voellmer, for their undying support, encouragement and love. I wouldn’t be where I am today, without them.

Lastly, I want to thank a *canis domesticus,* Tasa, who has inspired the creation of experimental stimuli for this project. His contribution is justly appreciated

Table of Contents

Preliminaries

1. Chapter 1

2. Chapter 2

3. Chapter 3

List of Tables

List of Figures

Acquiring a word

through the Speaker's false-belief

Chapter 1

Background

Theory of Mind (ToM) ability, the cognitive capacity to represent social agents' (including one's own) mental states and to understand their causal role in externally observable behavior (e.g. Premack & Woodruff, 1978), has been actively explored over more than 30 years. Despite the wealth of research that has been done on ToM, there are still large disagreements in the field regarding some of the key questions: what is the exact representational structure of ToM and when does this structure fully develop?

Most of the controversy revolves around seemingly incompatible findings that come from the false-belief task (FBT) paradigm, designed to test if a child understands that others' mental representations can differ from their own (e.g. Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). Typically, when the standard FBT is employed, children struggle to show appreciation of another's false belief until they are 4 years old, as prior to this age they mostly fail to predict the actions of an agent who is mistaken about a certain aspect of reality (see Wellman, Cross, & Watson, 2001; Wellman, 2014). This discrepancy in performance between younger children and those who are over the age of four has led a number of researchers to adopt the late-competence account, according to which ToM lacks a proper representational structure, viz. one that would enable representing the concept of false-belief, until the age of four. (e.g. Gopnik, & Meltzoff, 1997; Gopnik & Wellman, 1994; Perner, 1991; Perner, 1995; Perner & Ruffman, 2005; Wellman, Cross, & Watson, *Ibid*).

However, several studies that used modified versions of the standard FBT have shown that the performance of younger children can be significantly improved when the overall processing demands of the task are reduced (e.g. Bartsch, 1996; Carpenter, Call, & Tomasello, 2002; Zaitchik, 1991; Setoh, Scott, & Baillargeon, 2016; Grosso, Schuwerk, Kaltefleiter, & Sodian, 2018).

Curiously, a number of studies showed that even pre-verbal babies appear capable of attributing false beliefs to other agents, relying on different implicit measures of false-belief reasoning (e.g. Baillargeon, Scott, & Bian, 2016; Baillargeon, Scott, He, Sloane, Setoh, Jin, Wu, & Bian, 2015; Baillargeon, He, Setoh, Scott, Sloane, & Yang, 2013; Baillargeon, Scott, & He, 2010; Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Suhrke, & Buttelmann, 2015; Forgács, Gervain, Parise, Csibra, Gergely, Baross, & Király, 2020; Luo & Baillargeon, 2010; Onishi & Baillargeon, 2005; Scott, Baillargeon, Song, & Leslie, 2010; Surian, Caldi, & Sperber, 2007; Song, & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Southgate & Vernetti, 2014). These findings have supported the early competence approach, which posits that children younger than 4 fail the standard FBTs due to performance-related factors, rather than due to conceptual deficits of ToM competence (e.g. Baker, Leslie, Gallistel, & Hood, 2016; He, Bolz, & Baillargeon, 2011; Leslie, 1987; Leslie, 1988; Leslie & Frith, 1990; Leslie, 1991; Leslie & Thaiss, 1992; Leslie & Roth, 1993; Leslie, 1994a; Leslie, 1994b, Leslie, Friedman, & German, 2004; Leslie, 2000; Scholl & Leslie, 2001; Wang, Hemmer, & Leslie, 2019).

Although the empirical evidence in favor of the early-competence approach has been steadily accumulating, considerable skepticism regarding the nature of early belief sensitivity has remained in the field. Critics of the early-competence approach have argued that the evidence for early belief attribution could be explained away by appeal to simpler processes rather than by the attribution of mental states as such (e.g. Heyes, 2014a; Perner, 2010; Perner & Ruffman, 2005; Ruffman, 2014, Apperly & Robinson, 2003; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Surtees, Butterfill & Apperly, 2012; Low & Watts, 2013; Rakoczy, Bergfeld, Schwarz & Fizke, 2014). These lower level explanations ranged from attributing early successes to extraneous experimental factors and simple behavioral rules, to postulating a non-mentalistic ToM system, which allows for tracking (but not representing) others' mental states and is, hence, subject to a number of representational limits.

The long-lasting question of whether children's performance on various FBTs reflects a genuine, *metarepresentational* ToM competence (e.g. Leslie, 1987), or can largely be explained by simpler, non-(or low-)mentalistic processes, is one of the central questions I explore in this dissertation. I will argue for the position that metarepresentational ToM presents a part of children's core cognitive architecture (see Carey & Spelke, 1994; Spelke, 2000, 2007; also Leslie, 1994b) and, as such, is present from a very young age to support not only reasoning about social domain (i.e. prediction and explanation of agents' actions) but to enable acquisition of knowledge in other domains, such as language. Specifically, I argue that metarepresentational ToM plays an important role in the process of word acquisition, as it helps a young learner select the most likely candidate for a novel word's meaning.

The idea that young learners rely on profound ToM reasoning to acquire novel words is at odds with associationist approaches to word learning, according to which a word is acquired by virtue of a general, statistical learning mechanism, which enables a gradual formation of word-referent links (e.g. Richards & Goldfarb, 1986; Smith, 2000; Yu, 2008; Smith, Jayaraman, Clerkin, & Yu, 2018; Saffran, & Kirkham, 2018; also Locke, 1960). At the same time, this idea is well aligned with theoretical approaches and empirical evidence, which suggest that mere statistical tracking of word-world pairs won't suffice to fully explain the word learning process (e.g. Bloom, 2000; Gleitman, Liberman, McLemore, & Partee, 2019; Gleitman & Trueswell, 2018; Gleitman, 2009; Gleitman & Fisher, 2005; Gleitman & Gleitman, 1992; Landau & Gleitman, 1985; Medina, Snedeker, Trueswell, & Gleitman, 2011). Here, I aim to show that, indeed, a young word learner can override word-world associations by accounting for the (meta)representational context of the speaker's mind. The key idea is that the most likely candidate for a word's referent is a part of the speaker's representation of the world, rather than a part of the world itself; young children's ToM capacity enables them to "consult" the speaker's representation of the world in the process of word learning.

This dissertation is structured in the following way. I first provide a detailed overview of the existing accounts of children's performance on various FBTs. Then, I focus on the novel assumption that early ToM plays a key role in children's linguistic development, specifically, in their acquisition of novel words. I, then, present a set of experiments where I test young children's ability to employ early ToM reasoning in the process of word acquisition. I show new evidence that young children are capable of passing a verbal FBT prior to the age of four, and moreover, that they are capable of employing false-belief reasoning in the case that does not explicitly ask them to predict or explain agents' actions, but to learn a new label from them. This provides one of the first evidence for the role of ToM reasoning—specifically, false-belief reasoning—in the case of word acquisition.

1.1. The nature of representational structure of early Theory of Mind (ToM)

1.1.1. Conceptual-changes account vs. "Theory of Mind Mechanism" (ToMM).

The development of ToM has traditionally been tested by the FBT paradigm. Its standard version takes the form of a verbally narrated story in which a character forms a false belief about a certain aspect of the situation (e.g. Sally falsely believes her toy is in the box, but it is actually in the basket) and subjects are, then, asked to predict the character's behavior (e.g. Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). In order to pass the task, a child has to ignore her own (true) belief about the relevant aspect of the situation and, instead, rely upon the character's false belief when predicting her actions. However, a well-established finding has been that children massively struggle with the standard FBT until they are at least four years old (see Wellman, Cross, & Watson, 2001).

The performance discrepancy between younger and older children has led the advocates of the late-competence view to suggest that the key change in the representational structure of ToM occurs at the age of four; namely, the false-belief concept becomes available in its representational repertoire (e.g. Gopnik & Astington, 1988; Gopnik & Meltzoff, 1997; Gonpik & Wellman, 1994; Perner, 1991; Perner, 1995; Perner & Ruffman, 2005; Wellman, et al., 2001). Although theories encompassed within the late-competence approach differ in exact explanations of the nature of the representational change[1], they all agree upon the assumption that ToM computational structure gradually transforms into the mature form, as a child interacts with the surrounding environment.

However, challenging the idea that the false-belief concept is unavailable until the age of 4, a number of researchers pointed out that the standard FBTs are excessively complicated and, thus, underestimate young children's actual ToM competence (e.g. Bloom & German, 2000). According to these researchers, young children fail the standard FBTs because of extraneous factors associated with the standard FBTs, such as the proper mastery of syntax and pragmatics (e.g. Mitchell & Lacohce, 1991; Siegal & Beattie, 1990; de Villiers & Pyers, 2002; Hale & Tager-Flusberg, 2003)[2] and development of executive functions (e.g. Carlson, Moses, & Hix, 1998), rather than because of a conceptual deficit of ToM itself. Indeed, there have been a handful of studies that demonstrated considerable improvement of 3-year-olds when certain

[1] For instance, the change has been explained as a shift in the folk-psychology available to children; namely, from the simple desire reasoning to the belief-desire reasoning (e.g. Wellman & Woolley, 1990). Others characterize the change as a shift from non-representational (situation-based) to representational ToM (e.g. Perner, 1991).

[2] For instance, De Villiers and Pyers (2002) found that passing the traditional FBT was linked to the mastery of the syntax of sentential complements (i.e. "that" clauses) (see also de Villiers, 2005). Others suggested the link between having experience with mental state verbal terms and successes on FBTs (e.g. Dunn & Brophy, 2005; Taumoepeau & Ruffman, 2006). Although these studies are sometimes taken as the evidence of the formative role of language on ToM capacity, what they strictly speaking show is simply that mastering certain skills required for passing the traditional FBT will help children pass the traditional FBT.

modifications of the FBTs are made. For instance, Siegal and Beattie (1991) significantly improved 3-year-olds' performance by changing the test question from "Where will Sally look for the toy?" to "Where will Sally look *first* for the toy?" thus helping them to grasp the proper interpretation of the question. Sullivan and Winner (1993) demonstrated that 3-year-olds were capable of passing the FBT when they were required to trick another person into forming the false belief. Others found a link between the standard FBT performance and children's mastery of sentential complements, i.e. "*that*" clauses (de Villiers & Pyers, 2002; see also de Villiers, 2005), as well as children's passing the FBT and their experience with mental state verbal terms (e.g. Dunn & Brophy, 2005; Taumoepeau & Ruffman, 2006).

The assumption that standard FBTs obscure the actual ToM competence by performance factors, extraneous to ToM, is effectively captured by Leslie's (e.g. 1987; 1988; 1994a; Leslie & Frith, 1990) model of theory of mind reasoning. Leslie proposes that ToM ability rests on a specific neurocognitive mechanism, i.e. Theory of Mind Mechanism (or ToMM, for short), whose fundamental computational structure is innately specified and allows for representing *propositional attitudes,* such as beliefs and pretenses. To represent a propositional attitude, ToMM has to identify an agent and a particular attitude (informational relation) the agent holds towards a specific representation. Importantly, the representation in question is referentially opaque and is the product of ToMM decoupling the primary (literal) representation and placing it into the meta-representational context, where its semantic relations to reference, truth and existence presuppositions get suspended. This larger, meta-representational structure, consisting of an agent, a (decoupled) representation and an attitude that relates the agent and the representation, is a hallmark of ToMM. Importantly, a set of primitive attitudes, including the concepts of belief, desire and pretense, is available early on in one's life.

Leslie (1987, 1994a) argued that young toddlers overtly manifest their meta-representational ToMM as early as they start engaging in and (more importantly) recognizing pretense in others, which typically happens around the ages of 18 and 24 months. According to Leslie, what underlies both children's own pretense and their recognition of pretense in others is a single mental state, *pretense*, offered by ToMM. For instance, when a child sees her mom talking to a banana as if it were a telephone, ToMM spontaneously attributes a particular cognitive attitude (i.e. pretense) that the mom holds towards the decoupled representation (*this banana "is a telephone"*). By suggesting that it is a particular mental state—*pretense*—that underlies children's engaging in (and recognizing) of pretense behavior, Leslie's model was the first to offer a mentalist account of pretense (e.g. Friedman & Leslie, 2007). Namely, the traditional accounts of pretense conceptualize it as a form of behavior, enabled by emergence of a particular cognitive process (e.g. Harris, 1995; Nichols & Stich, 2000; Piaget, 1962; Walton, 1978), but not as a mental state. Because of that, these accounts suffer from the fundamental shortcoming in that they cannot account for the social nature of pretense. The mere ability to produce a specific behavior—namely, pretense behavior—will not enable a child to recognize pretense in others. If she lacked the concept of pretense, she could only rely on what was available through her perceptual apparatus—the literal description of mom's behavior: *she is talking to the banana.* Apparently, this would be very puzzling for a child (given that she knows a bit about bananas, phones and her mom); *is mom talking to the banana, because bananas and phones can be used for the* *same purpose?* Nonetheless, young children are not confused at all; they successfully recognize others' pretense and readily engage in mutually shared pretense activities, as demonstrated by a number of empirical studies (e.g. Bosco, Friedman, & Leslie, 2006; Harris, Kavanaugh, & Dowson, 1997; Onishi, Baillargeon, & Leslie, 2007; Walker-Andrews & Kahana-Klaman, 1999). One particularly notable study, favoring the meta-representational over the behavioral account, was done by Firedman, Neary, Burnstein and Leslie (2010). In their experiment, a researcher holding a puppet bear, was either uttering certain requests in her own voice or was pretending that the bear was uttering them. 2- and 3-year olds successfully fulfilled requests, regardless of whether they were "uttered" by the bear or uttered by the researcher. This suggested that they were going beyond mere behavior and were interpreting the speech as coming from a counterfactual source (i.e. the bear).

The fact that children can recognize others' pretense at such young age indicates that they are capable of understanding others' counterfactual representations way before the age of four. Then, why do they fail on the standard FBT until turning 4 years of age? According to Leslie (Leslie, 2000; Leslie & Polizzi, 1998; Leslie, German, & Polizzi, 2005), this is because, in addition to domain-specific ToMM, this task requires a domain-general component—the selection processor (SP). SP has to select between the two competing possible contents of the character's belief computed by ToMM, the true-belief (TB) and the false-belief (FB). SP favors the TB by default[3] and, hence, this bias has to

[3] The "true-belief default" is conceptualized as an inbuilt constraint of ToMM. Since in the majority of cases social agents indeed hold true-beliefs, this default is assumed to be the rational prior (Leslie, *ibid*). However, if true-belief is a default that needs to be inhibited for a child to succeed on an FBT, how do we explain toddlers and infants successes on implicit FBTs, when they lack sufficient inhibitory power? Namely, children around the age of 3.5 barely have enough inhibitory power to overcome the true-belief default, under specific conditions (e.g. lowering demands by removing the target object form the scene). Given that toddlers and infants lack sufficient inhibitory power, they should *always*

be previously inhibited to allow for the selection of the situation-appropriate mental state, i.e. the FB. Thus, it is because the process of selection-through-inhibition develops relatively slowly that young children fail on the standard FBT, rather than because of the lack of the false-belief concept.

The special value of Leslie's ToMM-SP is that not only does it have explanatory power, but it also generates testable predictions. For instance, the model predicts that if inhibitory demands of the FBTs are reduced, the performance of children younger than 4 will improve; similarly, if inhibitory demands of the FBTs are increased, the struggles will occur even in four-year-olds. Indeed, several studies that manipulated the inhibitory-selection demands of the standard FBT supported these predictions. For instance, it was shown that performance of children bellow the age of four improves significantly when the "low-demand" version of FBT is employed, in which the target object gets completely removed from the scene before the agent comes back to search for it (e.g. Bartsch, 1996; Carpenter, Call, & Tomasello, 2002; Setoh, Scott, & Baillargeon, 2016; see also Wang & Leslie, 2016). According to the ToMM-SP model, this happens because the default true-belief candidate is rendered unspecified; hence, the inhibitory demands no longer exceed a young child's processing resources.

Similarly, the ToMM-SP model can account for the fact that even 4-year olds

attribute a true-belief to an agent and, therefore, systematically fail FBTs. Yet this is not the case, as will be elaborated in the following section. A reason for this might lay in the nature of the elicited-prediction question, which could trigger a strategy to focus the child's attention to their own (true) representation of the scene to predict the character's action in a standard FBT (Baillargeon, p.c.). If this was the case, then any FBT that doesn't ask the prediction question would be less in demand for inhibitory control to defeat the true-belief default. This would, however, have certain implications for the true-belief default assumption, where this tendency to (incorrectly) attribute a true-belief would have to be triggered by a particular set of external circumstances (e.g. a prediction question), rather than acting as an actual default (Baillargeon, p.c.). Whether this is, indeed, the case is a question that requires explicit empirical testing in the future.

struggle when the character has a desire to avoid (rather than to approach) the target-object (e.g. Cassidy, 1998; Leslie, Friedman, & German, 2004). In this case, a subject has to apply a double inhibition, one for the belief and another for the desire, in order to succeed on the task; these increased inhibitory demands, thus, make the avoidance-desire FBT even more challenging than the traditional FBT (e.g. Leslie et al., 2004; 2005; Friedman & Leslie, 2004).

These studies indicate that the meta-representational structure of ToM is available *by at least* the age of 3. In the following section I review the studies employing implicit measures to test ToM reasoning even in younger children and babies and I examine the implications these studies have for the nature of early ToM competence.

1.1.2. A deeper look into early ToM competence: Implicit measures of ToM reasoning.

The assumption that the false-belief concept is available already in infancy has received compelling support by a number of studies that employ implicit measures of ToM reasoning. Unlike explicit FBTs, implicit tasks do not involve explicitly asking a child to predict the agent's likely behavior (e.g. "Where will Sally look for the toy?"). Rather, they rely either on elicited intervention (i.e. including a verbal prompt to elicit an action, but without directly acting a prediction question) or on spontaneous responses (i.e. responses gathered from subjects, without a verbal prompt) (see Scott & Baillargeon, 2017).[4]

[4] The distinction between implicit and explicit measures is sometimes understood as referring to non-verbal and verbal tasks, respectively. However, this analogy is not quite accurate, as implicit tasks can also be verbal (for instance, in elicited-intervention tasks where children's responses are elicited by a verbal prompt, but without asking them to predict a character's actions (e.g. Clemens & Perner, 1994; He, Bolz, & Baillargeon, 2012; Scott, He, Baillargeon, & Cummins, 2012)). In fact, there has been a growing number of implicit FBTs, both spontaneous-response and elicited-intervention, which renders the distinction between verbal and non-verbal tasks less useful for the field of ToM research (e.g. Baillaregon, p.c.).

Several notable studies employ the violation of expectation (VoE)[5] paradigm to explore “baby ToM” reasoning (e.g. Onishi & Baillargeon, 2005; Onishi, Baillargeon, & Leslie, 2007; Scott & Baillargeon, 2009; Scott, He, Baillargeon, & Cummins, 2012; Surian, Caldi, & Sperber, 2007; Song & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fischer, 2008). For instance, in the pioneering study, Onishi and Baillargeon (2005) presented 15-month-olds with a non-verbal analog of the FBT to show that infants were surprised (indicated by their longer looking times) if the actor reached for an object at its actual location, rather than at the location she believed it to be. Similarly, Surian et al. (2007) demonstrated that 13-month-olds expected an agent’s actions to be guided by its (true/false) beliefs, and Kovács, Téglás and Endress (2010) found evidence of automatic belief computation even in 7-month-old infants. Other VoE studies showed that by 18 months, children expect an agent to correct her false belief in virtue of a relevant verbal message (e.g. Song et al., 2008); they are sensitive to particular modes of belief formation (Song & Baillargeon, 2008); they can detect violations in pretense sequences (Onishi, Baillargeon, & Leslie, 2007); and they can attribute false-beliefs about object identities (Scott & Baillargeon, 2009).

The evidence from these studies indicates that sensitivity to others’ false beliefs is already available in infancy. However, some authors have questioned the validity of the VoE method and have argued that an infant could succeed on the VoE FBT simply by noticing *something* puzzling in the setup, and not necessarily by representing mental state as such (e.g. Perner & Ruffman, 2005; Sirions & Jackson, 2007). To strengthen the

[5] The rationale behind the VoE paradigm is that differences in infants’ looking times to specific events reveal whether they have prior expectations regarding these events. Hence, if an infant looks longer at an event A compared to an event B, it is taken as an indicator that the infant is *surprised* by A, presumably because it violated her initial expectations.

mentalistic interpretation of infancy and toddler data, one needs more active measures of ToM reasoning. For instance, we should be able to demonstrate that young subjects (at least) *anticipate* others' actions based on her mental states.

Indeed, a number of researchers managed to demonstrate false-belief understanding in young toddlers and infants using more active FBT measures. For instance, several studies relied upon the anticipatory looking (AL) paradigm[6] to show that subjects as young as 18 months do anticipate that an agent will reach for a desired object where she (falsely or correctly) believes it to be, as indicated by their first saccades to one of the possible target locations (e.g. Southgate, Senjy, & Csibra, 2007; Senjy, Southgate, Snape, Leonard, & Csibra, 2011; Wang & Leslie, 2016).

Another group of researchers relied upon the spontaneous helping paradigm[7] to demonstrate early false-belief understating (e.g. Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Suhrke, & Buttlemann, 2015; Knudsen & Liszkowski, 2012). For instance, Buttelmann et al. (2009) tested if 18-month-olds would help an agent who struggled to open a box and retrieve a desired object which she believed was in the box. They found that in the case in which the agent held a false-belief about the object location, toddlers (after being prompted to help the agent) approached the box the agent was not attempting to open (correctly recognizing that she wanted the toy, but wrongly thought it was in another box); by contrast, when the agent had a true-belief, the children approached the box she was trying to open.

Similarly, Southgate, Chavailer and Csibra (2010) riled upon eliciting

[6] AL presents another common, implicit measure, which (unlike VoE) requires a subject to make an anticipatory choice as indicated by her first saccade to one of the possible response-targets.

[7] This paradigm relies on young toddlers' well-evidenced tendency to spontaneously help others in need (e.g.Warneken & Tomasello, 2006; 2007).

reaching/pointing responses from babies to demonstrate that 17-month-olds understand others' false-beliefs in the context of referential communication. In their study, an agent pointed to one of two boxes, disclaiming that she put a *sefo i*n it and called a child to play with the sefo ("Shall we play with the sefo?"). They found that babies would reach for the non-referred-to box in the false-belief condition, whereas they would reach for the referred-to box in the true-belief condition, indicating their early belief understanding in the context of referential communication.

1.1.3. Is belief in the eye of the beholder? Non-mentalists vs. mentalists.

The fact that infants are not only sensitive to incongruences in others' belief-action patterns (as indicated by VoE), but moreover that they are capable of anticipating others' actions given their epistemic states (as indicated by AL), as well as properly responding themselves given what another had or had not seen (as indicated by spontaneous helping and reaching) provides compelling support for the early-competence view. However, it is still an open question whether what underlies these empirical findings converges to a uniform and specific ability to represent and attribute propositional attitudes.

A group of authors aim to explain early FBT successes by appeal to an entirely anti-mentalist view. Namely, according to this view, infants pass implicit FBTs by appeal to a set of rudimentary behavioral rules, such as *agents look for objects where they saw them last* (e.g. Perner, 2010; Perner & Ruffman, 2005; Ruffman, 2014). However, this view is insufficient to explain various implicit data, derived from the studies that go beyond non-verbal versions of the change-location FBTs; in fact, virtually any change to the traditional change-location FBT setup would require postulating an additional behavioral rule.[8] Hence, the behavioral-rules approach requires either that an infant possesses an astronomically large number of rules *a priori* (viz. inborn), or that she has to learn them at an exceedingly fast rate. Either way, this account seems implausible, given the scope of potential situational variations a child could possibly encounter (see Carruthers, 2013; Song & Baillargeon, 2008).

Relatively recently, a group of authors started favoring the dual-systems approach, which claims to offer a middle ground between the mentalist and the behavior-rules accounts of ToM (e.g. Apperly & Buterfill, 2009; Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Surtees, Apperly & Buterfill, 2012; Buterfill & Apperly, 2013; Low & Watts, 2013; Rakoczy, Bergfeld, Schwartz, & Fizke, 2014). They suggest that the discrepancy between young children's poor performance on the traditional FBTs and their successes on implicit measures could be accounted for by postulating two fundamentally distinct ToM systems. The full-blown, representational ToM is thought to emerge around the age of four, as reflected through the higher passing rate on the standard FBTs. The early, or minimal, ToM is responsible for infants' and young children's successes on implicit FBTs, yet it is essentially limited with respect to the kind of mental representation it employs. Namely, the minimal ToM does not represent genuine mental state concepts, but instead tracks others' minds through *belief-like* registrations (see Butterfill & Apperly, 2013). Unlike genuine mental state concepts, registrations are *non-cognitive* relations between agents, objects' locations and objects' properties and, as such, cannot account for *how* something is represented, but only for

[8] For instance, the behavioral-rules account would struggle to account for the findings by Scott and Baillargeon (2009), since they do not target false-beliefs about object location, but rather false-beliefs about object identity. Hence, the rule that states *agents look for the objects where they saw them last*, won't suffice to allow for the success on this task.

what is represented. Hence, minimal ToM would imply that if an agent registers X under description A, and X is also B, it follows that she registers X under the description B as well.[9] Because of the formal distinction between genuine mental state concepts and registrations, minimal ToM will suffice to allow for successes on certain kinds of implicit FBTs (i.e. those that involve false beliefs about objects' locations and properties), but will reveal a blind spot when it comes to the cases that involve identity false-beliefs.

This prediction of the dual-systems account was challenged by several studies, which demonstrated early understanding of identity false-beliefs (e.g. Buttelmann, et al., 2015, Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015; also Scott, Roby, & Setoh, 2020). For instance, Scott and Baillargeon (2009) found that 18-month-olds expected an agent to reach for an undesired toy, rather than a desired toy, if she falsely believed that the undesired toy was, in fact, the desired one. Similarly, in the helping study by Buttelmann et al. (2015), 18-month-olds helped an agent to get a desired object on the basis of the agent's false-belief about the identity of a deceptive object (e.g. a sponge that looks like a rock) that she was struggling to reach for. Finally, Scott et al. (2015) provided the evidence that not only can 17-month-olds understand identity false-beliefs, but they also understand an agent's intention to implement identity false-beliefs in others. Nonetheless, the advocates of the dual-systems approach argue that these studies fail to provide a compelling evidence of early identity-belief understanding, as they are based on restricted sets of stimuli, which, presumably, allow for more

[9] This is an instance of the well-known *Frege's Puzzle* (see Frege, 1948), which concerns identity statements and propositional attitude reports. It yields that co-referring terms, such as "Hesperus" and "Phosphorus" that refer to a numerically identical entity, cannot be substitutable *salva veritate* within a propositional content of the mental state reports.

parsimonious interpretations (Fizke et al., 2017; see also Butterfill & Apperly, 2013; Low, Apperly, Butterfill, & Rakoczy, 2016).

There have been few attempts to empirically demonstrate a "blind spot" in early identity false-belief reasoning (e.g. Low & Watts, 2013; Low, Drummond, Walmsley, & Wang, 2014; Fizke et al., 2017). One such attempt was made by Low and Watts (2013), who claimed to demonstrate that, whereas 3-year-olds were capable of passing an AL change-location FBT, they could not spontaneously anticipate where the agent will search for the toy in an identity FBT. Similarly, Fizke et al. (2017) employed a spontaneous helping paradigm to show that toddlers would take into account another's false-belief about a location of the target object, but would fail to do so if the false-belief involved object identity. Nonetheless, these studies appear insufficient to provide compelling evidence in favor of the dual-systems approach. In particular, they have been criticized for falling short of disentangling whether children's poor performance on identity FB tasks was caused by early ToM's representational limit, or was a consequence of performance demands extraneous to ToM reasoning.[10]

Taken together, the existing empirical studies are still far from offering conclusive evidence regarding the question of whether early ToM can represent identity false-beliefs (and consequently, of whether early ToM represents genuine propositional attitudes or, alternatively, employs non-mentalistic registrations). Now, considering the theoretical ground solely, it appears unclear what would make the dual-systems model advantageous

[10] For instance, apart from the demands on ToM reasoning, the experiment by Low and Watts (2013) significantly taxes subjects' working memory, involving mental rotation of the visual image of the object (see Carruthers, 2015). Similarly, the "dual-identity" objects used in Fizke et al. (2017) might have been puzzling for young children who, themselves, could have struggled to represent different aspects of an object as pertaining to a single identity.

over the assumption of a single ToM mechanism. As the advocates of this model posit, the primary motivation lies in the "quest" for parsimony: presumably, propositional attitudes are too complex to be computed within the early ToM system, especially given that this system ought to operate efficiently to account for the rich dynamics of social interactions (see Butterfill & Apperly, 2013; Apperly & Butterfill, 2009). Although this appears as a valid theoretical motive, it is unclear how parsimony is established by introducing an additional ToM construct—viz. the "Minimal ToM"—apart from the late ToM, which itself still remains to be explained.[11] As a matter of fact, the very conceptualization of the "minimal ToM" appears overly complicated, as it rests on a set of new (and somewhat vague) notions, each of which require additional postulates to be properly defined. For instance, Butterfill and Apperly explain early successes on FBTs by introducing *registrations.* However, this basic notion of *registrations* requires a set of additional principles and notions, such as the *goal directed action as a unit of bodily movements directed to goals, encountering* an object within the agent's *field,* conditions that would determine *success of a goal directed action,* etc. (for the full discussion of these principles, see Butterfill & Apperly, 2013). For comparison, to account for the same phenomenon (viz. early FBT successes) Leslie's ToMM model requires a single postulate: the primitive "belief concept" offered by the fundamental computational structure of the mechanism (Leslie, 1987; 1994a).

It appears, therefore, that "minimal ToM" lacks sufficient conceptual, as well as empirical power to be taken as a favored model of early successes on FBTs. However,

[11] It is unclear, for instance, what exactly contributes to the emergence of the full-blown ToM around the age of four. The proponents of the dual-systems approach appear to embrace the assumption that language development presumably plays a role in the formation of the late ToM system. Nonetheless, this assumption remains vague and largely underexplained.

whether the form of the underlying mental representation employed on FBTs prior to the age of 4, indeed, exhibits the key properties of propositional attitude reports (viz. the property of referential opacity) and whether it is, consequently, capable of accounting for identity false-beliefs is still an empirical question. In this dissertation, I propose a new way of exploring this question, by appeal to *proper names* as indexes of particular individuals' identities. Specifically, I ask the following question: can a young child learn whom a new proper name refers to solely by virtue of attending to another's false-belief about the referent's identity? This leads us to the second major postulate of this dissertation, which is that ToM presents one of the key components in the process of mapping new labels to their referents. I discuss this idea in the next section.

1.2. ToM as a vehicle for mapping words and their referents

In this part, I explore the relationship between metarepresentational ToM and word learning. In particular, I propose the idea that meta-representational ToM (in the sense of Leslie's ToMM) is a critical element that is required for young learners to narrow down the space of a label's possible word meanings, in the process of word learning. I will first review the main approaches to the word learning process, after which I will focus on the role of ToM in this process.

1.2.1. Dominant approaches to the word-learning process.

The acquisition of new words is an enormously complex process and it requires effective tools for solving the problem of referential ambiguity. A novel label may refer to infinitely many things[12] and figuring out the correct meaning-referent link presents a daunting task especially given the complexity of the situational contexts in which the first words are usually encountered (see Gleitman & Trueswell, 2018). Nevertheless, young children are able to master it in a strikingly effortless and efficient manner. At about 16 months, they start to demonstrate an astonishing talent for word learning and are capable of acquiring up to 20 new words per day (e.g. Bloom, 2000). What underlies this capacity has been the focus of large debates among cognitive scientists, yet the consensus hasn't been reached yet.

A common approach to the word learning process posits that it is in virtue of a general, associative learning mechanism that children form links between words and meanings (e.g. Richards & Goldfarb, 1986; Smith, 2000; Yu, 2008; Smith, Jayaraman, Clerkin, & Yu, 2018; Saffran, & Kirkham, 2018). This associationists perspective assumes that the early vocabulary growth relies mainly upon statistical learning of co-occurrences between labels and their referents across different settings. According to this account, young learners store multiple hypotheses of what a novel label refers to when they first encounter it; as they re-encounter the same label across different contexts, they update the weights of these concurrent hypotheses, eventually rendering one of them the most likely referent, hence, establishing the label-meaning link.

Several researchers have attempted to demonstrate the probabilistic and gradual nature of word learning in laboratory by presenting the subjects with multiple object-

[12] This problem echoes Quine's (2013) doctrine of the indeterminacy of translation, which holds that there cannot be a way to fully translate the meaning of a word from one language to another, as there might be multiple, *equally correct,* yet *semantically distinct* translations of the same utterance. Applied to the problem of early word learning, suppose that a young, pre-verbal child sees an image of a rabbit and hears her mother providing the linguistic input *("A rabbit!")*; the question arises how the child understands what "rabbit" means, given the large scope of potential referents she gathers simply from facing the picture of a rabbit (viz. "rabbit" can refer to the picture as a whole, to the animal's ears, fur, etc.).

images on the white background and pairing them with novel words settings (e.g. Yu & Smith, 2007; Smith & Yu, 2008; Vouloumanos, 2008). Although they seemed to successfully demonstrate that subjects learned word-referent pairings in virtue of tracking the patterns of their co-occurrences across multiple trials, the ecological validity of these studies has been problematic, as they oversimplify the complexity of the actual word learning contexts. Namely, the actual word learning setting assumes that a young learner encounters an enormously large number of potential referents across environments full of noise and variance (see Trueswell, Lin, Armstrong, Cartmill, Goldin-Meadow, & Gleitman, 2016). Given that hundreds of potential referents are in view when a learner hears a new label, storing all the word-meaning hypotheses for the future updating would dramatically exceed the capacities of the human memory system. Another problem with these laboratory demonstrations of cross-situational statistical learning is that they mainly relied on identical images (instances) of the same word. Medina et al. (2011) pointed out that this method neglected the variability accompanied with the actual word learning process. For instance, a young learner sometimes hears the word "cat" when there is a partially occluded cat in her visual field, or when there is an orange cat, or a drawing of a cat, or when there is *no* cat presented whatsoever ("*We are going to feed the cat.")*; hence, repeating the word "cat" cross-contextually would (if the statistical learning paradigm was correct) in fact increase the hypotheses space of possible meanings, rather than narrowing it down (Medina, et al., 2011).

The alternative comes from a group of authors who suggest the "propose-but-verify" model, according to which a learner immediately forms a single hypothesis about the meaning of a word when she first encounters it (e.g. Medina et al., 2011).

Importantly, it is only this hypothesis that is stored and which will be evaluated against the future evidence, which enables the learner to rapidly select the most likely candidate, without exceeding her limited memory capacities. This model has been supported by a number of compelling studies (Medina et al*.,* 2011; Woodard, Gleitman, & Trueswell, 2016; Trueswell, Medina, Hafri, & Gleitman, 2013; Trueswell, et al., 2016). For instance, Medina et al. (2011) relied on the Human Simulation Paradigm (HSP; Gillette, Gleitman, Gleitman, & Lederer, 1999) to test adults' resolution of referential ambiguity. They exposed the subjects to a set of short videos of toddler-parent interactions, where a parent would utter a "mystery word" to their toddler and the subjects were to guess the meaning of the word. Importantly, the videos differ in respect to referential clarity[13], so for each mystery word participants would see a single high-informative video, and four low-informative videos, the order of their presentation being manipulated. The pattern of results suggested that participants were, indeed, forming a single hypothesis about the word's meaning and were retaining this hypothesis until it was disconfirmed, instead of storing and updating multiple simultaneous hypotheses. For instance, if their guess was wrong on the first trial, their performance on the subsequent trials would not steadily improve, as predicted by the cross-contextual statistical learning (Medina et al., *Ibid*).

Another study by Woodard et al. (2016) demonstrated that 2- and 3-year olds rely on the same strategy. They first presented children with pairs of pictures, showing unfamiliar animals and asked them to point to "dax". In the next phase, children either

[13] In the first experiment, Medina et al. (2011) explored the referential clarity of the natural contexts in which the word learning takes place. They asked a group of participants to guess a "mystery word" from the videos, and used their accuracy to classify the videos based on their informativeness. Only a small percentage (about 7%) of videos were classified as high informative (i.e. above 50% of accuracy in guessing the "mystery word"), whereas most of the videos were classified as "referential junk" (i.e. bellow 33% accuracy in guessing the "mystery word"). (Medina et al., 2011).

saw the animal they previously selected as “dax” together with a novel one, or they saw the animal they previously had not selected as “dax” alongside with the a new one, and were, again, asked to point to “dax”. Children would point to the previously selected animal when it reappeared but were at chance when the non-selected animal reappeared alongside with the new one (Woodard et al., 2016). This indicates that when a new label is first introduced, children rapidly form a *single* hypothesis of what it refers to and use the later occurrences to either support or reject this hypothesis. However, alternatives are not stored and, hence, are not available for the future updating, contrary to what statistical learning models assume.

An important virtue of the “propose-but-verify” model is that it allows for a rapid pace of the word learning process (contrasting the gradual and slow progression, implied by the cross-contextual statistical models), which is a key constraint word-learning models need to account for. After all, a young learner only has a couple of years to master the vocabulary of her language; moreover, young children typically exhibit striking proficiency with their language by the age of 3 (e.g. Bloom, 2000). As the model suggests, an optimal learning process, given the time constraint, would be the one that enables a learner to immediately form a referent-word hypothesis, and evaluate only this hypothesis against the evidence; in the case of disconfirming evidence, the hypothesis instantly gets discarded, and a new candidate gets selected. In addition, the model takes into account another important constraint—human mind’s limited memory-attentional resources—as it assumes that the learner does not need to store and update numerous simultaneous hypotheses at the same time, but rather stores a single most likely referent of the word. The question arises, however, of how exactly the learner selects the most likely referent when she encounters the word for the first time. This question becomes especially pressing when considering the amount of noise that is typically associated with the natural learning contexts (e.g. Medina et al., 2011).

1.2.2. The role of ToM in selecting the most likely referent of a novel label.

There have been a number of authors exploring what makes a certain context highly informative for a young word learner. Many of them pointed out the critical role of social-attentive cues, emphasizing inherently social nature of language (e.g. Baldwin, 1991, 1993a; 1993; Baldwin, Markman, Bill, Desjardins, & Irwin, 1996; Baldwin & Tomasello, 1998; Bruner, 1974/1975; Bloom, 2000; Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, & Trueswell, 2013; Gleitman & Trueswell, 2018; Tomasello & Kruger, 1992; Koening & Echols, 2003; see also Papafragou, Friedberg, & Cohen, 2017).

For instance, Baldwin (1991) famously demonstrated that infants in the second year of life would track what the speaker attends to (i.e. what their eye-gaze directs at, what their head is oriented towards, etc.) to figure out the referent of a novel word. Hence, they would pair a word with what the speaker was attending to when she uttered it, rather than with what was more directly dominating their own perceptual space (e.g. a toy in a bucket right in front of them). Baldwin et al. (1996) extended these findings as they showed that 18- to 20-month old children formed an object-label link *only* when the speaker utters the label in the full view of the child, suggesting the critical role of agents' social cues in referential disambiguation. Similar findings come from Koening and Echols (2003), who showed that 16-month-olds children were sensitive to an agent's knowledge states (based on their perceptual access to the scene) in the true/false labeling events. The importance of social-attentive cues was demonstrated in Cartmill et al. (2013)

as well, who relied on the HSP to demonstrate that both quantity and quality (i.e. overt and frequent usage of social-attentive cues) of the talk predicts children's vocabulary size three years later. In another study, Trueswell et al. (2016) altered the temporal aspects of the social-attentive cues to the reference in HSP videos, to show the key role of their precise timing for referential clarity.

The fact that social-attentive cues have the key role in the process of forming a word-meaning hypothesis is in line with Grice's (1969) intention based semantics. Grice grounds semantic content in the process of mutual intention recognition among interlocutors in a communicative setting: by performing a certain utterance, the speaker intends for their interlocutor to believe that the utterance refers to something, in virtue of the interlocutor's recognition of this intention. According to this perspective, for establishing the word-meaning link, one ought to take into account the speaker's mental states in the process of word acquisition. Notably, although the role of social-attentive cues—such as the sensitivity to others' eye-gaze, joint attention, etc.—has been demonstrated as a key factor in the process of resolving the referential ambiguity problem, it is still not clear whether young word learners take into account others' *epistemic states* (e.g. beliefs) about referents of novel labels in the process of referential disambiguation. In other words, the question of interest here is the role of meta-representational ToM in the process of narrowing down a young learner's hypothesis space of a novel's words possible meanings. Here, we postulate that ToM makes a critical contribution in establishing what a novel word might possibly refer to and that the most plausible hypothesis that a word learner might form is that the word means *what the speaker believes it to refer to.*

There have been several studies exploring young children's performance on the FBTs in a communicative setting, involving the aspect of word learning (e.g. Carpenter, Call, & Tomasello, 2002; Happé & Loth, 2002; Southgate, Chevallier, & Csibra, 2010). For instance, Happé and Loth (2002) tested 3- to 5-year-old children on the standard FBT and a special "word-learning" FBT, in which children were required to assign the reference of a novel label in virtue of recognizing the speaker's false-belief about the content of a box. They found that children's false-belief reasoning was enhanced in the word-learning context, compared to their performance on the traditional FBT. Similarly, Carpenter et al. (2002) found the evidence that a communicative setting— viz. reasoning about others' communicative intentions—helped 36-month-olds to account for another's false-belief. Southgate et al. (2010) found a similar improvement in false-belief reasoning in even younger, 17-month-old children.

These studies suggest that communication might be a privileged domain of ToM reasoning, i.e. that children might more readily account for others' false-beliefs if they are placed in the context of communicating information between social agents (e.g. Grice, 1969; Roth & Leslie, 1991; Southgate et al, 2010; also see Sperber, 2000; 2002). Note, however, that the focal question here is that of the actual role of ToM in the process of word learning. The studies cited above do not truly capture this question for several reasons. In Happé and Loth (*ibid*), for instance, children could have used ownership to figure out the referent of a new label rather than her false-belief about the referent of the label.[14] Similarly, in Carpenter et al. (*ibid*) and Southgate et al. (*ibid*) there was no

[14] In this study, the agent (A1) puts her own toy in a box and leaves, and then another agent (A2) replaces A1's toy with their own toy. A1, then, comes back and refers to the toy in the box with a new label. The problem with this design is that children could simply associate the target toy with A1 (*A1's toy*) and use this association to solve the task, rather than relying on A1's epistemic state to figure out the referent of the

difference between a “new-word” condition (e.g. where the agent points to the box and refers to the object inside it as *a sefo*, before inviting the child to play with the sefo/help them find a sefo) and a “generic expression” condition (e.g. where the agent points to the box and refers to the object inside it in a generic way, as “it” or “a toy”, before inviting the child to play with it/help them find it). So, these studies, in fact, do not answer the question of whether ToM reasoning (specifically, epistemic states attribution) plays an important role in children’s learning of a new word.

Our key assumption is that meta-representational ToM is already available to young children not only for the purposes of interpreting and/or predicting other social agents’ behaviors, but also to enable them to efficiently receive novel information from them—specifically, meanings of their communicative acts. Upon this assumption, when a learner encounters a new label, a number of lexical constraints are first applied to narrow down the space of possible meanings, such as the mutual exclusivity principle (tendency to interpret a label as referring to an object that does not already have a label) and the whole object bias (tendency to interpret a label as referring to the whole object, rather than its parts)

We propose that the additional element that is needed to guide the word-meaning label is the meta-representational ToM, which takes into account the speaker’s intentional and epistemic states to discern the meaning of a label. Following Gricean view, in order to figure out the meaning of an utterance (e.g. a label), the learner has to employ ToM to infer the complex intentional structure of the speaker: 1) the speaker wants to refer to “X” by uttering “Y” (referential intention) and that 2) the speaker wants the learner to believe

label (for the similar criticism see Houston-Price, Goddard, Séclier, Grant, Reid, Boyden & Williams, 2011).

that “Y” means “X” (communicative intention). What is important to notice here is that in order to discern what a new label refers to, a young learner has to take into account both the speakers’ intentional states and her *epistemic* states. In other words, the learner needs to recognize that when the speaker utters a new label, she *intends* to refer to a certain entity that she *believes* to be the referent of this label. Upon this account, the candidate referent of a label is determined not only by what the speaker highlights through social-attentive cues, such as eye-gaze, referential pointing, etc. (Baldwin, 1991; Baldwin & Tomasello, 1998; Tomasello & Kruger, 1991), but also by recognizing her beliefs about the referent. Hence, the suggestion is that a young learner is a profound mentalist in the process of word learning, as she consults a decoupled, meta-representational context, apart from the overt first-order representation, in the process of selecting the word’s referent. If this is correct, then a young learner should be able to understand that when Sally points to a doll-A and refers to her as *Ann,* it is actually another identical doll (doll-B) whose name is Ann—Sally was mistaken when uttering the name, as she (falsely) believed of the doll-A that she was the doll-B.

If a young learner takes into account the speaker’s beliefs about whom the label refers to in the process of word learning, this entails that the learner can already represent the speaker’s identity false-beliefs. This brings us back to the question raised in the previous section, which concerns the nature of early ToM and its development. If the fundamental meta-representational structure of ToM is innately specified, it should allow for early representations of propositional attitudes that involve individuals’ identities as their propositional content (the developmental continuity hypothesis, see Leslie, 1987); by contrast, if the meta-representational structure of ToM emerges only after the child turns four, she should be “blind” to the scenarios that involve identity false-beliefs (the dual systems hypothesis, see Butterfill & Apperly, 2013).

In the next part, I turn to the study that tests the two main questions of this dissertation:

1) Can children younger than four represent identity false-beliefs?

2) (If so) can children younger than four select the referent of a novel word by correcting for the agent’s identity false-belief?

Affirmative answer to the first question would support developmental continuity of ToM competence, and affirmative answer to the second question would indicate its important role in the process of word acquisition (suggesting a profound mentalism involved in this process). To test these questions, we developed a new, naming version of the FBT, which presents a young child with a story in which the speaker uses a new label to refer to a certain object that she *falsely believes* is in a box (the intended referent), when in fact it contains a different object (the unintended referent). At the end of the task, we ask children to point to the referent of the label. To figure out the correct referent-label mapping, children have to rely solely on the speaker’s false-belief about the referent, rendering the one she *intended* to refer to (*not* the one she *actually* referred to when she first uttered the label) the correct response. If young children succeeded on such a task, this would indicate the key role of ToM in the process of word acquisition and would suggest that false-belief reasoning is already available prior to the age of four, not only for prediction/explanation of others’ actions, but to enhance the acquisition of relevant information in other domains of knowledge as well.

Chapter 2

The Study

This study includes a set experiments in which we explore whether meta-representational ToM underlies the process of mapping a new word onto its referent in young 3-year-olds and 2-year-olds. As highlighted in the previous part, demonstrating that a young child represents a speaker's false-belief about the referent's identity and uses this representation to discern the word-referent link in a situation that does not explicitly require her to think about the agent's actions[15] bears important implications both for ToM development and for early word learning.

In the first three experiments, we focus on the case of *proper names*, to answer the question of whether ToM can represent identity false-beliefs prior to the age of four. Proper names are particularly relevant for this question as they mark individuals' identities rather than category-based commonalities. Hence, they allow us to create a scenario in which an agent has a false-belief about *whom* she is indexing by uttering a certain proper name (e.g. *"It is Fido in the box",* when it is *not* Fido in the box). To succeed on such a task, children have to account for the speaker's identity false-belief and understand that she was representing a different individual's identity as the target of the name from the one she actually referred to. The fourth experiment extends the findings from Experiments I, II, and III beyond the case of proper names, and focuses on the role of ToM in selecting a correct referent of a new common noun. Together, these

[15] Word learning does not necessarily require prediction/explanation of agents' actions, which are considered traditional roles of ToM.

experiments allow us to test the hypothesis that meta-representational ToM is a key element that guides young learners in the process of mapping new words onto their referents.

2.1. Experiment I

Three-year-olds' performance on a proper names learning FBT

We test whether young three-year-olds can map a proper name and its referent through the speaker's identity false-belief. Since proper names single out particular individuals' identities (see Kripke, 1980; Putnam, 1975, but also Russell, 1905) they allow us to probe whether early ToM can represent identity false-beliefs. Recall that the advocates of the "minimal ToM" account claim that early ToM system (prior to the age of four) is not meta-representational, hence it is subject to a series of "signature-limits", such as the incapacity to account for the cases that involve identity false-beliefs (e.g. Butterfill & Apperly, 2013). Showing that a young child could entertain another's false-belief about an individual's identity to discern the referent of a novel proper name would, thus, challenge the assumption that identity-beliefs present a "blind spot" of the early ToM. Instead, it would support the claim that early ToM involved in the process of referent-name mapping is, indeed, meta-representational and that it allows for attribution of full-blown mental states (i.e. propositional attitudes).

Previous studies on lexical development have provided evidence that children start understanding the nature of both proper names and common nouns rather early on in life (e.g. Hall, 2009; Macnamara, 1982; Nelson, 1973; Tincoff & Jusczyk, 1999; Tincoff & Jusczyk, 2000). For instance, in a series of looking time experiments, Tincoff and

Jusczyk showed that 6-month-olds would understand that certain proper names (specifically, “Mommy”/”Daddy”) apply exclusively to specific individuals (Tincoff & Jusczyk, 1999), whereas they would expect that common nouns (specifically “feet”/”hands”) exhibit cross-exemplar generalization (Tincoff & Jusczyk, 2000). Furthermore, several studies showed that children in their second year of life understand that *novel* proper names (e.g. “Daxy”) refer to specific individuals rather than to a category based commonality (Katz, Baker, & Macnamara, 1974) and that this understanding perseveres even if the potential referents are property identical (Bélanger & Hall, 2006; Hall, Lee, & Bélanger, 2001).

Here, we tested whether young children can map a referent and its proper name by correcting for the speaker’s false belief about the referent’s identity. We explored this question with 3-year-olds that we presented with the naming FBT in which an agent had a false-belief about the identity of a dog hidden in a box, when she named it, “Fido”. The goal was to see if children would be able to figure out whom “Fido” refers to solely by correcting for the speaker’s identity false-belief. If young children succeeded, this would indicate developmental continuity of meta-representational ToM competence and would, at the same time, imply its critical role in word acquisition.

2.1.1. Method

Participants

Participants were 25 English-speaking three-year-old children who ranged from 36- to 47-months (M = 41 months, SD = 3.59 months; 8 female). One additional child was tested, but excluded from the sample, since they failed to cooperate.

Stimuli and Procedure

Children were tested individually, in a quiet environment (either at their preschools or in the lab). Each child was presented with an illustrated story, displayed on a laptop computer through Microsoft Power Point. The story was advanced by an experimenter who was changing the slides on the screen by pressing a button on the laptop's keyboard. The child was seated in front of the laptop, facing the screen and the experimenter was seated on the floor, next to the child. Each slide of the story was verbally narrated by the experimenter (the same experimenter was running the study for all the subjects).

The story unfolded around a girl, Jane, who had two identical puppies. Since one of the puppies was tired, Jane put him in a box to sleep, closed the lid and left. In her absence, this puppy jumped out of the box and ran away from the scene, and then the other one jumped into the box (the lid of the box closes subsequently). At this point, children were asked the "See" control question ("Did Jane see what just happened?"). If the child answered incorrectly, the experimenter would correct them and start the story again. If the child answered correctly, the experimenter would proceed with the story, showing Jane returning with her brother who, then, pointed to the box and asked who was inside it. Jane responded, "Fido is in the box! I put Fido in the box to sleep!" after which the siblings left the scene. In the test scene, the first dog (Jane's intended referent) returns and stands next to the box, and the other dog (whom Jane actually referred to) pops out of the box and remains sitting in it. Subjects were then asked, "Can you point to Fido?" (for schematics of the stimuli see Figure 1).

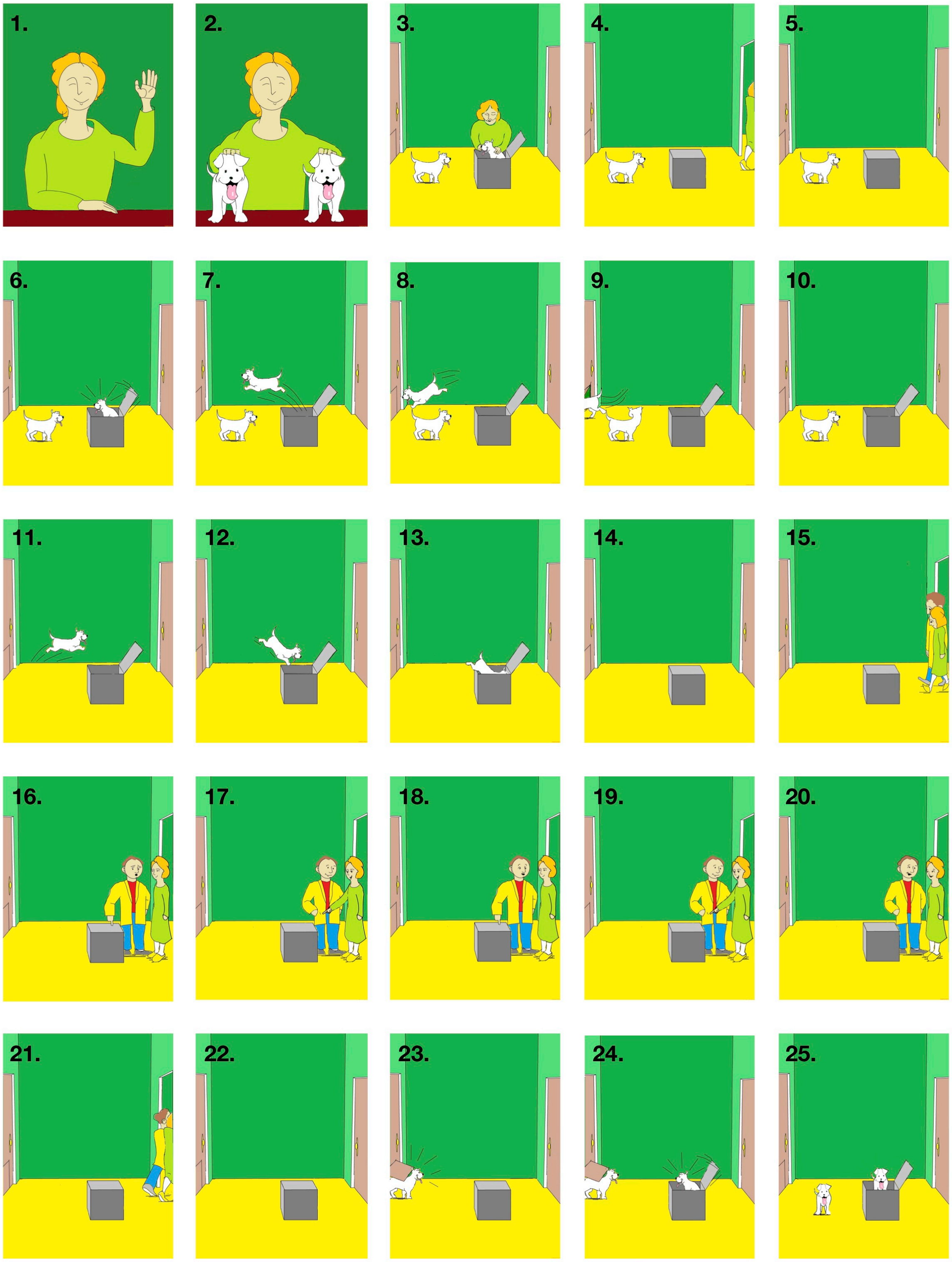
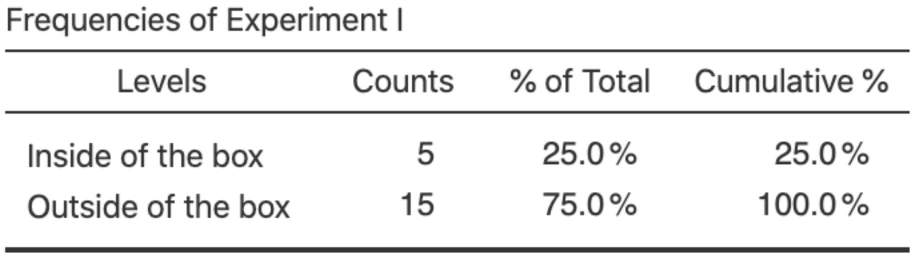


Figure 1. Schematics of the stimuli used in Experiment I and Experiment II (False-Belief).

2.1.2. Results and Discussion

Out of 25 tested children, 5 failed the control “See” question, hence we focused the main analyses on 20 children who passed this question (M = 41 months, SD = 3.41 months, 6 female). Children’s responses are summarized in Table 1. We found that 15 three-year-olds (75%) pointed to the dog outside of the box, that is, to Jane’s intended referent (Binomial *p* = .04, two-tailed). Bayesian analysis revealed moderate evidence in favor of the alternative over the null hypothesis (BF = 3.22 in favor of H1)[16] (Figure 2).



Frequencies of Experiment I

Levels	Counts	% of Total	Cumulative %
Inside of the box	5	25.0 %	25.0 %
Outside of the box	15	75.0 %	100.0 %

Table 1. Number of children pointing to the dog outside of the box vs. the dog inside of the box in Experiment I.

[16] The classification scheme for Bayes factors was based on Lee and Wagenmakers (2013, p. 105), adjusted from Jeffreys (1961).

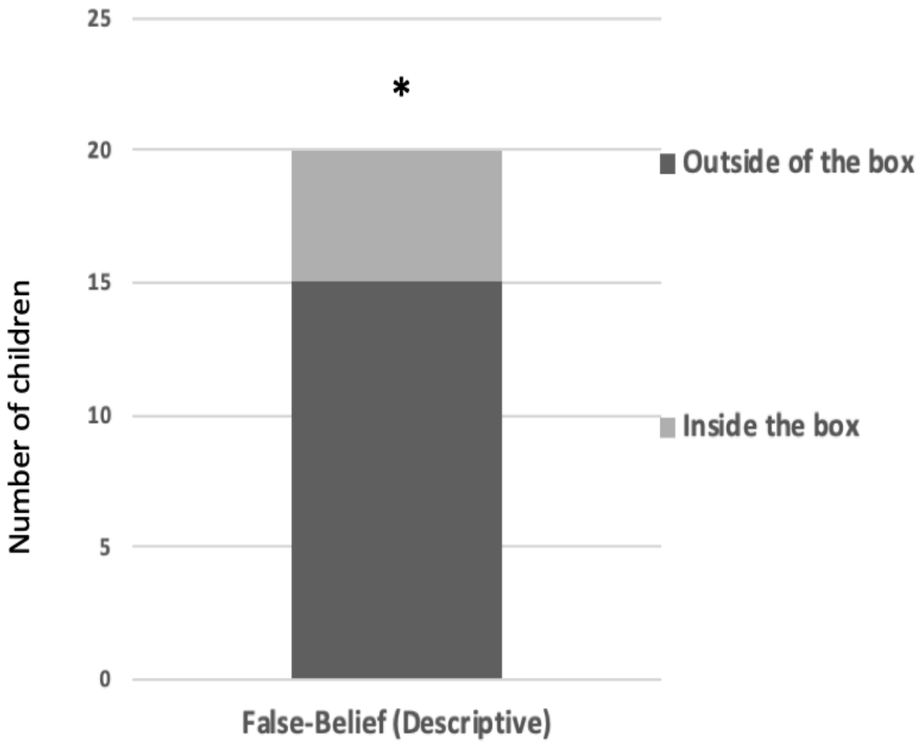


Figure 2. Number of children pointing to each of the potential referents in Experiment 1. * = p < .05, two-tailed.

The fact that majority of 3-year-olds successfully mapped the proper name “Fido” with its referent (that is, with the dog that the speaker had in mind as the referent when she provided the name for the first time) indicates that children took into account the speaker’s false-belief about the referent’s identity, together with her referential intention, in order to solve the referential ambiguity problem (i.e. to figure out who Fido is). This supports the idea that young children rely on ToM reasoning in the process of word acquisition, as they attend to the speaker’s (false)belief to select the most likely referent of a new name. In addition, the results suggest that children understand that someone can hold a false-belief about an individual's identity ("*The one that is in the box is Fido.*") prior to the age of four, which contradicts the minimal ToM approach's prediction that early ToM will reveal a "blind spot" in identity false-belief scenarios (Butterfill & Apperly*,* 2013). In contrast, the results suggest developmental continuity of ToM competence (Leslie, 1987, 1994a, 1994b).

However, there is a possibility that children inferred the referent of "Fido" through Jane's description, rather than through her identity false-belief. Namely, what Jane said in the naming scene (when the proper name was introduced for the first time) was, *"Fido is in the box! I put Fido in the box for sleep!".* Thus, children could have used this description ("*Fido" = the one that Jane put in the box for sleep*) to track the referent of the proper name, without engaging in ToM reasoning at all. The aim of the second experiment was to rule out this possibility and to replicate the findings from Experiment I on a larger sample.

We also included an additional condition in which Jane had a true-belief about the identity of the intended referent, to control for potential low-level explanations of children's pointing responses (e.g. preferential factors, side bias).

2.2. Experiment II

Replication of Experiment I on a larger sample and with additional controls

The goal of this experiment was to replicate findings from Experiment I, controlling for a potential influence of the description that the agent provided during the naming scene, thus reassuring that children could pass the task solely by accounting for the agent's identity false-belief. We also included the true-belied (TB) condition in which the agent held a true-belief about the identity of the dog that she referred to in the naming scene. We predicted that if young three-year-olds, indeed, engage in false-belief reasoning to discern the referent of a proper name, they will show the same pattern of responses as children in Experiment I, when tested on the FB condition. Specifically, they will tend to point to the dog that was outside of the box in the naming scene (when the agent points to the box and says, "Fido is in the box!"). By contrast, children should show the opposite pattern of responses in the TB condition, where the agent's intended referent matches the dog that was *actually* in the box at the time of naming. In this condition, they should show a tendency to point to the dog in the box when asked if they can point to Fido.

2.2.1. Method

Participants

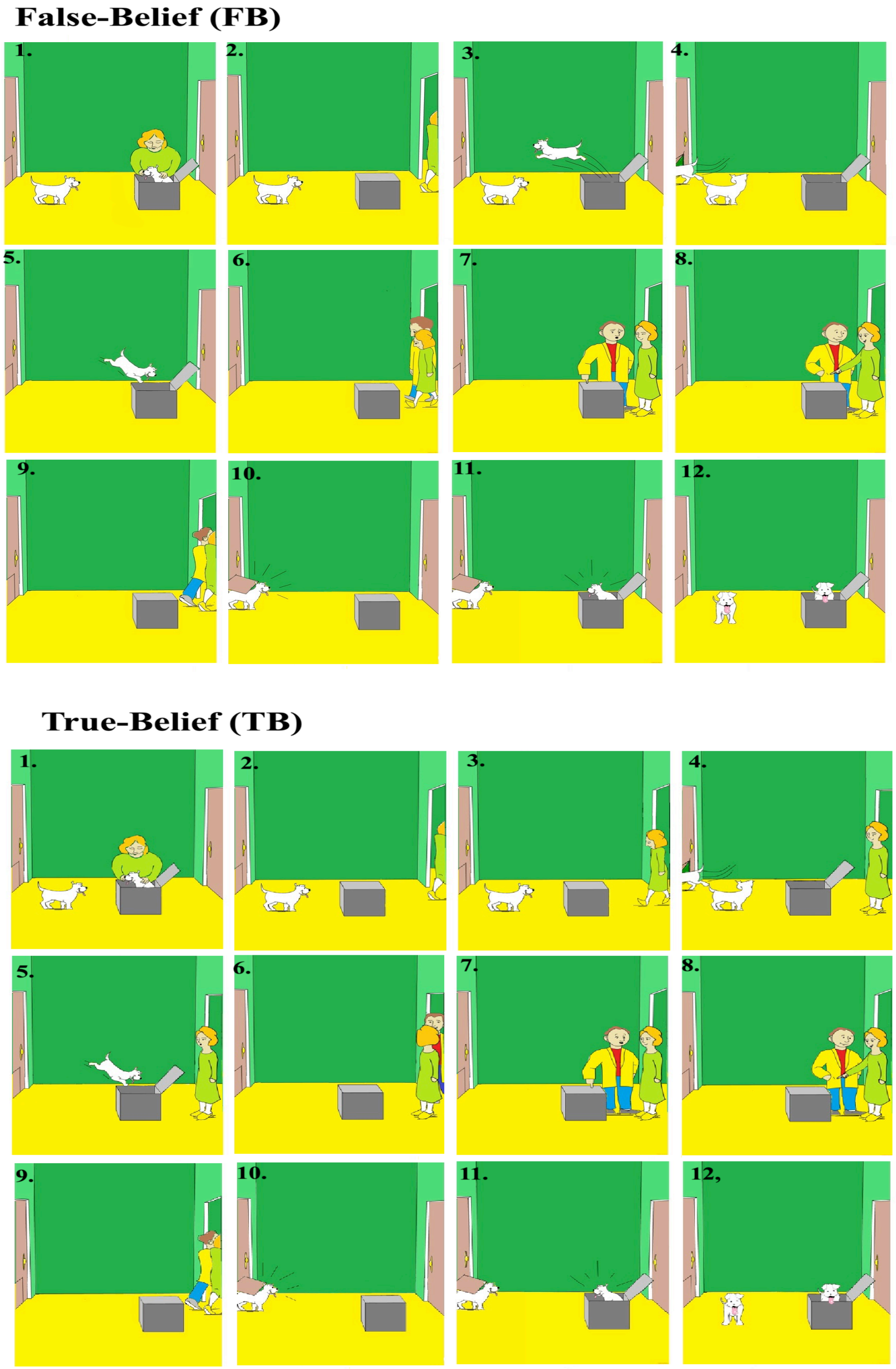
We tested a new group of 88 three-year-old children, ranging from 35- to 48-months (M = 40 months, SD = 3.6 months, 48 female). Nine additional children were tested but excluded from the analyses because they failed to cooperate (n = 2), did not produce/comprehend English (n = 2), parental interference (n = 1) and experimenter error (n = 4).

Out of 88 children, 48 were randomly assigned to the False-Belief (FB) condition (M = 40 months, SD = 3.5 months, 25 female), and 40 were assigned to the True-Belief (TB) condition (M = 41 months, SD = 3.6 months, 23 female). Again, children were tested individually, in a quiet environment (either at their preschools or in the lab).

Stimuli and Procedure

False-Belief (FB) condition. The stimuli and the procedure in the FB condition were the same as in Experiment I, but for one key difference: in the naming scene, when the proper name "Fido" is introduced for the first time, we dropped the description "*I put Fido in the box to sleep*" from the script. So, following her brother's question, "Who is in this box?", Jane only says, "*Fido is in the box!",* without providing any potential cues to the referent's identity.

True-Belief (TB) condition. The stimuli and the procedure in this condition were the same as in the FB condition, but for one key difference: Jane returns to the room before the dogs switch their places and observes the following sequence of events, thus ending up with a true-belief about the identity of the dog in the box (see Figure 3).



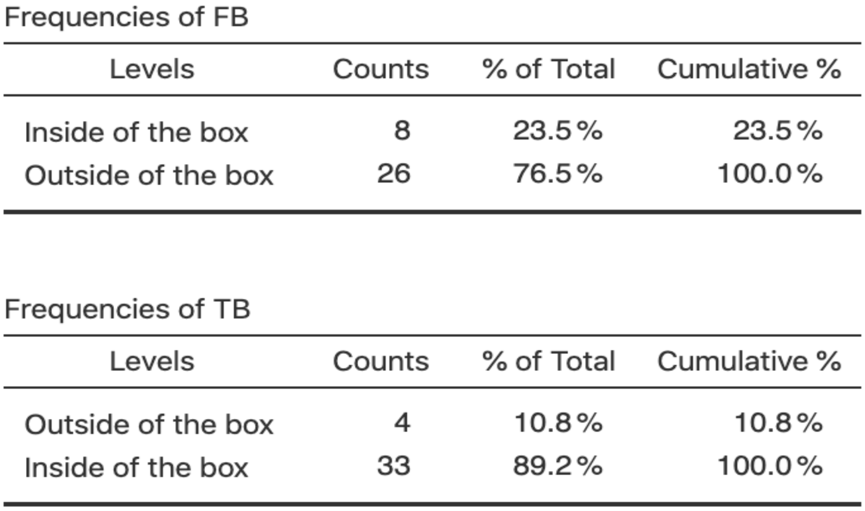
False-Belief (FB)

True-Belief (TB)

Figure 3. Schematics of the stimuli used in experiment II. The figure includes the key events. The panel on the bottom shows the TB condition, where Jane returns to the scene before the dogs switch their places.

2.2.2. Results and Discussion

Out of 48 children in the FB condition, 14 failed the control “See question” and out of 40 children in the TB condition 3 failed this question. The following analyses, thus, exclude these children from the sample. Children’s responses are summarized in Table 2. We found that 26 out of 34 (76.5 %) three-year-olds in the FB condition pointed to the correct referent—the dog outside of the box at the time of naming and Jane’s intended referent (Binomial p = .003, two-tailed; BF = 27 in favor of H1). By contrast, in the TB condition, 33 out of 37 (89.2 %) children pointed to the dog inside of the box at the time of naming—Jane’s intended and the actual referent (Binomial *p* < .001, two-tailed, BF = 54762 in favor of H1). Children’s pattern of responses differed significantly between the FB and TB conditions (Fisher’s exact test, *p* < .001, two-tailed) (Figure 4).



Frequencies of FB

Levels	Counts	% of Total	Cumulative %
Inside of the box	8	23.5 %	23.5 %
Outside of the box	26	76.5 %	100.0 %

Frequencies of TB

Levels	Counts	% of Total	Cumulative %
Outside of the box	4	10.8 %	10.8 %
Inside of the box	33	89.2 %	100.0 %

Table 2. Number of children pointing to the dog inside of the box vs. outside of the box in the False-Belief condition (FB) and True-Belief condition (TB), Experiment II.

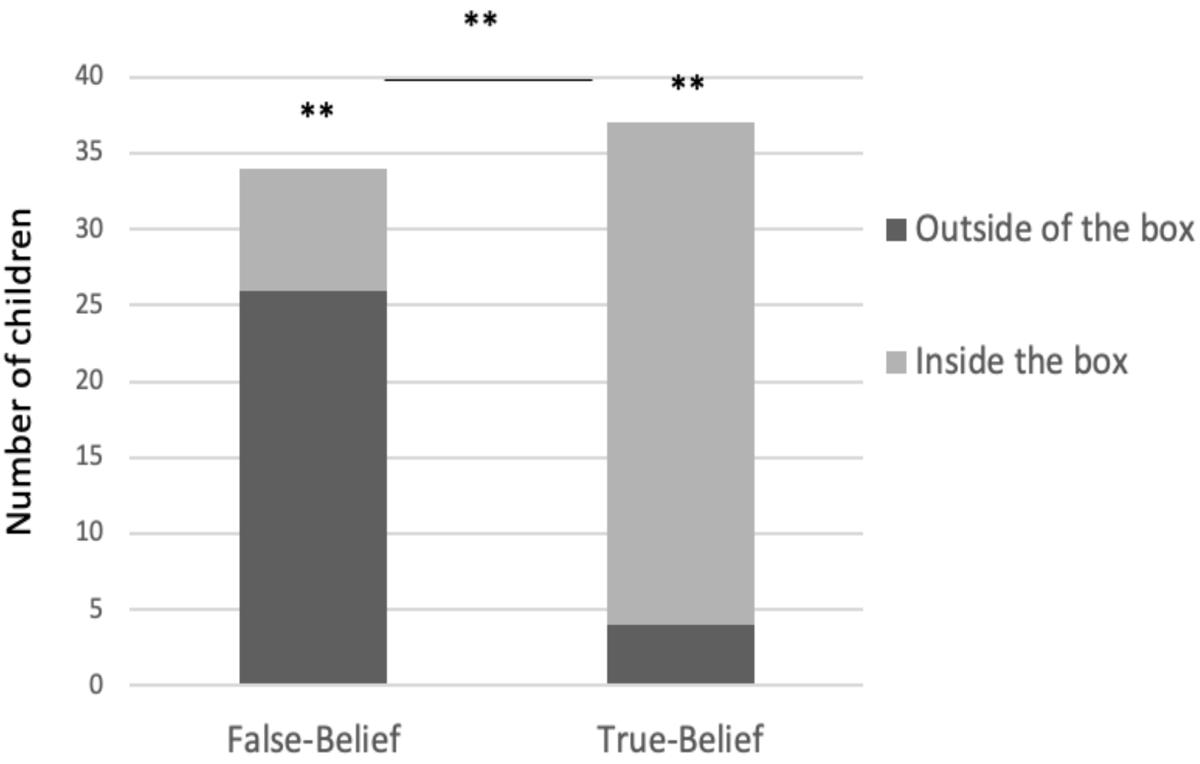


Figure 4. Number of children pointing to the dog outside of the box (correct in False Belief condition) and to the dog inside of the box (correct in True-Belief condition), excluding those who failed the "See" question. ** = p <.01, two-tailed.

We also analyzed the data including those children who failed the "See" question[17] and obtained the same pattern of results. Children's responses are summarized in Table 3. In the FB condition, 32/48 children (66.7 %) pointed to the dog outside of the box at the time of naming—the correct referent (Binomial *p* = .029, two-tailed, BF = 2.55 in favor of H1), whereas in the TB condition 34/40 children (85 %) pointed to the dog

[17] Within the 14 children who failed the "see" question in the FB condition, there was no significant trend towards pointing to either of the possible referents (6/14 children pointed to the dog outside of the box, Binomial *p* = 0.8, two-tailed).

inside of the box at the time of naming—the correct referent (Binomial *p* < .001, two-tailed, BF = 6986.6 in favor of H1); the difference between FB and TB conditions was significant (Fisher's exact test, *p* < .001, two-tailed) (Figure 5).

Frequencies of FB

Levels	Counts	% of Total	Cumulative %
Inside of the box	16	33.3 %	33.3 %
Outside of the box	32	66.7 %	100.0 %

Frequencies of TB

Levels	Counts	% of Total	Cumulative %
Outside of the box	6	15.0 %	15.0 %
Inside of the box	34	85.0 %	100.0 %

Table 3. Children's pointing responses to the dog inside of the box vs. the dog outside of the box in the FB and TB conditions, including children who failed the "see" control question, Experiment II.

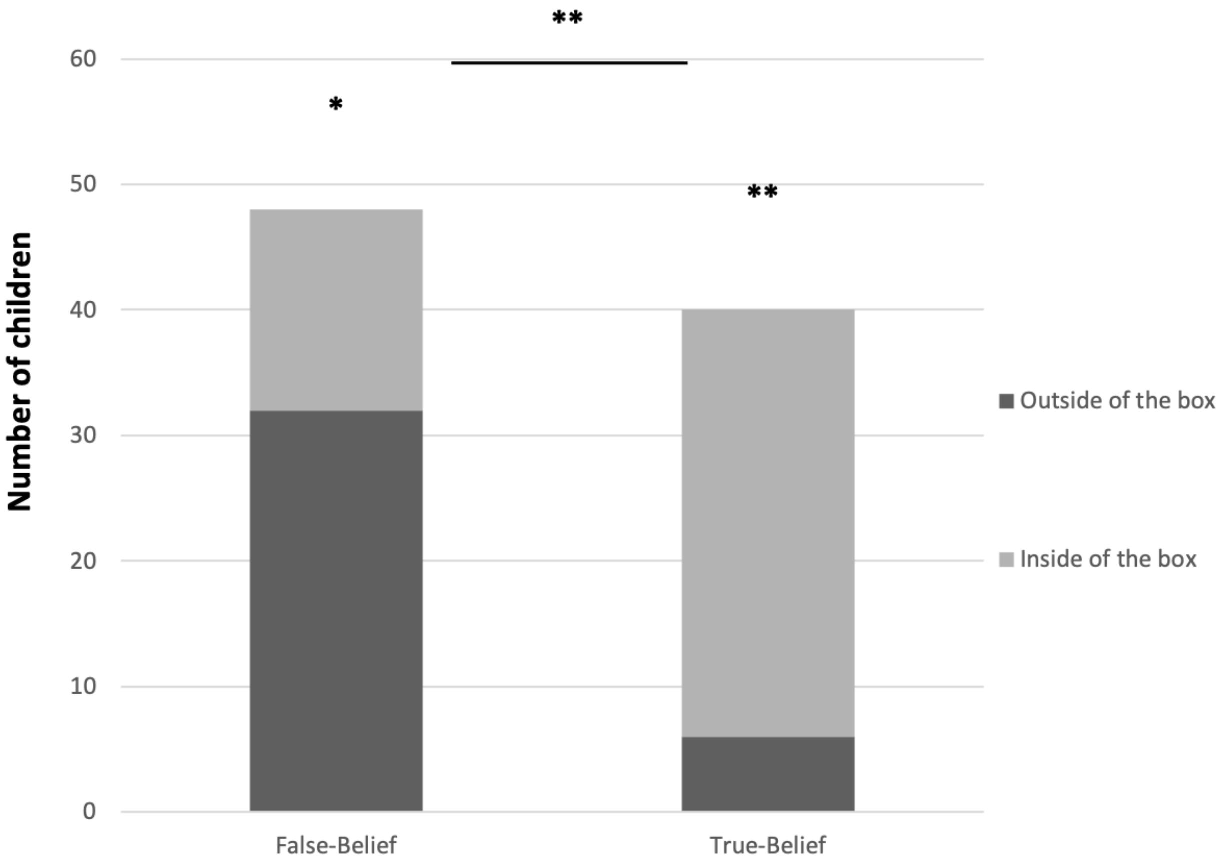


Figure 5. Number of children pointing to the dog outside of the box (correct in False-Belief condition) and to the dog inside of the box (correct in True-Belief condition), including those who failed the "See" question. ** = p <.01, two-tailed.

Overall, the results are in line with those from Experiment I (see Figure 6). Importantly, in the case of Experiment II, there was no description accompanied with the naming act, hence children had no other potential cues to Fido's identity. Rather, to discern who Fido is in the test scene, they had to rely solely on what the speaker had in mind (i.e. her belief about the identity of the dog in the box) when she pointed to the box and uttered the proper name. We found that, indeed, majority of three-year-olds in

Experiment II pointed to the dog outside of the box in the test scene, indicating that they took into account what Jane had in mind when she (mistakenly) named the one in the box "Fido". Bayesian analyses supported this conclusion, revealing strong evidence for H1, both when those who failed the "See" question were included and when they were excluded from the sample.

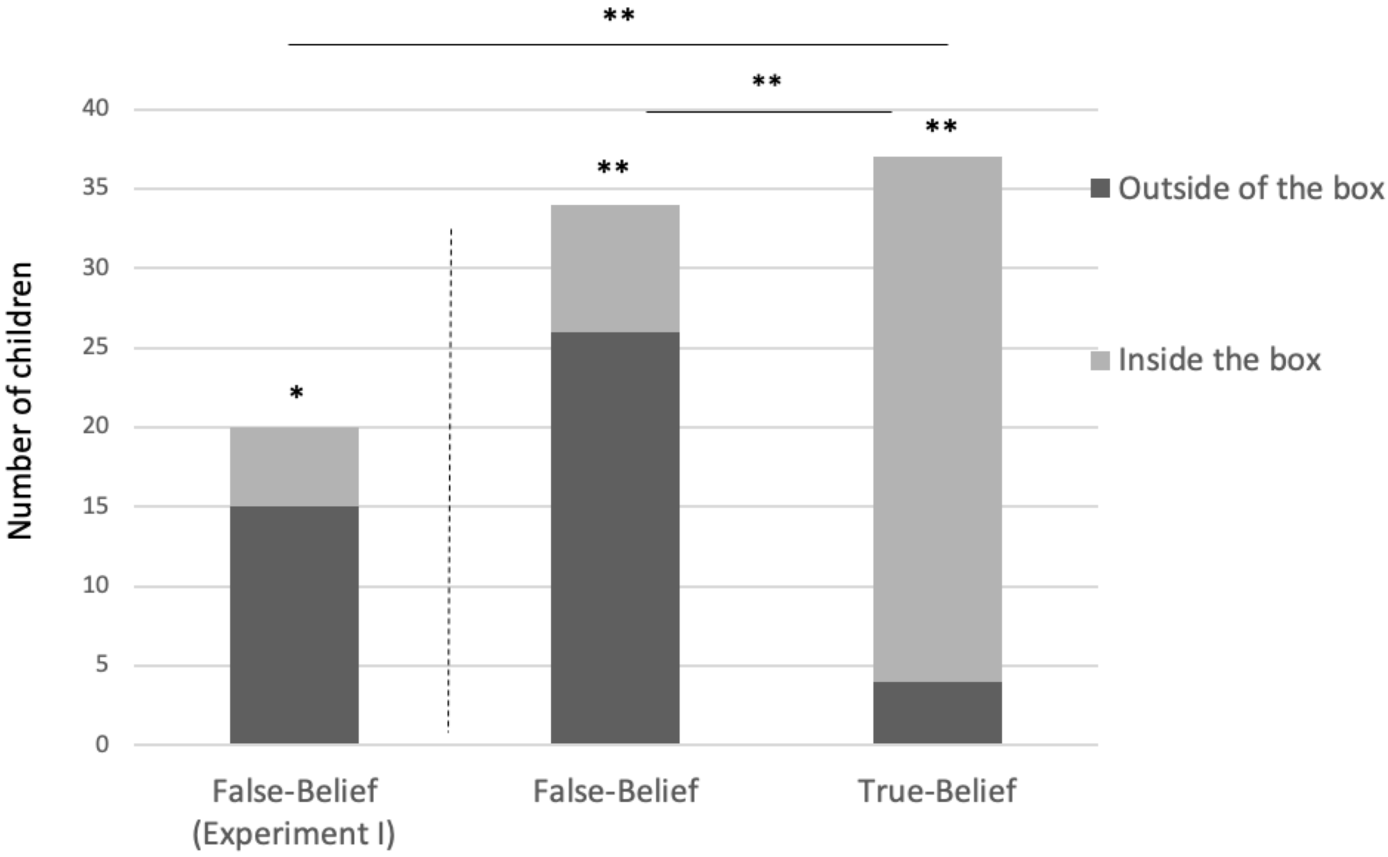


Figure 6. Patterns of children's pointing responses across experiments I and II. Majority of three-year-olds pointed to the dog outside of the box in Experiment I and FB condition of Experiment II; by contrast, majority of children in the TB condition (Experiment II) pointed to the dog inside of the box. * = p < .05, two-tailed; ** = p <.01, two-tailed.

Taken together, the results from Experiment I and Experiment II provide a strong evidence for the role of ToM reasoning in the process of mapping a proper name and its referent; overall 41 out of 54 (75.9 %) children[18] correctly inferred that the referent of the name "Fido" would be the dog that Jane *intended to* refer to, but that she *did not actually* end up referring to (cumulative Binomial $p < .001$, two-tailed, BF = 296 in favor of H1, which is considered extreme evidence, e.g. Lee & Wagenmakers, 2013).

In addition, the opposite pattern of pointing responses across FB (both experiment I and II) and TB conditions, rules out low-level, non-mentalistic explanations (such as side-bias) of children's successes in discerning the referent of the novel name.

Experiments I and II suggest that young children can engage in false-belief reasoning not only for the purposes of prediction and explanation of other agents' actions, but for acquiring new information from them—in this case, for discerning the link between a proper name and its referent. These experiments, thus, provide the first evidence for the role of genuine ToM in word acquisition, specifically, for the role of false-belief reasoning in mapping a label (i.e. a proper name) and its referent.

As we discussed earlier, the fact that three-year-olds were successful in attributing an identity false-belief in a verbal task suggests developmental continuity of ToM competence. Indeed, as elaborated in the discussion of Experiment I, in order to pass our task, children had to understand that the agent was mistaken about the identity of the dog that she was naming. Recall that some researchers suggested that, although children can still perform well on FBTs that involve location change of the target object, this need not

[18] Excluding children who failed the "See" question.

reflect genuine belief-reasoning (Butterfill & Apperly, 2013; Low & Watts, 2013). The test case, according to these researchers, would be an FBT that requires reasoning about others' beliefs about identities; they considered this capacity unavailable prior to the age of four, when the minimal ToM (but not the "full blown" one) is claimed to be the only system available to a young child. Our results challenge this approach, showing children's successes in identity false-belief reasoning before the age of four, strengthening the assumption that a single ToM accounts for both early and late successes on FBTs.

However, how young is young *enough* to grant developmental continuity of ToM competence? Indeed, to make a stronger case for developmental continuity of genuine ToM capacity (and, moreover, for its crucial role in word learning) we would have to demonstrate successes on a similar task in even younger children. This was the main goal of Experiment III, where we tested 2.5-year olds on a similar version of the naming FBT that we developed for Experiments I and II. In addition, we wanted to provide a replication of the study by Setoh et al. (2016; see also Grosso et al, 2019 for a recent replication), who demonstrated that when processing demands are sufficiently reduced, even young 33-month-olds can pass the traditional, verbal version of an FBT. These results provided a significant support for the hypothesis that the actual reason behind young children's struggles with traditional FBTs does not lay in their limited ToM capacity, but in processing demands associated with these tasks.

Given the significance of these findings and given that we were testing 2.5-year olds on a novel, naming FBT, we wanted to reassure that they were able to pass the traditional FBT (following Setoh et al., *ibid*), in order to make sense on their performance on the naming FBT.

2.3. Experiment III

Exploring 2.5-year-olds' performance on both the proper names learning FBT and on a traditional FBT with reduced demands

The aim of the third experiment was two-fold. First, we wanted to test if 2.5-years old children would also be successful on the naming FBT used in the first two experiments, i.e. whether they could map a proper name and its referent through the speaker's identity false-belief. If they, indeed, performed successfully, this would provide an even stronger support for developmental continuity of meta-representational ToM and its early capacity to represent identity false-beliefs.

The second aim was to replicate Setoh, Scott and Baillargeon (2016) who found that even 2.5-year olds could succeed on a traditional false-belief task if its demands are *sufficiently* reduced. According to Setoh et al., traditional FBTs are taxing for younger children not only because they pose demands on inhibitory control, but also because of their significant demands on response generation: young children have to correctly interpret the test question and generate an adequate response, given a representation of the agent's mental states. When only inhibitory, but not response generation demands are reduced, young children's performance does not improve above chance (e.g. Devine & Hughes, 2014). To reduce demands of response generation, Setoh et al. embedded specific practice trials in their FBT. They presented young subjects with a narrated picture-book story, where toddlers were prompted to point to where an agent will look for her apple ("*Where will Emma look for her apple?*"). Importantly, when two practice trials were embedded within the story, which allowed children to practice response generation (e.g. "*Where's Emma's apple?*", "*Where's Emma's ball?*"), even 33-month-olds were successful in correctly predicting where Emma will look for her ball. Setoh et al. reasoned that this manipulation sufficiently reduced younger children's processing demands, by allowing them to practice generating pointing responses.

Following Setoh et al., we wanted to maximize young subjects' successes on our naming FBT, by reducing potential response-generation demands. To do so, and to provide additional evidence in favor of 2.5-year-olds' capacity of passing a traditional FBT with reduced demands, we included the adaptation of Setoh et al. task in our experiment, so that all subjects were first presented with this task ("Emma task") followed by our naming FBT ("Fido task"). The reasoning was that the "Emma task" could serve as a warm-up trial where 2.5-year-olds could practice generating pointing responses. This would potentially reduce performance demands and would make the "Fido task" simpler for children as young as 2.5-years old, allowing them to manifest their actual capacity.

2.3.1. Method

Participants and Procedure.

Twenty-four 2.5-year-olds were tested individually, in a quite environment, either in the lab or at their preschools. Four additional children were tested, but excluded from the final sample due to failure to cooperate (N = 3) and due to experimenter error (N = 1). Participants were first presented with a traditional, FBT—"Emma task" (N = 24 children, 12 female, M = 32.6 months, SD = 2.95 months, age range = 25- to 36-months), adapted according Setoh et al. (*Ibid*). After this task, subjects were presented with the FB condition of the "Fido task" (N = 19, 10 females, M = 32.4 months, SD = 2.98 months, age range = 25- to 36-months)[19]. Thus, each child received two stories in total.

Traditional FBT ("Emma task").

The story was presented in form on a set of illustrated slides, on a laptop computer, using Microsoft PowerPoint. The child was seated in a small chair, centered in front of the screen. The experimenter was seated on the floor, on the right side of the child's chair and was advancing the slides manually, by pressing the right arrow on the laptop's keyboard. The experimenter narrated the story about a girl, Emma, who finds an apple in a bowl and puts it in a box for later. The child, then, saw a slide presenting two pictures side by side (i.e. the picture of Emma's apple and the picture of a banana) and will be asked to point to Emma's apple (*"Where is Emma's apple?"*). This was the first practice trial, in which children practice generating the pointing response. After this, children saw Emma, putting her apple in a box. Then, she was presented outside of her room, playing with a ball. The next slide was the second practice trial in which children saw two pictures side by side (Emma's ball and a freebee) and were asked to point to Emma's ball (*"Where is Emma's ball?"*). After this slide, children saw Emma's brother,

[19] Five additional children were tested but excluded from the "Fido task" analyses (2 female, M = 33.06, SD = 3.06, range = 28- to 35-months), since they received a slightly different version of the original "Fido task". In particular, they had embedded pointing trials in the "Fido task" as well, where they were asked to point to the pictures presented on the wall of the scene, prior to the test question, i.e. "Can you point to Fido?" Since it appeared that these subjects were getting distracted from the practice question embedded prior to the test question, we excluded them from the main analyses of the second task (i.e. the "Fido task"). The rest of the children received the original version of the "Fido task", with no practice trials embedded within it.

Ethan, who found the apple and took it away.[20] Then, hungry Emma came back to look for her apple. The test trial showed two pictures, side-by-side, one presenting the box and the other presenting the bowl, and children were asked, *"Where will Emma look for the apple?"* (For schematics of the task, see Figure 7)

Naming FBT "Fido task".

Following "Emma task", subjects were immediately presented with the second, "Fido task" (FB condition), which followed exactly the same form as the one in Experiment II. As in "Emma task", the child was seated on the small chair, facing the screen of the laptop, while the experimenter, seated on the floor next to the child, was advancing the slides manually, using the laptop's keyboard.

[20] Note that this makes the task a low-demand FBT, since the target-object is removed from the scene when the actor comes back to search for it. Prior research as discussed, showed that this modification improves 3-year-olds' performance on verbal FBTs (Bartsch, 1996). However, Setoh et al.'s reasoning was that this was not sufficient to allow for *younger* children's successes; hence they added the practice trials to reduce demands coming from response generation.

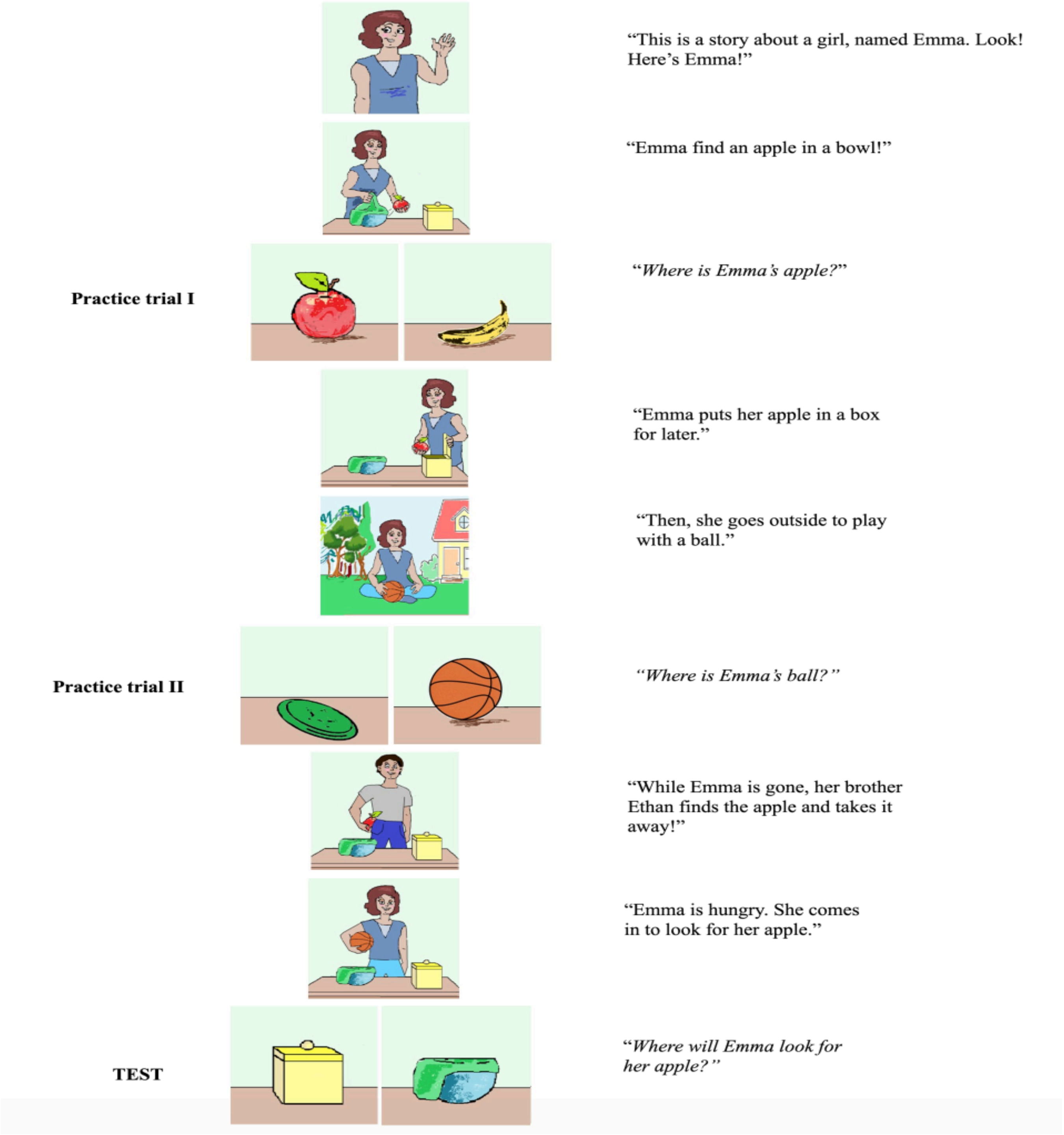

"This is a story about a girl, named Emma. Look! Here's Emma!"
"Emma find an apple in a bowl!"
Practice trial I
"Where is Emma's apple?"
"Emma puts her apple in a box for later."
"Then, she goes outside to play with a ball."
Practice trial II
"Where is Emma's ball?"
"While Emma is gone, her brother Ethan finds the apple and takes it away!"
"Emma is hungry. She comes in to look for her apple."
TEST
"Where will Emma look for her apple?"

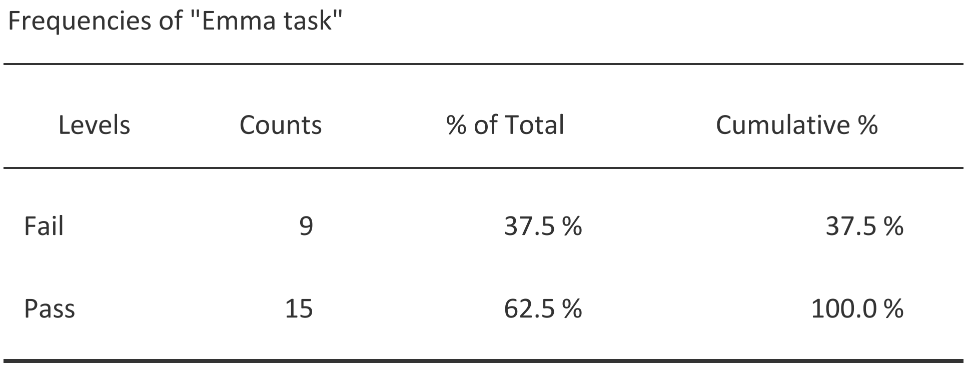
Figure 7. Schematics of the stimuli used in Experiment III, "Emma task". The task was adapted following Setoh et al. (2016). The sequence of the events and the script were exactly the same as in the original study; drawings were made to match the photographs of the original study.

2.3.2. Results and Discussion

Preliminary analyses revealed no effect of children's gender on their performance on neither "Emma task" nor "Fido task", so we collapsed the data across this factor. We first focus on children's performance on "Emma task". Children performed reliably above chance on both the first and the second practice trials (20/24 children pointed to the picture of the apple in the first practice trial, 20/24 children pointed to the picture of Emma's ball on the second practice trial, p <.001, two-tailed). However, children were not significantly better than chance when it comes to predicting where Emma will search for her apple: 15 out of 24 children (62.5 %) pointed to the container that was congruent with Emma's belief in response to the test question (Binomial p = .15, two-tailed, BF = 1.95 in favor of H0).

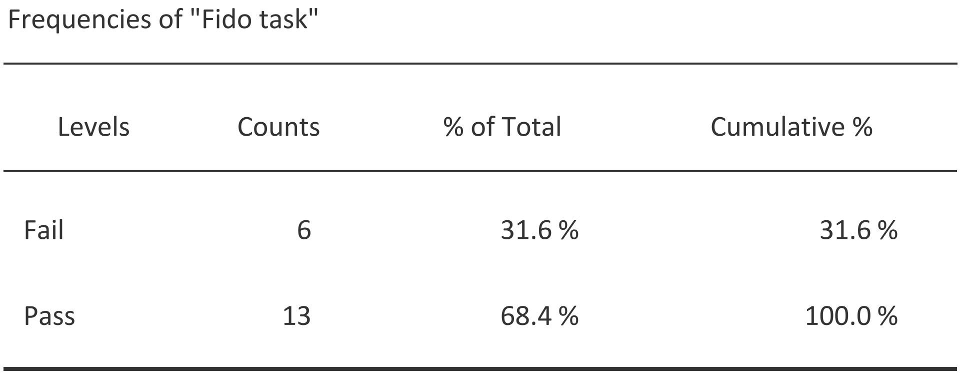
We, then, focus on performance on those children who completed "Fido task", following "Emma task". We found a trend toward choosing the intended (correct) referent of the proper name, but the results failed to reach significance: out of 19 children, 13 (68.4 %) pointed to the dog that was outside of the box at the time of naming (Binomial p =.08, one-tailed, BF = 0.966 in favor of H1) (See Figure 8). Children's responses on the two tasks are summarized in Table 4.

Since 19 subjects were tested on both "Emma task" and "Fido task" in a repeated measures design, we also conducted McNemar's Test of Changes to see if there was a consistent pattern of succeeding/failing between the two tasks in individual subjects. This analysis revealed no significant consistency in subjects' responses to the test question between the two tasks (McNemar Binomial, p = .1, two-tailed, n.s.).



Frequencies of "Emma task"

Levels	Counts	% of Total	Cumulative %
Fail	9	37.5 %	37.5 %
Pass	15	62.5 %	100.0 %



Frequencies of "Fido task"

Levels	Counts	% of Total	Cumulative %
Fail	6	31.6 %	31.6 %
Pass	13	68.4 %	100.0 %

Table 4. Number of children who passed/failed the test question on “Emma task” and “Fido task” in Experiment III.

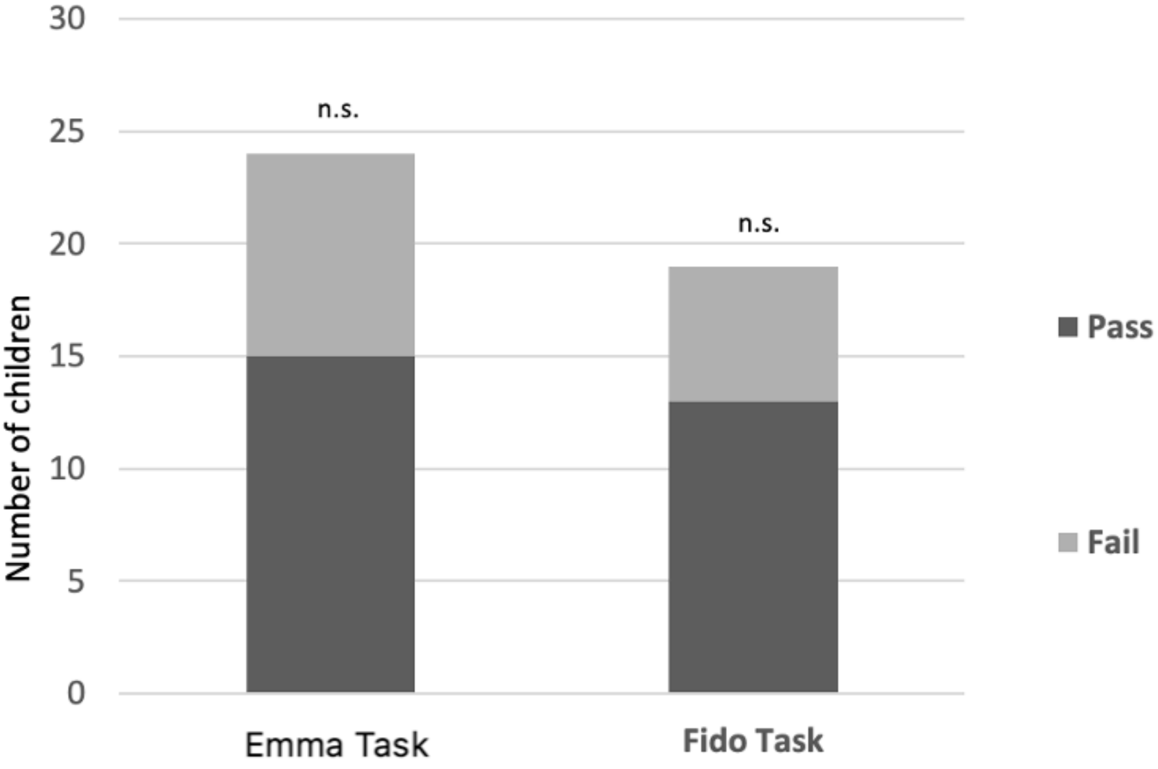


Figure 8. Number of 2.5-year-olds passing the "Emma task" and number of children passing the "Fido task", in Experiment III.

Overall, we found that 33-month olds performed at chance on both "Setoh task" and "Fido task" (although there was a marginal trend towards pointing to the correct referent on the later). Since we did not reproduce the original study by Setoh et al. (2016), which has received a number of replications from other researchers (e.g. Grosso et al., 2019; Scott, Roby, & Setoh, 2020) we now turn onto a discussion of the potential reasons that could have contributed to the negative outcome in our experiment.[21]

[21] We thank Renée Baillargeon and Rose Scott for valuable comments and suggestions regarding this matter, which were communicated in an email interchange. They pointed out to us that the original task

Story-telling platform (medium). One of the potentially important differences between the original task, used by Setoh et al. and the task we used in Experiment III was the medium through which the story was presented. Setoh et al. relied on a fully live environment to deliver the story, with and actual picture book and a live narrator (experimenter). This medium could have created a more natural story-telling environment and could have eased young 2.5-year-olds' attention to the storyline. By contrast, we presented the visual aspect of the story (i.e. illustrations) via an electronic medium, while the narration was done live by an experimenter. It is possible that this mixed media in delivery of the story made it more difficult for young children to properly attend to the storyline. There is, indeed, some evidence suggesting that electronically delivered stories might be cognitively taxing for young children (e.g. Krcmar & Cingel, 2014), so relying on this medium could have canceled out beneficial effects of the two practice trials, which Setoh et al. included to reduce the processing demands.

Size of the stimuli. The images that we presented to children, as well as the space between the displayed images (e.g. in practice and test trials) were smaller compared to the images (and the overall display) used in Setoh et al., which could have made our task more difficult for young children. (Specifically, Setoh et al. used 20 cm X 25 cm colored photographs, whereas we used 13 cm X 14 cm drawings).

Delivery. The fact that we relied on PowerPoint to present the story was associated with another potentially relevant factor that could have increased the overall processing demands of the task. Namely, the story in our setup was advanced by changing the slides, so that a new slide would instantly appear to replace the previous

was, indeed, sensitive to details, so that any changes that might appear insignificant could, as a matter of fact, have detrimental effects (Baillargeon & Scott, p.c.).

one. By contrast, in Setoh et al., the story was advanced by flipping the pages that were bound on the top of the story book, which could have allowed for more processing time and, consequently, for children's better integration of the story sequences into a uniform narrative. Given that this was a verbal task presented to 2.5-year-olds who lacked the sufficient mastery of language, slower progression of the story might be critical to allow for their sufficient comprehension.

The way in which the experimenter advanced the story in our experiment—i.e. by pressing a button on the laptop's keyboard—was also associated with her being positioned next to the child and facing the screen. By contrast, in Setoh et al., the experimenter was centered behind the story book, facing the child and flipping the pages that were bound on its top. This difference, again, could have resulted in our task being cognitively more taxing than Setoh et al., as the experimenter's position could have created a side bias, which younger children struggled to overcome.

Language variables. In the original study by Setoh et al., the narration was always done by a native speaker of English. In our experiment the narrator was fluent in English, but was not a native speaker, and had an accent that could have been unfamiliar to younger children, which could have made it more difficult for them to process the story. It should be noticed that the same narrator was telling the story in Experiment I and II, where we found the effects; however, the subjects in these experiments were older and had a better mastery of language compared to 2.5-year-old subjects who participated in Experiment II. Younger children are more sensitive to accents, and this could have impeded their understanding of the story in our experiment, especially if they were not familiar with the narrator's accent.

It is possible that some (or all) of these factors had contributed to the fact that we did not find the effects with 2.5-year-olds. Testing this assumption would require a follow-up study that would systematically manipulate these factors to see how (and if) they affect children's performance on the type of FBT originally designed by Setoh et al. In addition, it would be beneficial to include a TB condition in the "Fido task" with 2.5-year-olds; comparison between their performance on FB and TB could reveal a significant difference in their pattern of responses, which would provide insights into the underlying competence, even without single conditions reaching significance.[22]

2.4. General Discussion (Experiments I-III)

Together, the results from Experiment I, Experiment II and Experiment III provide a solid evidence that by at least the age of three, children can map a proper name and its referent through the speaker's false-belief about the referent's identity. This has two important implications: 1) children appear capable of representing identity false-beliefs prior to the age of four, which supports developmental continuity of ToM and 2) ToM—specifically false-belief reasoning— might play an important role in word learning, which suggests strong mentalism of this process. Bellow, I discuss both of these implications.

Representing identity false-belief prior to the age of four. The question of whether identity false-beliefs are within the scope of a young child's ToM is particularly important for understanding the conflicting findings coming from the FBT paradigm. As

[22] A follow up study that would follow these directions was in preparation as a part of this project, but the data-collection phase was prevented by COVID-19.

discussed before, there are disputes among researchers regarding whether children's successes on implicit FBTs reflect the same, unique ToM competence that allows for later successes on explicit (traditional) FBTs. The dual-systems approaches assume that two fundamentally distinct systems are reflected by implicit and explicit FBTs, where only the later ones could be taken as reflecting a "full blow", meta-representational ToM (e.g. Butterfill & Apperly, 2013). According to this view, early ToM lacks the capacity to flexibly represent *how* someone represents an entity (e.g. the registered star *as* "Hesperus", but not as "Phosphorus"), which is a necessary component in understanding how others represent individual identities. Hence, demonstrating early understanding of identity false-beliefs could be taken as a test case for developmental continuity of ToM competence.

In this respect, our findings provide a challenge for the dual-systems approaches and support developmental continuity of ToM competence. Young 3-year-olds in our experiments demonstrated understanding of a speaker's false-belief about the identity of a pet that she intended to refer to. Moreover, they could use this understanding to solve the problem of referential ambiguity, as they correctly inferred who the referent of a proper name was in our naming FBT. That their pointing responses were driven by identity false-belief understanding was supported by the findings from Experiment II, where no description of the referent was provided and where we included a TB condition as well (which, as we saw, yielded the opposite pattern of children's pointing responses). However, although these findings do provide a support for developmental continuity of ToM capacity, we still need to be cautious in respect to embracing this interpretation as we did not find the evidence of identity false-belief attribution in younger children, i.e.

2.5-year-olds. As we saw, Experiment III revealed a trend towards pointing to the correct (i.e. intended) referent of the proper name, but the results failed to reach significance.

Could one argue that the negative results of Experiment III were due to younger children's deficiency in ToM competence? Although this is a possible hypothesis, our results do not provide a support for it; specifically, Bayesian analyses that we conducted revealed no support for H0 over H1 (BF = 0.966 in favor of H1). Hence, we are inclined to interpret the negative results as a result of the task related noisiness, rather than as a result of younger children's lack of genuine false-belief understanding. In addition, our subjects were at chance on the traditional FBT as well, which was previously shown to demonstrate successes in 2.5-year-olds' false-belief reasoning (Setoh et al, 2016; Grosso et al., 2019; Scott et al. 2020). Given that other studies demonstrated FB reasoning in 2.5-year-olds using a similar method, and given that we did not find a strong evidence for the *actual* lack of an effect (BF = 1.95 in favor of H0), we conclude that Experiment III wasn't powerful and/or precise enough to properly capture 2.5-year-olds competence. We discussed some of the potential factors that could have contributed to this outcome in the previous section; as noted, future research is required to systematically test these factors.

As pointed out, "Emma task" is sensitive to details, hence seemingly trivial modifications of the original task (e.g. presenting the story on a screen, instead of using the actual picture book) might have resulted in impeding young children's performance. Similarly, 2.5-year-olds' poorer performance on "Fido task" compared to three-year-olds' successes (Experiments I and II), could be a result of the fact that younger children are more prone to be influenced by performance factors, extraneous to ToM reasoning (e.g. language comprehension, as discussed above). Also, unlike children who participated in

Experiments I and II, children in Experiment III were always tested on “Fido task” only after previously completing “Emma task”. Although including “Emma task” prior to “Fido task” was intended to reduce potential processing demands on the later task, we could have, in fact, created the opposite effect, resulting in younger children’s cognitive fatigue. To test this possibility, a future study is required where younger children will be tested on “Fido task”, without previously participating in another task. Finally, since younger children’s performance tends to be more fragile in face of extraneous and/or task-related (but not ToM competence related) factors compared to older children, a larger sample size might be necessary to capture the same effect.

Learning a new word via FB reasoning. Experiments I and II together provide a strong evidence that, by at least the age of 3.5 children can take into account a speaker’s false-belief to map a label and its referent. This bears significant implications for the current theories of word acquisition, suggesting an important role of ToM reasoning in solving the problem of referential ambiguity. As we already discussed, previous research indicated that young word learners attend to various social-attentive cues to select a likely candidate of an unfamiliar label in a noisy learning environment (e.g. Baldwin, 1993; Baldwin et al., 1996; Baldwin & Tomasello, 1998; Bloom, 2000; Cartmill et al., 2013; Tomasello & Kruger, 1992; Koening & Echols, 2003; see also Papafragou, Friedberg, & Cohen, 2017). Our findings extend the social learning accounts, suggesting that children go even deeper in attending to agents in the process of referent selection—they attend to the agents’ *epistemic states*, such as (false-)beliefs, indicating a profoundly social nature of word acquisition.

However, notice that so far, we only focused on how ToM reasoning affects children's capacity to map a proper name and its referent. In order to make a stronger claim for the central role of ToM reasoning in word acquisition, we have to go beyond the case of proper names and demonstrate that children can as well acquire other kinds of words by correcting for the speaker's false-belief. We aim to make a further step in this direction with the next experiment (Experiment IV), where we test whether children younger than four can also learn a new *common noun* through the speaker's false-belief about its referent. Showing that this was the case would, hence, strengthen the assumption that ToM might be a key vehicle for mapping new labels onto their referents; in addition, it would add to the growing body of empirical evidence revealing young children's successes in FB reasoning prior to the age of four, when tested on a verbal version of a FBT.

2.5. Experiment IV

Three-year-olds' performance on a common nouns learning FBT

The goal of Experiment IV was to extend the findings from the first two experiments and to provide a stronger support for the claim that ToM plays an important role in a young child's language development, specifically in her acquisition of a novel label.

As we previously discussed, a number of researchers have explored the relation between early agency reasoning and development of certain aspects of language, such as vocabulary growth. These researchers demonstrated that young children attend to various social cues to select the most likely referent of a new label (e.g. Baldwin, 1991; Baldwin et al., 1996; Cartmill et al., 2013; Gleitman & Trueswell, 2018; Trueswell et al., 2016). However, they still left the question of whether young children also attend to the speaker's epistemic states—such as beliefs—to narrow down the space of potential referents of a new label.

Several studies explored more closely the link between attending to a speaker's mental states (such as intentions and beliefs) and labeling events (e.g. Carpenter et al., 2002; Jin & Song, 2014; Koening & Echols, 2003; Papafragou, Fairchild, Cohen, & Friedberg, 2016; Southgate et al., 2010). For instance, Southgate et al. (2010) presented 17-month-olds with a scenario where an agent places two novel objects in two separate containers and leaves, after which the objects switch locations. When the agent returns, she points to one of the boxes, says that a "sefo" is in there, leans the boxes towards the child and ask her/him to help her get the *sefo*. Children tended to reach for the box that the agent did not point to, when she provided the label, suggesting that they took into account the agent's false-belief to assign reference in a communicative setting. Similar findings were obtained with older children, i.e. 3-to 5-year-olds (Carpenter et al. 2002; Happé & Loth, 2002; Houston-Price, Goddard, Séclier, Grant, Reid, Boyden, & Williams, 2011).

It should be noted that the focus of these studies was on the question of whether communicative setting—such in which the agent interacts with a child and uses referential intention to convey relevant information to her—facilitates ToM (specifically, false-belief) reasoning, rather than whether ToM facilitates word acquisition. As a matter of fact, both Southgate et al. (2010) and Carpenter et al. (2002) included a "novel word" condition (where the actor referred to the desired toy with a novel word, e.g. “A sefo is in here, shall we play with the sefo?”) and a “generic expression” condition (where the actor referred to the desired toy in a generic way, e.g. “Do you know what’s in here? Shall we play with *it?”*). Importantly, they did not find a difference between the two conditions, which leaves the question of whether ToM facilitates word learning (or vice versa) opened.

By contrast, we focus on the aspect of word acquisition. Precisely, we test whether young children can assign reference through the speaker’s false-belief, when the speaker (unlike in the cited studies) does not interact with the child at all (hence, the role of the communicative context is rendered less salient). In addition, in this experiment the test question is not a ToM question, but a pure word learning question: the child is simply asked to point to the referent of a new label, in a complete absence of any of the characters from the story. Note that this was not the case with the studies cited above, as they asked children to do something in respect to a collaborative activity with the agent (e.g. helping her get the toy she desired). Hence, our experiments aim to extract the role of ToM reasoning in assigning the reference, when other factors (such as engaging the child in the communicative context) are controlled for.

Here, we tested whether 3-year-olds can learn a new common noun through the speaker’s false-belief, using the same method we developed for Experiments I, II, and III. We presented our subject with a narrated, illustrated story (displayed through PowerPoint slides) in which a character places one of the two unfamiliar creatures in a box, and then she leaves. Then, the creature that was in the box flies out of the window, and the other one replaces it in the box. The character returns with her friend, points to the box and announces that “a daxy is in the box!”. In the test scene, two creatures are presented side-by-side and children are asked to point to the daxy. In addition, we included the True-Belief (TB) condition, in which the only difference was that the character returned to the room before the creatures switched their places. We predicted that if children spontaneously compute false-beliefs as they listen to the story, and if they can use this computation to infer the referent of a new label, they should point to the creature outside of the box in the FB condition and inside of the box in the TB condition.

2.5.1. Method

Participants

Participants were 38 English-speaking three-year-old children (M = 40 months, SD = 3.2 months, 20 female), ranging from 34- to 46-months. Out of 38 children, 20 were tested on the False-Belief (FB) condition (M = 39 months, SD = 3.2 months, 11 female) and 18 were tested on the True-Belief (TB) condition (M = 41.7 months, SD = 2.6 months, 9 female).

Stimuli and Procedure

The procedure was similar to the one used in the Experiment I – III. Participants were tested individually, either in a quite environment in their preschools or in the lab. Children were presented with an illustrated story that was displayed through PowerPoint slides, on a laptop computer. They were seated in front of the laptop, facing the screen (centered in respect to it). The story was narrated by the same experimenter for all participants, who was seated next to the child and advancing the story by pressing a key on the laptop’s keyboard. (For schematics of the task, see Figure 9.)

False-Belief (FB) Condition. The children listened to a story about a girl, Sarah, who has two strange creatures (pseudo-animals that belong to two different, unfamiliar kinds). Sarah puts one of the creatures in the box, closes the lid and leaves the scene. Once Sarah is gone, the creature that was placed in the box pops out of it and flies out of the room, through a window. Then, the other creature jumps into the box and the lid closes afterwards. Sarah, then, comes back with her friend, Billy, who points to the box and asks, "What is in this box?". Sarah points to the box and says, "A *daxy*! A *daxy* is in the box.", providing a new label ("*daxy*") for the first time in the story. The friends, then, leave the scene. Once they are gone, the first creature (Sarah's intended referent of the label "daxy") flies back to the room, right through the window, and the other creature pops out of the box. Children are, then, asked, "Can you point to the daxy?".

True-Belief (TB) Condition. This condition was the same as the FB condition, but for one difference: Sarah returns to the room before the creatures switch their places, hence she ends up with the true-belief about what creature is in the box. In this case, Sarah's intended referent is the same as the actual referent (i.e. the referent she ostensibly labels during the naming scene), that is the creature inside of the box.

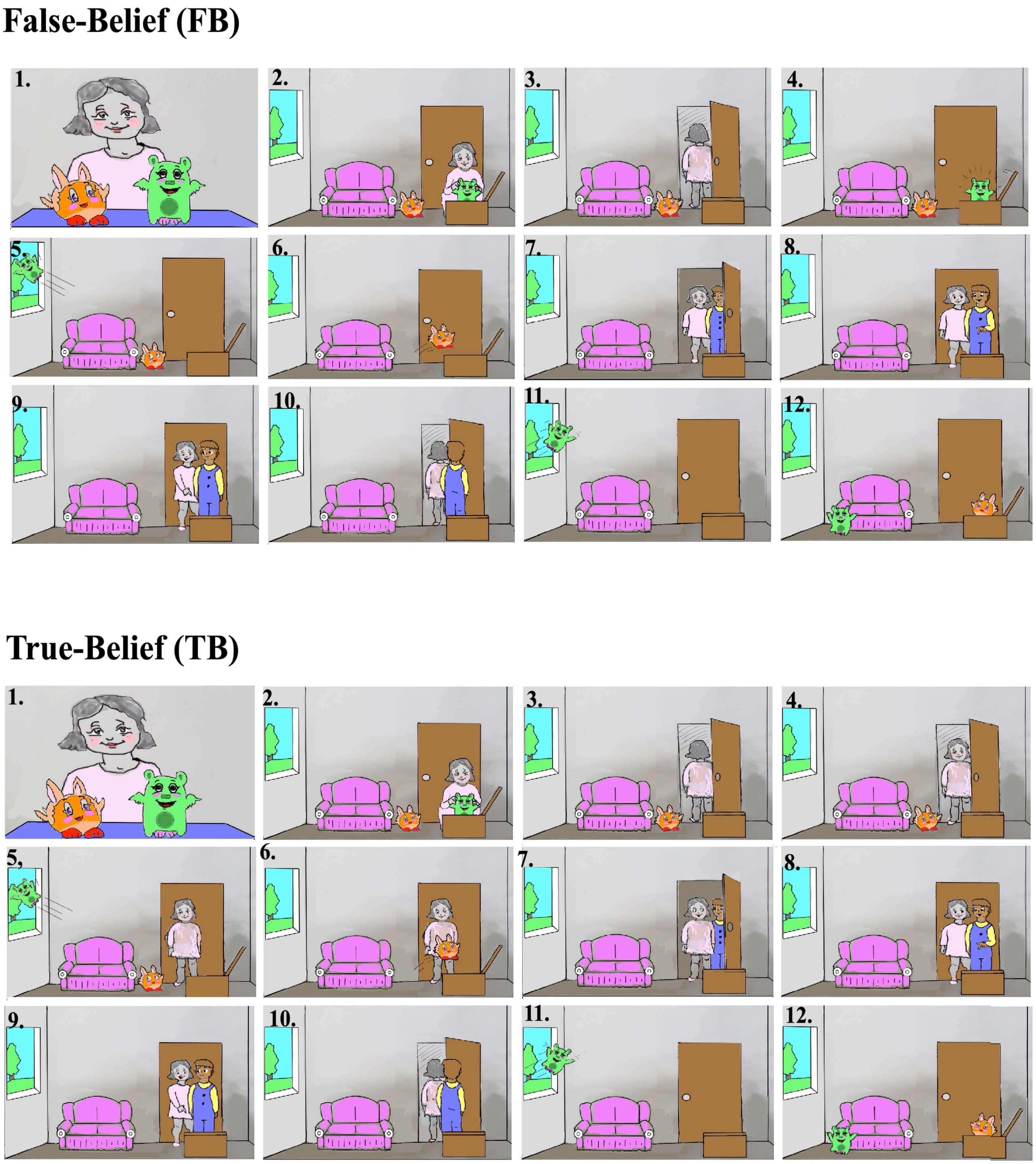
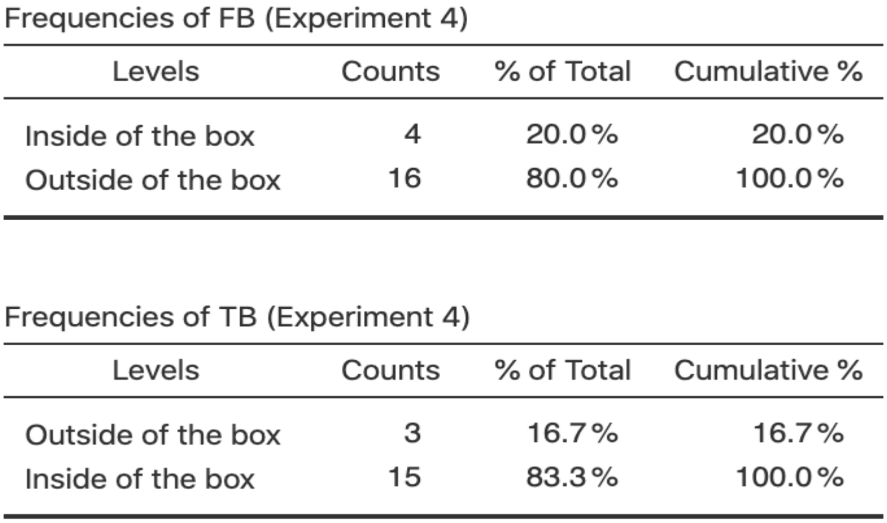


Figure 9. Schematics of the stimuli used in Experiment IV. Figure includes the key events. The bottom panel is the True-Belief (TB) condition, where Sarah comes back before the creatures switch their places.

2.5.2. Results and Discussion

Preliminary analyses of children's performance on FB and TB conditions revealed no significant effect of gender, so we collapsed the data across this factor. Children's responses are summarized in Table 5. The main analyses reveled that out of 20 children in the FB condition, 16 (80%) pointed to the creature that was outside of the box during the scene of labeling (Binomial *p* < .01, two-tailed, BF = 10.306 in favor of H1), suggesting that they successfully mapped the category-label with the referent through the speaker's false-belief. By contrast, 15 out of 18 children (83%) in the TB condition pointed to the creature that was in the box at the time of naming (Binomial *p* = .008, two-tailed, BF = 16.908 in favor of H1). The two conditions differed significantly (Fisher's Exact test, *p* <.001, two-tailed), which rules out potential low-level explanations, such as preference and side-bias, of children's pointing responses in the FB condition (see Figure 10).



Frequencies of FB (Experiment 4)

Levels	Counts	% of Total	Cumulative %
Inside of the box	4	20.0 %	20.0 %
Outside of the box	16	80.0 %	100.0 %

Frequencies of TB (Experiment 4)

Levels	Counts	% of Total	Cumulative %
Outside of the box	3	16.7 %	16.7 %
Inside of the box	15	83.3 %	100.0 %

Table 5. Number of children who pointed to the creature outside vs. the one inside of the box in the False-Belief (FB) and True-Belief (TB) conditions, Experiment IV.

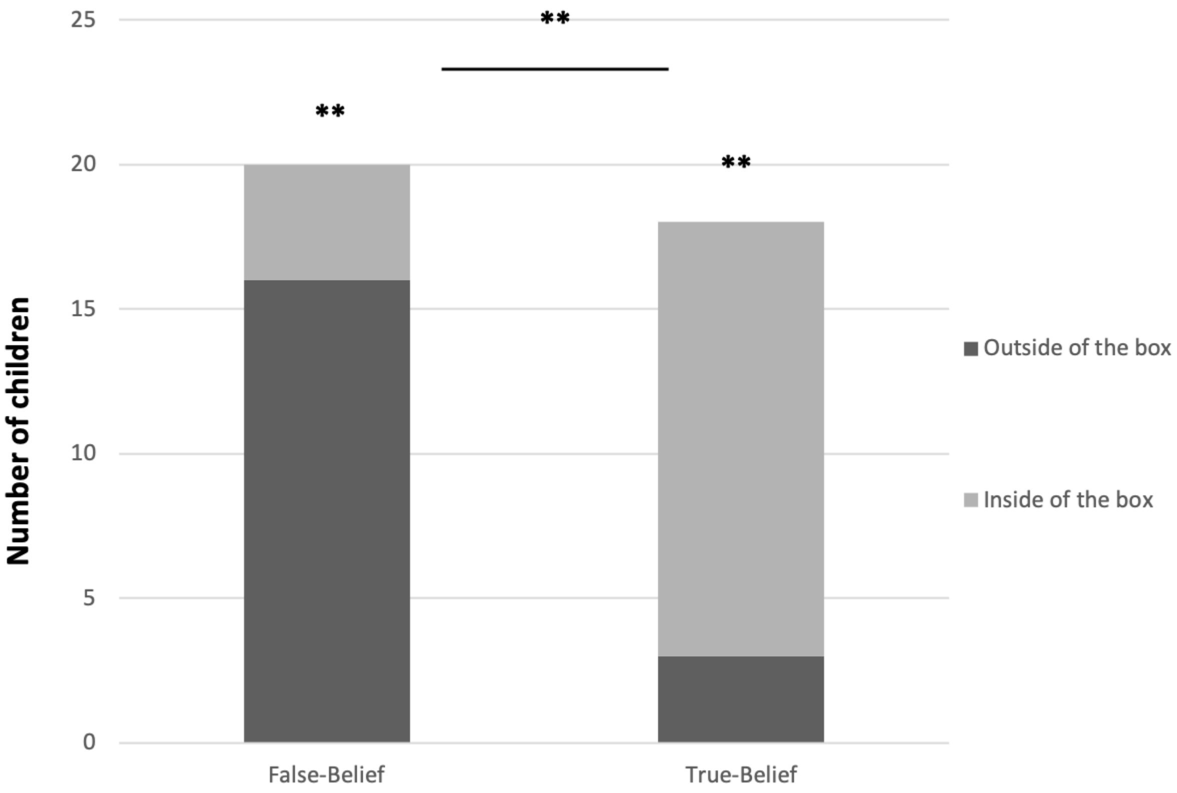


Figure 10. Number of children pointing to the creature outside of the box (correct in False-Belief condition) and to the creature inside of the box (correct in True-Belief condition) in the Experiment III. ** = p =<.01, two-tailed.

These results demonstrate that by the age of 3.5 children can map a new common noun and its referent by correcting for the speaker's false-belief. This is in line with the findings from Experiments I and II, in which 3-year-olds children discerned the referent of a proper name, by attending to the speaker's beliefs in the naming scene. Hence, Experiment IV provides a further support for the assumption that ToM plays an important role in referential disambiguation, and that its facilitating effects apply to different kinds of words, such as proper names (as demonstrated by the Experiments I and II) and common nouns (as demonstrated by Experiment IV). However, to strengthen our conclusion, further studies should attempt to replicate our findings from Experiment IV with a larger sample size.[23]

It should be noted that the way in which we differentiated between the two kinds of words (i.e. proper names and common nouns) across the set of experiments (I-IV) involved a minimal modification: adding an article (“*a”*/ “*the”*) before the label in Experiment IV, to indicate its grammatical category (i.e. that it is a common noun). Previous research suggest that this is, indeed, an effective manipulation to induce a common noun/proper name interpretation of a novel label in young children (e.g. Bélanger & Hall, 2006). For instance, in a preferential looking study, Bélanger and Hall (2006) demonstrated that 16- to 20-month old children expected the same label (e.g. “DAXY”) to extend to different instances of the same category, only if it was modeled as a common noun by an indefinite article (“a DAXY”); however, they expected the same label to apply to only a single instance of the same category when it was modeled as a proper name, by omitting an article (“DAXY”). This suggests that in their second year of life, children are sensitive to differences between grammatical categories of words as indicated by articles. This provides some confidence that children in our experiments were also interpreting a label modified with an article as a common noun in Experiment IV (as contrasted with a proper name interpretation in Experiments I-III) and, as such, as

[23] Indeed, the sample size in Experiment IV was modest, as we had to prematurely stop the data collection due to COVID-19. However, Bayesian analyses allow us to be confident in our findings, as they revealed a strong evidence in favor of H1, both the FB condition (where H1 was around 10 times more likely than H0) and in TB (where H1 was around 17 times more likely than H0).

being generalizable across different members of the same category. This would, if true, entail that they thought of the two creatures as representatives of *distinct kinds.*

However, it should be noted that this assumption remains untested with the current set of experiments included in this project. It would be interesting to see if young children would also expect a novel common noun to generalize across different instances of the same category, but not to apply to any of the members of a different category in the naming FBT we employed here. For instance, in the beginning of the story, the agent could be presented with multiple green creatures and multiple orange creatures and she would put one of them (say, a green one) in a box and leave; in her absence an orange creature would replace it in the box and the rest of them would leave the scene. In the test scene, following the naming event, multiple instances of both categories would return to the scene and the child would be asked to point to "a daxy". If children really interpreted the label as a common noun (as we assume they did in our Experiment IV) we would expect their points to be about equally distributed among the creatures for one of which the agent believed to be in the box at the time of naming; by contrast we would expect them not to point to any of the creatures of the different kind (i.e. whose member was not the agent's intended referent). This would provide a stronger support for the assumption that children in our Experiment IV were really learning a common noun, and this is something that the future research should address to foster the claim that ToM reasoning presents a vehicle for mapping words onto their referents in a broader sense (i.e. extending beyond the case of proper names).

Chapter 3

General Discussion

The main aim of this thesis was to provide answers to two central questions: 1) can ToM represent identity false-beliefs prior to the age of four, and 2) can children learn a new word through the speaker's false-belief. Our findings provide affirmative answers to these questions, which bears important implications for both the nature of ToM and its development, as well as for the theories of word learning. Bellow, we discuss these implications in detail, as well as additional methodological and conceptual implications that stem from our findings.

3.1. Developmental continuity of ToM.

We elaborated earlier how demonstrating that identity false-beliefs are available in ToM's conceptual repertoire before the age of four, would challenge the dual-systems accounts (e.g. Butterfill & Apperly, 2013; Low, Butterfill, Apperly, & Rakoczy, 2016) and would support the alternative view, according to which ToM development involves a sophistication of a single ToM system (e.g. Leslie, 1987; 1994; Baillargeon, Scott, & Bian, 2016; Scott, Roby, & Baillargeon, *in press*).

Results from the set of experiments presented here, indeed, favor the single system view, which assumes that the fundamental properties of ToM reasoning—including the ability to reason about identity false-beliefs—are specified well before the age of four (Leslie, *ibid,* Baillargeon et al., *ibid*). We showed that young 3-year-old children can learn a new proper name solely through ToM reasoning—i.e. by correcting for the speaker's identity false-belief. Overall, we found that 76% of three-year-old children (Experiments I and II combined) pointed to the dog that the speaker falsely believed she was referring to at the time of naming, while, accidentally, referring to a different dog. We emphasize that the naming FBT we employed here, indeed, required children to understand the speaker's identity false-belief; when Jane refers to the dog in the box as "Fido", she believes that she is referring to a *different individual*, i.e. she has an identity false-belief. Hence, not only did this method allow us to obtain additional evidence for FB reasoning in children bellow the age of four (undermining the conceptual changes approach), but moreover, it provided the evidence that they can succeed on a special kind of false-belief scenario—one that involves mistaken identity (undermining the dual-systems approach).

We note, however, that although our results do provide support for developmental continuity of ToM as demonstrating identity false-belief reasoning in young 3-year-olds, future research is required to determine whether even younger children have such capacity. As we saw, Experiment III we conducted with 2.5-year-olds did not provide a conclusive evidence in this respect: although there was a slight trend towards pointing to the speaker's intended referent in the "Fido task", the results did not reach significance. Surprisingly, 2.5-year-olds also performed at chance on our adaptation of the traditional FBT with reduced demands, originally developed by Setoh et al. (2016). Recall that Setoh et al. demonstrated that even 2.5-year-olds can succeed on a verbal FBT if the processing demands are sufficiently reduced by 1) removing the target object from the scene and 2) allowing children to practice response generation and selection. This finding has been replicated by Grosso et al. (2019) and more recently by Scott, Roby and Setoh (2020). These findings are also congruent with Leslie's ToMM-SP model, according to which performance on FBTs involves both a domain specific Theory of Mind Mechanism (presumably innately specified) and a domain general Selection Processor; once the inhibitory demands are sufficiently reduced (e.g. by removing the target object from the screen and reducing demands on selecting a proper response) young children can manifest their actual ToM competence on a verbal FBT.

Successful replications of the original Setoh et al. (2016) makes us skeptical towards the hypothesis that the reason why 2.5-year-olds did not perform significantly above chance in our Experiment III was because of a conceptual deficit in their ToM reasoning. Rather, we are inclined to interpret this outcome as a result of the tasks related noise. This interpretation is also supported by Bayesian analyses that did not provide sufficient evidence for the null hypothesis, in neither the “Emma task” nor the “Fido task”. We already discussed several factors that could have impeded children’s performance in Experiment III, such as the format of the task (especially relevant for the “Emma task”), linguistic factors (relevant for both the “Emma task” and the “Fido task”) and cognitive fatigue (especially relevant for the “Fido task”). As emphasized, future research is needed to systematically test how each of these factors could affect young children’s FBT performance.

Importantly, Scott et al. (2020) also demonstrated that 2.5-year-olds can succeed on an *identity* FBT if the processing demands are reduced in the same way as in the “Emma task”—i.e. by including two practice trials and allowing children to practice response generation and selection, and by removing the target object from the scene. These findings provide additional support for early identity false-belief reasoning, with young 2.5-year-olds and are in line with the findings from our Experiments I and II, which support developmental continuity of a single ToM system. However, we have to point out that although our results with three-year-olds (Experiments I and II) are in line with Scott et al. (*Ibid*), we did not find the same successes in identity FB reasoning in 2.5-year-olds (Experiment III, “Fido task”). Why did 2.5-year-olds succeed in identity false-belief reasoning in Scott et al. (*Ibid*), but not in our Experiment III? This could be explained by methodological differences between the two identity FBTs. For instance, Scott et al. included the practice trials *within* the identity FBT. By contrast, in our Experiment III, children did not get to practice response generation and selection during “Fido task”, but only during “Emma task” that was preceding it. Since children performed at chance at our “Emma task” as well, it is reasonable to assume that it did not help reduce processing demands on the subsequent, “Fido task” either. Notably, we always presented “Fido task” after “Emma task”, which could have produced cognitive fatigue in young 2.5-year-olds, hence, impeding their performance instead of enhancing it. This assumption requires explicit testing in the future, where the order of the tasks would be counterbalanced.

Finally, we used proper names to index individual identities, while Scott et al. used the dual-aspects objects (e.g. a ball that can turn into a bunny). It is unclear, however, why this difference *per s*e would make “Fido task” more demanding compared to the identity task by Scott et al. Namely, three-year-olds performed better on “Fido task” compared to how they typically perform on both traditional and low-demand versions of FBTs (e.g. Barsch, 1996; Devine & Hughes, 2014; Wellmann et al., 2001), which suggests that “Fido task”, if anything, facilitates and does not hinders children’s performance. We discuss this point more in the following section.

3.2. The nature of the naming FBT.

As pointed out in the previous section, 3-year-olds performed better on our naming FBT compared how children typically perform on traditional FBTs, including its low-demand versions (e.g. Barsch, 1996; Devine & Hughes, 2014). Children younger than four typically perform well below chance on the standard FBT (e.g. Wellmann et al., 2001), and although their performance improves when low-demand versions are employed, they still don't perform significantly better than chance (Devine & Hughes, 2014). However, 3-year-olds performed significantly above chance on the naming FBT we employed in experiments I, II and IV, which suggests that certain aspects of this task facilitated the expression of ToM competence.

One possibility is that the naming task created a communicative setting, by including the aspect of a speaker communicating her referential intention to another agent. Previous research suggests that communicative setting, indeed, enhances ToM performance (Carpenter et al. 2002; Happé & Loth, 2002; Houston-Price, Goddard, Séclier, Grant, Reid, Boyden, & Williams, 2011; Southgate et al., 2010, see also Sperber, 2000; 2001). For instance, Carpenter et al. (2002) found that communicative setting helped 36-month-olds to account for an agent's false-belief in an FBT that included a word-learning aspect. In this study, 36-mointh-olds watched an actor putting an unfamiliar object in a box and leaving, after which the object was replaced with another one; actor then returns and expresses her desire for the object in the box, either by using a novel word (e.g. "a daxy") or a generic word ("a toy") and struggles to open the box. Children, then, choose which of the two object they will give to the actor. 36-month-olds were significantly better at this task than at the traditional FBT, indicating that the communicative setting facilitated their FBT performance. Similarly, in Southgate et al. (2010) an actor pointed to one of two boxes where she falsely-believed her desired toy to be, and asked 17-month-olds to give her that toy. Children were successful in choosing the desired toy over the one that was actually in the box the actor pointed to, suggesting that this communicative setting helped them account for the actor's false-belief. (Note that the study also included both a novel label condition ("shall we play with the *sefo*"?) and a generic condition ("shall we play with *it*?), but there was no difference between these conditions.)

Our experiments (I-IV) are in line with these studies, but they include several important differences that we discuss here. One such difference is the nature of the communicative context in our naming task, compared to the ones in the cited studies. Namely, our task included the aspect of the speaker's communicative and referential intention (e.g. Jane pointing to the box and announcing that Fido is in the box), as did the tasks used in the cited studies. However, in our task, communication occurs between the two characters in the story, but without inviting the child to participate. Hence, the child is in the role of an observer, whereas in Carpenter et al. and Southgate et al., the child takes an active (participative) role in communicating with the agent and helping her retrieve her desired toy. Hence, we can say that these other studies included a communicative setting in a stronger sense, as they relied on an interactive FBT where reasoning about agents' (communicative partners') mental states is highlighted. hence potentially advantageous for children's FBT performance.[24] By contrast, our task is a more passive version of an FBT, as the child is not one of the communicators, but an

[24] This is in line with other studies that found increased performance in even younger children, when an interactive, helping FBT is employed (e.g. Buttelman et al., 2009; Buttelman et al., 2015).

observer of a communicative situation between other agents. It is unclear whether observing a communicative context would have the same facilitating effect on false-belief reasoning as actively participating in it. Hence, although three-year-olds' performance was also enhanced in our naming FBT, we believe that communicative context did not play the same facilitative role in our case, as it presumably did in the case of Carpenter et al. (2002) and Southgate et al. (2010).

Another possibility is that our naming FBT is, as a matter of fact, closer to implicit FBTs than explicit FBTs. As pointed out earlier, the distinction between explicit and implicit FBTs is not fully clear, as it is sometime equated with verbal/non-verbal dichotomy and sometime with whether subjects' responses are prompted by an elicited-intervention question or by an explicit prediction question (see Scott & Baillargeon, 2017). So, although our naming FBT takes the form of a verbally narrated story, it does not require subjects to answer an explicit prediction question, which qualifies it for an implicit FBT. As a matter of fact, (and critically) it does not explicitly require subjects to think about the agent *at all*, as the agent is entirely absent from the test scene: subjects are simply asked to point to "Fido/ the daxy". This is an especially interesting feature of our naming FBT: the only way to succeed in referential disambiguation is through the speaker's false-belief. Yet in the test scene, children are neither explicitly nor implicitly prompted to predict/explain the agent's actions (nor to think about the agent at all).

The fact that young three-year-olds successfully discerned the correct (intended) referent, suggests that they had implicitly computed the speaker's false-belief during the naming scene (when the word was uttered for the first time) and spontaneously used this representation to answer the question that was not even concerning the agent (or agents in general). This finding is of a great importance for current conceptualizations of early ToM ability as it extends its role beyond predicting/explaining agents' actions to learning about other domains of knowledge, such as language. In addition, it suggests that ToM computes agent's mental states online, as the child observes the unfolding sequence of events, instead of computing it retrospectively, once she is explicitly asked the prediction question. This is, again, in line with Leslie's (*ibid*) ToMM hypothesis, according to which ToM ability rests on a domain specific mechanism that attributes mental states to agents *spontaneously,* rather than on an explicit request. However, the question of when exactly the speaker's false-belief is computed in our naming FBT needs to be explicitly tested in the future.

3.3. ToM as a vehicle for word learning.

One of the main goals of this dissertation was to explore the intersection between ToM and word learning— specifically, to test the idea of ToM as vehicle for establishing word-meaning links. Testing this assumption is of a fundamental importance for our understanding of early word acquisition, as it provides insights into how profoundly mentalistic this process really is.

Our results, indeed, provide compelling evidence for a strong mentalism in word learning. We demonstrated that young three-year-old children successfully learned a new label-referent mapping solely by appeal to the speaker's epistemic state (i.e. false-belief) about the referent. Importantly, children successfully formed a label-referent link, both in the case when the label was modeled as a proper name (Experiments I and II) and when it was modeled as a common noun (Experiment IV), which suggests a more general role of ToM in referential disambiguation.

Our findings are in line in line with the social learning accounts of word learning (e.g. Baldwin et al., 1996; Baldwin & Tomasello, 1998; Bloom, 2000; Tomasello & Kruger, 1992) as they suggest the key role of agency reasoning in word learning and, at the same time, they undermine associative learning accounts (e.g. Smith & Yu, 2008, Skinner, 1957). Indeed, it is clear that associative learning wouldn't be sufficient to explain children's successes in the set of experiments presented here, given that all the externally observable evidence in the naming scene pointed in the direction of a *wrong* referent (i.e. the entity that was in the box when the actor uttered the label, while at the same time pointing to the box). However, children were not mapping the label onto this actually cued referent.[25] Instead, they were correctly selecting the intended referent in response to the test question, although the intended referent was not even present when the label was first introduced. The fact that children were mapping the label with the *intended* rather than with the *actual* (given the directly observable evidence) referent, indicates that young children are not acting as behaviorists in the process of word learning.

This is congruent with previous research, which demonstrated that young learners don't simply track co-occurrences of labels and referents across multiple contexts, but that they rely on quantity and quality of various social-attentive cues that the speaker produces (e.g. Medina et al., 2011; Trueswell et al., 2016; Tomasello & Kruger, 1992). Importantly, young learners seem to take into consideration agents' intentional states (Jin & Song, 2017) and their perceptual access to the scene (Koening & Echols, 2003) in labeling events, suggesting that they go beyond directly observable cues provided by the

[25] Nor onto the box, as, perhaps, an extreme version of associationism would predict.

speaker. This is, again, compatible with our suggestion that young children are mentalists in the process of word learning. However, *how profound is this mentalizing?*

In the set of experiments presented in this dissertation we demonstrated that at least by 40-months of age, word learners are, indeed, quite profound mentalists as they manifest quite sophisticated level of agency reasoning. They seem to ascribe a complex propositional attitude (i.e. a false-belief) to the speaker, which is then used to guide the process of referential disambiguation. In our naming FBT, the object that the speaker cued as the referent (indicated by social-attentive cues—ostensive pointing and the eye-gaze direction) was *not* the correct referent, as she had a false-belief about whom/ what was in the box. So, to succeed on this task, children had to override directly observable evidence provided by the speaker's socio-attentive cues; in addition, they had to couple the speaker's referential intention (*Jane intends to refer to X by "Fido"daxy"*) with her false-belief about the referent (*Jane (falsely) believes that it is X in the box*) to form the correct word-referent mapping. Our findings reveal that young three-year-old children indeed engage in such sophisticated and flexible ToM reasoning to select the correct referent of a new label in a referentially ambiguous context.

We can, hence, conclude with confidence that the set of experiments presented in this dissertation provides a strong evidence that young preschoolers spontaneously engage in profound ToM reasoning to discern the correct referent of a novel label. It appears that for a young child the most likely candidate for a word's referent is not simply an object that the child attends to herself as she hears the label, but it is also not necessarily the one that the speaker attends to in the factual context as she utters the label. By contrast—and as indicated by results of this research—the most likely candidate for the label's referent is an object that appears in a context decoupled from the factual world; that is, as the content of the speaker's representation of the world.

Conclusion

This thesis provides a bridge between two major questions of developmental psychology: how a young child's ToM develops and how a young child learns novel words. I argued that these two questions are, in fact, fundamentally intertwined, as young word learners critically attend to speakers not only as physical agents who act in the surrounding world, but as mental agents who represent this world in particular ways.

I argued that genuine, meta-representational ToM—in a sense of Leslie's domain specific, neurocognitive Theory of Mind Mechanism, ToMM (Leslie, 1987)—is already available in children younger than four (contrary to the claims of the conceptual shifts approaches (e.g. Gopnik & Wellmann, 1994; Perner, 1991) and of the dual-systems approaches (e.g. Butterfill & Apperly, 2013; Low & Watts, 2013)). Moreover, and critically, not only is it available prior to the age of four for the purposes of predicting and/or explaining others' actions, but to support learning in different domains of knowledge, such as learning of new words. This extends what were traditionally thought to be the key roles of early ToM and suggests its important role in narrowing down the space of possible meanings of a new word.

To test the idea of ToM as a vehicle for learning new words, I developed a new version of the naming false-belief task, where young children ought to rely on the speaker's false-belief in order to figure out whom/what a novel label denote in a referentially ambiguous context. I conducted a set of experiments using this method to show that by at least the age of 40-months, children successfully correct for the speaker's false belief to learn a new word. As elaborated in the text, these findings bear implications of fundamental significance for both the nature of ToM ability and for conceptualizations of the process of word acquisition— they indicate that 1) young children spontaneously engage in genuine meta-representational ToM before the age of four, even if they are not explicitly required to think about agents at all, and 2) they indicate that this genuine ToM reasoning underlies the process of mapping a word and its referent, suggesting a profoundly social nature of word acquisition. It appears, hence, that young word learners prioritize the abstract context of the speaker's mind, over the factual context of directly observable evidence to infer the meaning of a word.

Bibliography

Apperly, I.A. & Butterfill, S.A. (2009). Do humans have two systems to track beliefs And belieflike states? *Psychological Review, **116***(4), 953–70.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science, **14***(3), 110–118.

Baillargeon, R., Z. He, P. Setoh, R. Scott, S. Sloan, & D. Yang. (2013). False-belief understanding and why it matters: the social-acting hypothesis. In M. Banaji & S. Gelman (Eds.), *Navigating the social world* (pp. 88-95). Oxford: Oxford University Press.

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., Wu, D., & Bian, L. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer, P. R. Shaver, E. Borgida, and J. A. Bargh (Eds.), *APA Handbook of Personality and Social Psychology: Attitudes and Social Cognition* (Vol 1, pp. 79-150). Washington, DC: American Psychological Association.

Baillargeon, R., Scott, R.M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology**,** **67**,* 159-186.

Baker, S.T., Leslie, A.M., Gallistel, C.R., & Hood, B. (2016). Bayesian change-point analysis reveals developmental change in a classic theory of mind task. *Cognitive Psychology, **91**,* 1-26.

Baldwin, D.A. (1991). Infants' contributions to the achievement of joint reference. *Child Development, **62**,* 875-890.

Baldwin, D.A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, **20***(2), 395-418.

Baldwin, D.A. (1993a). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, **29***(5), 832-843.

Baldwin, D.A., Markman, E.M., Bill, B., Desjardins, R.N., & Irwin, J.M. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, **67***(6), 3135-3153.

Baldwin, D.A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In E. Clark (Ed.), *Proceedings on the Twenty-ninth Annual Child Language Research Forum, **29*** (pp. 3-23). Cambridge, UK: Cambridge University Press.

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

Bartsch, K. (1996). Between desires and beliefs: Young children' s action predictions. *Child Development, **67**,* 1671-1685.

Bélanger, J., & Hall, D.G. (2006). Learning Proper Names and Count Nouns: Evidence From 16- and 20-Month-Olds. *Journal of Cognition and Development, 7*(1), 45-72.

Bloom, P. (2000). *How Children Learn the Meanings of Words.* The MIT Press. Cambridge, Mass.

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.

Bruner, J. (1974/1975). From communication to language – A psychological perspective. *Cognition, **3***(3), 255-287.

Buttelman, D., Carpenter, M, & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, **112**,* 337-342

Buttelmann, F., Suhrke, J. & Buttelman, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, ***131***, 94–103.

Butterfill, S.A., & Appely, I.A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language, **28***, 606-637.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). Cambridge University Press.

Carlson, S.M., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development, **69***, 672-691.

Carpenter, M., Call, J. & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology, **20***, 393-420.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences, **110***(28), 11278-11283.

Carruthers, P, (2013) Mindreading in infancy. *Mind and Language, **28***, 141-172.

Carruthers, P. (2015). Two Systems for Mindreading? *The Review of Philosophy and Psychology, 7*(1), 141-162.

Cassidy, K.W. (1998). Three- and 4-year-old children's ability to use desire-and belief based reasoning. *Cognition, **66***, B1-B11.

Clemens, W.A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, **9**,* 377-395.

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

De Villiers, J.G. (2005). Can language acquisition give children a point of view? In J.W. Astington & J.A. Baird (Eds.), *Why language matters for theory of mind* (pp. 123-143). New York: Oxford University Press.

Devine R.T. & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Development, **85**, 1777-94.*

Dunn, J., & Brophy, M. (2005). Communication, relationships, and individual differences In children's understanding of mind. In J.W. Astington & J.A. Baird (Eds.), *Why language matters for theory of mind* (pp. 50-69). New York: Oxford University Press.

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E. & Rakoczy, H. (2017). Are there signature limits in early theory of mind?. *Journal of Experimental Child Psychology, **162**, 209-224.*

Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J. & Király, I. (2020). Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience, **43**,* 1-8.

Frege, G. (1948). Sense and Reference. *The Philosophical Review, **57***(3), 209-230.

Friedman, O., & Leslie, A. M. (2004). Mechanisms of belief-desire reasoning. Inhibition and bias. *Psychological Science, **15***(8), 547-52.

Friedman, O., & Leslie, A.M. (2007). The conceptual underpinnings of pretense: Pretending is not 'behaving-as-if'. *Cognition, **105**,* 103-124.

Fridman, O., Neary, R.K., Burnstein, C.L., & Leslie, A.M. (2010). Is young children's recognition of pretense metarepresentational or merely behavioral? Evidence from 2- and 3-year-olds' understanding of pretend sounds and speech. *Cognition, **115**,* 314-319.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer. (1999). Human simulations of vocabulary learning. *Cognition, 73,* 135-176.

Gleitman, L.R. (2009). The learned component of language learning. In M. Piattelli-Palmarini, P. Salaburu, & J. Uriagereka (Eds.), *Of minds and language: Encounters with Noam Chomsky* (pp. 239-256). Oxford: Oxford Press.

Gleitman, L.R., & Gleitman, H. (1992). A picture is worth a thousand words – but that's the trouble: Conceptual and structural factors in vocabulary acquisition. *Current Directions in Psychological Science, **1***(1), *31-35.*

Gleitman, L.R., & Fisher, C. (2005). Universal aspects of word learning. In McGilvray (Ed.), *The Cambridge Companion to Chomsky.* Cambridge: Cambridge University Press.

Gleitman, L.R., & Trueswell, J.C. (2018). Easy Words: Reference Resolution in a Malevolent Referent World. *Topics in Cognitive Science*, pp.1-26. ISSN:1756-8765 online DOI: 10.1111/tops.12352.

Gleitman, L.R., Liberman, M.Y., McLemore, C.A., & Partee, B.H. (2019). The Impossibility of Language Acquisition (and How They Do It). *Annual Review of Linguistics, **5**, 1-24.*

Grice, H.P. (1969), Utterer's Meanings and Intentions, *The Philosophical Review, **78**, 147-177.* Reprinted as ch 5. Of Grice 1989, 86-116.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J. & Sodian, B. (2016). 33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016). *Infant Behavior & Development, **54,*** 151-155.

Gopnik, A. & Wellman, H. (1994). The "theory theory". In L. Hirschfiled & S. Gelman (Eds.) *Domain specificity in culture and cognition.* New York: Cambridge University Press.

Gopnik, A. & Meltzoff, A.N. (1997). *Words, Thoughts and Theories. Cambridge*. MA: MIT Press.

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

Harris, P.L. (1995). Imagining and pretending. In M. Davies & T. Stone (Eds.). *Mental simulation: Evaluations and applications*. Oxford, Blackwell.

Harris, P.L., Kavanaugh, R.D., & Dowson, L. (1997). The Depiction of imaginary transformations: Early comprehension of a symbolic function. *Cognitive Development, **12**,* 1-19.

Hall, D.G., Lee, S., & Bélanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology, **37**,* 298-307.

Hall, D.G. (2009). Proper Names in early Word Learning: Rethinking the Theoretical Account of Lexical Development. *Mind and Language, **24***(4), 404-432.

Happé, F., & Loth, E. (2002). 'Theory of Mind' and Tracking Speakers' Intentions. *Mind and Language, **17**,* 24-36.

He, Z. Bolz, M., & Baillargeon, R. (2012). 2.5-year-olds succeed at a verbal anticipatory-looking task. *British Journal of Developmental Psychology, **30**, 14-29.*

He, Z., Bolz., M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: evidence from change of location and unexpected-contents violation of expectation tasks. *Developmental Science*, ***14,*** 292–305.

Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental Science, **17**,* 647-59.

Houston-Price, C., Goddard, K., Séclier, C., Grant, S.C., Reid, C.J.B, Boyden, L.E., & Williams, R. (2011). Tracking speakers' false beliefs: is theory of mind available earlier for word learning? *Developmental Science, **14**(4), 623-34.*

Jin, K., & Song, H. (2017). You changed your mind! Infants interpret a change in word as signaling a change in an agent's goals. *Journal of Experimental Child Psychology, **162**, 149-162.*

Jin, K., Kim, Y., Song, M., Kim, Y., Lee, H., Lee, Y., Cha, M. & Song, H. (2019). Fourteen- to Eighteen-Minth-Old Infants Use Explicit Linguistic Information to Update an Agent's False-Belief. *Frontiers in Psychology, **10**, 1-12.*

Jeffreys, J. P. (1935). Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society, **31**(2), 203-222.*

Katz, N., Baker, E., & Macnamara, J. (1974). What's a name? A study of how children learn common and proper names. *Child Development, **45**,* 469-473.

Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science, **15**,* 113-122.

Kovács, A.M., Téglás, E. and Endress, A.D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, **330***, 1830–4.

Kripke, S.A. (1980). *Naming and Necessity.* Harvard University Press.

Krcmar, M., & Cingel, D.P. (2014). Parent-Child Joint Reading in Traditional and Electronic Formats. *Media Psychology, **17**, 262-281.*

Koenig, M.A., & Echols, C.H. (2003). Infants' understanding of false labeling events: the referential roles of words and the speakers who use them. *Cognition, **87**,* 179-208.

Landau, B., & GLeitman, L.R. (1985). *Language and experience: Evidence from the blind child.* Cambridge MA: Harvard University Press.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge: Cambridge, University Press.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, **94***(4), 412-426.

Leslie, A. M. (1994a). Pretending and believing: Issues in the theory of ToMM. *Cognition, **50***, 211–238. (Reprinted in J. Mehler & S. Franck (Eds.), *Cognition on Cognition* (pp. 193–220). Cambridge, MA: MIT Press, 1995).

Leslie, A. M. (1994b). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp.119–148). New York: Cambridge University Press.

Leslie, A.M. & Frith, U. (1990). Prospects for a Cognitive Neuropsychology of Autism: Hobson's Choice*. Psychological Review, **1**,* 122-131.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, **1***(2), 247-253.

Leslie, A.M. (2000). 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp. 1235–1247). MIT Press.

Leslie, A.M., Friedman, O., & German, T.P. (2004). Core Mechanisms in "Theory of Mind". *Trends in Cognitive Sciences*, ***12***, 528-533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, **50***(1), 45-85.

Locke, J. (1968). *An essay concerning human understanding.* Cleveland, Ohio: World Publishing Co. Original publication 1690.

Low, J., & Watts, J. (2013). Attributing False Beliefs About Object Identity Reveals a Signature Blind Spot in Humans' Efficient Mind-Reading System. *Psychological Science, **24***(3), 305-311.

Low, J., Drummond, W., Walmsley, A. & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. Child Development, 85, 1519-1534

Low, J., Apperly, I.A., Butterfill, S.A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer of the two-systems account. *Child Development Perspectives, **10**,* 184-189.

Luo, Y., & Baillargeon, R. (2010). Towards a mentalistic account of early psychological reasoning. *Current directions in psychological science, **19***(5), 201-307.

Macnamara, J. (1982). *Names for Things.* Cambridge, MA: MIT Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's Use of Mutual Exclusivity to Constrain the Meanings of Words. *Cognitive Psychology, **20***, 121-157.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences, **108**,* 9014–9019.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, **38**,* (Serial No. 149).

Nichols, S., & Stich, S. (2000). A cognitive theory of pretense. *Cognition, **74**,* 115-147.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, **308***(8), 255-258.

Onishi, K.H., Baillargeon, R., & Leslie, A.M. (2007). 15-month-old infants detect violations in pretended scenarios. *Acta Psychologica, **124***, 106-128.

Papafragou, A., Friedberg, C., & Cohen, M. L. (2017). The role of speaker knowledge in children's pragmatic Inferences. *Child Development, **89***(5), 1642-1656.

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Perner, J. (1995). The many faces of belief: reflections on Fodor's and the child's theory of mind*.* *Cognition, **57***, 241–69.

Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science, **308**,* 214–216.

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anticognitivism. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology international perspectives on psychological science* (Vol. 1, pp. 246-261). Hove, UK: Psychology Press.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York, Norton.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, **1**,* 512-526.

Putnam, H. (1975). The Meaning of "Meaning". *Minnesota Studies in the Philosophy of Science, 7, 131-193.*

Quine, W.V.O. (2013). *Word and Object.* Cambridge, Massachusetts: MIT Press ISBN 9780262670012

Rakoczy, H., Bergfeld, D., Schwarz, I, & Fizke, E. (2014). Explicit Theory of Mind is Even More Unified Than Previously Assumed: Belief Ascription and Understanding Aspectuality Emerge Together in Development. *Child Development*, ***86***(2), 486-502.

Richards, D. & Goldfarb, J. (1986). The episodic memory model of conceptual development: An Integrative viewpoint. *Cognitive Development, **1**,* 183– 219.

Roth, D., & Leslie, A.M. (1991). The recognition of attitude conveyed by utterance: A study of autistic and preschool children. *British Journal of Developmental Psychology, **9**,* 315-330.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, **34***(3), 265-293.

Russell, B. (1905). On Denoting. *Mind, New Series, **14***(56), 479-493.

Saffran, J.R., & Kirkham, N.Z. (2018). Infant Statistical Learning. *Annual Review of Psychology, **69**,* 181-203.

Scott, R.M., & Baillargeon, R. (2009). Which Penguin is This? Attributing False-Beliefs About Object Identity at 18 months. *Child Development, **80***(4), 1172-1196.

Scott, R.M., Baillargeon, R., Song, H., & Leslie, A.M. (2010). Attributing false-beliefs about non-obvious properties at 18 months. *Cognitive Psychology, **61***(4), 366-395.

Scott, R.M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-Belief understanding in 2.5-year-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental Science, **15***(2), 181-193.

Scott, R.M., Richman, J.C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology, **82**,* 32-56. doi: 10.1016/j.cogpsych.2015.08.003

Scott, R.M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Science, **21***(4), 237-249.

Scott, R.M., Roby, E., & Setoh, P. (2020). 2.5-year-olds succeed in identity and location elicited-response false-belief tasks with adequate response practice. *Journal of Experimental Child Psychology, **198**, 1-13.*

Scott, R. M., Roby, E., & Baillargeon, R. (in press). How sophisticated is infants' theory of mind? To appear in O. Houdé & G. Borst (Eds.-in-chief), *Cambridge handbook of cognitive development*. Cambridge, England: Cambridge University Press.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-Month-Olds Really Attribute Mental States to Others? A Critical Test, *Psychological Science*, ***22***(7), 878-880.

Setoh, P., Scott, R.M., & Baillargeon, R. (2016). 2.5-year-olds succeed at traditional false-belief tasks with reduced executive demands. *Proceedings of the National Academy of Science, **113***(47), 13360-13365.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition**, 38***(1), 1-12.

Sirois, S., & Jackson, I. (2007). Social cognition in infancy: A critical review of research on higher order abilities. *European Journal of Developmental Psychology, **4***(1), 46-64.

Skinner, B. F. (1957). *Verbal Behavior.* Engle-wood Cliffs, NJ: Prentice Hall.

Smith, L.B. (2000). How to learn words: An associative crane. In R. Golinkoff and K. Hirsh-Pasek,(Eds.), *Breaking the word learning barrier* (pp. 51– 80). Oxford: Oxford University Press

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, **106**,* 1558-1568.

Smith, L.B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The Developing Infant Creates a Curriculum for Statistical Learning. *Trends in Cognitive Sciences, **4**, 325-336.*

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

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

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, **18**(7), 587-92.*

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science, **13**,* 907-912.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, **18**, 587-592.*

Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: the influence of trickery. *Journal of experimental child psychology, **56***(2), 135-148.

Surtees, A.D.R., Butterfill, S.A., & Apperly, I.A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, **30***, 75-86.

Spelke, E. S. (2000). Core Knowledge. *American Psychologist, **55**(11), 1233-1243.*

Spelke, E.S. & Kinzler, K.D. (2007). Core knowledge. *Developmental Science, **10**(1), 89-96.*

Sperber, D. (2000). *Metarepresentations in an evolutionary perspective*. In D. Sperber (Ed.) *Metarepresentations: A Multidisciplinary Perspective*. New York: Oxford University Press.

Sperber, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language, **17**,* 3-23.

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development, 77*, 465-481.

Tincoff, R., & Jusczyk, P. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychologic al Science, **10**,* 172-175.

Tincoff, R., & Jusczyk, P. (2000). Do 6-month-olds link sound patterns of common nouns to new exemplars? *Paper presented at the International Conference of Infant Studies, Brighton, UK.*

Tomasello, M.. & Kruger, A.C. (1992). Joint attention on actions: acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language, **19***(2), 311-333.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, **66***(1), 126–156. Doi:10.1016/j.cogpsych.2012.10.001

Trueswell, J.C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L.R. (2016). Perceiving referential intent: Dynamics of reference in natural parent child interactions. *Cognition, **148***, 117-135.

Vouloumanos, A. (2008) Fine-grained sensitivity to statistical information in adult word learning. *Cognition, **107**,* 729–742.

Walker-Andrews, A., & Kahana-Kalman, R. (1999). The understanding of pretence across the second year of life. *British Journal of Developmental Psychology, **17**,* 523-536.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, **311***, 1301-1303.

Warneken, F., & Tomasello, M. (2007). Helping and Cooperation at 14 Months of Age. *Infancy, **11***(3), 271-294.

Walton, K.L. (1978). Fearing fictions. *The Journal of Philosophy, **75**,* 5-27.

Wang, L. & Leslie, A.M. (2016). Is Implicit Theory of Mind the ‘Real Deal’? The Own-Belief/True-Belief Default in Adults and Young Preschoolers. *Mind & Language*, ***31***(2), 147-176.

Wang, L., Hemmer, P., & Leslie, A.M. (2019). A Bayesian framework for the development of belief-desire reasoning: Estimating inhibitory power. *Psychonomic Bulletin & Review, **26**, 205-221.*

Wellman, H.M. & Woolley, J.D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, **35**, 245-275.*

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: the truth about false-belief. *Child Development, **72***, 655–84.

Wellman, H.M. (2014). *Making Minds: How Theory of Mind Develops*. New York, NY: Oxford University Press.

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

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track single meaning during word learning: Evidence for propose-but-verify. Language Learning and Development. *Language learning and development, **12***(3), 252-261.

Yu, C., & Smith, L.B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, **18**,* 414-420.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development, **4***(1), 32-62.

Zaitchik, D. (1991). Is only seeing really believing? Sources of true belief in the false belief task. *Cognitive Development, **6***(1), 91-103.

APPENDICES

1. Appendix A

Script of the naming FBT ("Fido task") in Experiment I and Experiment II (False-Belief).

"This is a story about a girl named Jane. Here's Jane! And look what Jane has! What are these? That's right, they're doggies. And look! They look the same! But you know what? One of these doggies is tired, so Jane puts him in the box to sleep. And then she goes away! So, Jane is no longer in the room, she cannot see what's gonna happen. And look what's happening while Jane is gone! This doggy… jumps out of the box… and runs away! And then… this other doggy… jumps into the box! Wow, that was fun! But did Jane see what just happened? That's right, she didn't see because she was not in the room! But look, Jane is coming back with her brother, Timmy! Timmy points to the box and asks, "Who is in this box?". Jane says, "Fido! Fido is in the box!" Timmy asks, "Really, Fido is in the box?" Jane says, "Yes, Fido is in the box. I put Fido in the box to sleep" (*in Experiment II, Jane only says, "Yes, Fido is in the box!"*). "Alright!", they say, "Let's get some food!". And look, they're leaving! Look they're gone! But look who is back! And look who popped out! Now, that's the end of my story… But can I ask you something? Can you point to Fido?"

2. Appendix B

Script of the naming FBT in Experiment IV.

“This is a story about a girl named Sarah. Look, here’s Sarah! And look what Sarah has! Hmm.. they’re cute! And look, Sarah puts one of these in a box! See? And then, Sarah goes away! So, Sarah is no longer in the room, she cannot see what’s gonna happen! And look what happens while Sarah is gone! This one… flies out of the box… right through the window! And then, this other one… jumps into the box! Wow, that was fun! But Sarah didn’t see what just happen, because she was not in the room! But look, Sarah is coming back now, with her friend, Billy! Billy points to the box and asks, “What is in this box?” Sarah says, “A daxy! A daxy is in the box!” Billy asks, “Really? A daxy is in the box?” Sarah says, “Yes, a daxy is in the box!” “Alright,” they say, “Let’s get some food!” And look, they’re leaving! Look they’re gone. But look, this one flies back! And look, this one pops out! Now, that’s the end of my story... But can I ask you something? Can you point to the daxy?”