MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

A nonverbal Theory of Mind hiding paradigm: Ambiguous displacement in preschoolers

verfasst von / submitted by

Karla Arzberger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / Degree programe code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / Degree programe as it appears on the student record sheet:	Master's programe in Psychology UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Claus Lamm
Mitbetreut von / Co-Supervisor:	Dr. Catherine-Noémie Alexandrina Guran, B.Sc. MSc

Abstract

We implemented a nonverbal Theory of Mind hiding paradigm in a preschool sample (N = 33) to investigate whether children followed a confederate's misleading cue under a true belief and a false belief condition. A child witnessed a confederate (hider) hiding a sticker in one of two containers and another confederate (communicator) cuing the container holding the sticker during familiarization trials. In two belief trials, the communicator cued the wrong container after having either witnessed (true belief), or not having witnessed (false belief) the hider displacing the sticker. Under the false belief condition, the children were able to choose the correct container holding the sticker significantly above chance level. Under the true belief condition, the children were unable to choose the correct container above chance level. The belief the communicator held did not significantly influence the choices the children made. We detected a main effect of age, indicating that the older children in our sample were more likely to choose the correct container under either belief condition. Our results indicate that some children aged five to six years old struggled with attributing mental states in our paradigm or did not base their container choice on the belief the communicator held. We argued that the children in our sample were in a sensitive age window regarding the development of the Theory of Mind abilities needed to master both belief trials successfully. An exploratory analysis revealed that speech comprehension and inhibitory control did not influence the children's choices, allowing preliminary conclusions about the paradigm's nonverbal nature and lowered task demands to be drawn.

Table of Contents

1. Introduction

A substantial part of successfully navigating and interpreting everyday life consists of representing the mental states of others, which allows individuals to make sense of the actions around them. For example, one could interpret a situation such as "Why is Simon reaching into his pocket after ordering coffee?" with the conclusion "Because he wants to pay and *thinks* he put his wallet there."

Interpreting and understanding others' intents, wishes, and beliefs is a social-cognitive skill that has been subsumed under the term *Theory of Mind* (ToM). This concept has been crucial in comparative and developmental psychology ever since its emergence in 1978, when Premack and Woodruff asked the question "Does the chimpanzee have a Theory of Mind?" They demonstrated that chimpanzees showed an understanding of human intentions (Premack & Woodruff, 1978).

Some years later, Wimmer and Perner pioneered investigations of ToM in a broad age span of children. They provided evidence that this social-cognitive skill develops fundamentally between the ages of four and six years (Wimmer & Perner, 1983). Later research has shown that crucial developments that are prerequisites to ToM acquirement occur in early infancy. Infants appear to interpret humans as influenceable by communicative signals and their behavior as goal directed (Brandone & Wellman, 2009; Csibra, 2010; Flavell, 2004; Meltzoff, 1995; Reid et al., 2007). By the age of three, children seem to have developed an *implicit ToM* (Flavell, 2004). The term implicit ToM is used to describe the stage in ToM development in which children have acquired an early, inchoate, unconscious understanding of mental states, but lack an explicit understanding of *beliefs*, which they demonstrate in later developmental stages (Apperly & Butterfill, 2009; Low & Perner, 2012; Schneider et al., 2014). Beliefs are mental states meant to accurately represent the world (James, 1889). A sign of implicit ToM is children's correct use of desire terms (e.g., "want" and "see") and them starting to perceive relationships among desires, emotions, actions, and their results (Bartsch & Wellman, 1995). Children understand that they and another person can have different desires about an object (Wellman & Liu, 2004), while being unable to understand beliefs explicitly, which is indicated by young children's failure on explicit belief tasks (Kristen et al., 2006; Low & Perner, 2012; Wimmer & Perner, 1983).

A finding important to the current study is that children start to explicitly understand beliefs in subsequent stages of ToM development (Kristen et al., 2006; Wellman & Liu, 2004; Wellman et al., 2001). Between the ages of three and four years, children become aware that their belief and another person's belief can differ (Wellman & Liu, 2004), for instance, if they know something their counterpart does not.

Wimmer and Perner (1983) were the first to apply *false beliefs* (FBs) in ToM paradigms. A false belief is a belief that does not conform to reality and is in opposition to a *true belief* (TB). In FB tasks, a story is usually told about a protagonist experiencing a certain incident that leads them to hold a FB. An example is the Maxi task (Wimmer & Perner, 1983), in which the protagonist, Maxi, hides his chocolate in a box and leaves. While he is absent, a friend removes the chocolate from the box and hides it in another box. Afterwards, Maxi returns to the scene. The child is then asked where Maxi would search for his chocolate. To answer correctly, the child must hold two opposing mental states: their own (knowing the location of the chocolate switched in Maxi's absence) and Maxi's (not knowing the location of the chocolate switched in his absence). The child is asked to anticipate the protagonist's actions based on his FB.

Classic FB tasks have been criticized for three main reasons. First, the tasks impose high linguistic demands (Call & Tomasello, 1999; Krachun et al., 2009; Russell et al., 1991). They ask children how a fictional protagonist would act or what they would believe under the explained circumstances. For the child to answer, they must comprehend the imaginary story and articulate a correct answer. As a result, FB understanding and speech comprehension are intertwined in such FB tasks. Second, authors have claimed that many tasks demand children to inhibit salient responses (Apperly, 2012; Flavell, 2004; Russell et al., 1991), imagine something that is in clear opposition to reality, or imagine people or figures behaving differently than they normally would (Russell et al., 1991). These scenarios demand children to inhibit salient information for the sake of the correct answer, which requires inhibitory control, which is an important part of executive function (Chasiotis et al., 2006; Posner & Rothbart, 2000). Hence, children may fail FB tasks due to their lack of executive function rather than FB understanding (Carlson & Moses, 2001; Carlson et al., 1998). Third, classic verbal FB tasks are valueless for investigating nonverbal individuals. The ToM development of nonverbal children, adults, and primates has been questioned, which evinced the need for nonverbal tasks (Baron-Cohen et al., 1986; Call & Tomasello, 1999; Figueras-Costa, 2001; Krachun et al., 2009; Lonardo et al., 2021; Porter et al., 2008; Simlesa et al., 2021). These three criticisms raise the question of whether classic verbal FB tasks schould remain the preferred tool to investigate ToM development. Additionally, previous attempts to implement nonverbal ToM paradigms include questionable approaches (see Figueras-Costa, 2001; Simlesa et al., 2021), and further revision is needed.

Adjustments to classic FB tasks have already been made by Call and Tomasello (1999), who approached ToM paradigms in an active, engaging manner. They advocated for tasks that did not require imagining epistemic states. In their paradigm, children observed the behavior of a person who held a FB and reacted based on their comprehension of the FB. Furthermore, Call and Tomasello (1999) criticized that previous nonverbal tasks required children to master complicated task demands (e.g., memory, attention etc.) while simultaneously demonstrating an understanding of mental states. They suggested an experimental setup in which subjects would undergo multiple familiarization trials to ensure that the task requirements were sufficiently met (Call & Tomasello, 1999). The crucial FB trial would be conducted only after the child mastered the familiarization trials. With these adjustments, they sought to eliminate the possibility of children's failure in FB tasks due to extraneous task demands.

A similar perspective regarding FB tasks has been expressed by Schilbach et al. (2013), who proposed a shift in research towards a "second-person" engagement of the subject. They argued that observing or thinking about others cannot be compared to the fundamentally different experience of interacting with them (Schilbach et al., 2013) and consequently supported shifting classic FB tasks to active paradigms in which the subject has a role to play. Relating to another person by engaging directly and emotionally allows subjects to experience the other as naturally human without the need to superficially infer mental states (Schilbach et al., 2013). This novel approach facilitates shared goals, intentions, and actions, which enable participants to create inartificial experiences from which ToM development can be concluded (Byom & Mutlu, 2013).

The current study has implemented these considerations by lowering speech comprehension and task demands as well as by engaging our subjects in a real-life, engaging scenario. Rather than asking children to infer an imaginary protagonist's mental state in a verbal story, we integrated a nonverbal ToM hiding paradigm in a motivating environment. In this approach, the children are invited to play a game in which they can collect rewards and are then confronted with a FB and a TB condition to which they must react. The created environment thus allows belief comprehension to be investigated with the advantages of second-person engagement.

The proposed hiding paradigm is similar to that of Call and Tomasello (1999) and implements their suggested adjustments as well. In their study, a confederate (hider) hid a reward in one of two identical containers while another confederate (communicator) observed the hiding. A child observed both confederates but could not directly see in which container the reward was placed. In the familiarization trials, the communicator marked the location of the reward with a wooden block. The child learned to trust the communicator's correct cue to obtain the reward. In the FB trials, after the hider hid the reward while beeing watched by the communicator, the communicator left the room. While she was outside, the hider switched the position of the containers. When the communicator returned, she marked the incorrect container as the reward location had switched in her absence. If the child identified that the communicator held a FB about the location of the reward, they should nevertheless have been able to retrieve the reward from the correct container. Children aged five years old were able to do so, while younger children and apes failed. Krachun et al. (2009) transformed the paradigm into a competitive context and reached the same conclusions about five-year-olds as well as younger children and apes. Importantly, Krachun et al. (2009) included a TB condition that contrasted the FB condition. Under the TB condition, the communicator left the room after witnessing the hiding, but nothing occured while she was outside. When she reentered the room, the hider switched the reward. The communicator watched the switching, held a TB about the location of the reward, and consequently cued the correct container.

The current experimental setup, which is based on an analogous experiment by Lonardo et al. (2021), changed two important aspects of the nonverbal ToM hiding paradigm. First, the children explicitly observed where the hider placed the sticker in most of the familiarization trials and both belief trials. Therefore, the executive demands were lowered as the children did not have to solely trust the communicator's cue to infer where the sticker must have been placed. Second, the TB condition was adjusted to a condition that could be interpreted as a passive deception condition or a condition under which the communicator was displaying unclear intentions. Under the adapted TB condition, when the communicator reentered the room and the hider switched the location of the reward in front of her (all the hiders and communicators were female), she cued the wrong, empty container anyway. The fact that the communicator gave a wrong cue despite having witnessed the displacement of the sticker under the TB condition could have been interpreted by the children as the communicator trying to deliberately deceive them. The adapted TB condition could therefore be described as tapping into children's ability to identify whether a confederate is acting dishonest towards them. The TB conditon was passive in its nature, as the children had to react to a potentially deceiving cue. The passive implementation of the TB condition is in opposition to experimental conditions in which the children had to deceive an opponend actively (Shultz & Cloghesy, 1981). Couillard and Woodward (1999) summarized that when deceiving another person, the child's goal is to make the person believe something that is not true, which ability reflects an inherent understanding of another person's mental states. Active deception conditions as well as passive deception conditions have been implemented experimentally to learn about childrens emerging deceptive abilities and their reaction to dishonest behavior throughout different stages of ToM development (Chandler et al., 1989; Palmquist & Jaswal, 2012; Russell et al., 1991; Shultz & Cloghesy, 1981). Krachun et al. (2009) used the TB condition (i.e., the communicator sees the switching and cues correctly) as a control trial for task demands. The fact that TB conditions have been primarily used as control trials has been criticized as wasting the potential of acquiring further information about ToM ability (Hedger & Fabricius, 2011). Thus far, the systematic implementation of TB trials has only been attempted in verbal tasks by Hedger and Fabricius (2011). They concluded that five-year-olds err consistently in TB versions of prominent verbal FB tasks. They argue that this tendency reveals an intermediate level of understanding the mental states of others, in which children reason by *Perceptual Access Reasoning* (PAR). PAR is situated before the level of ToM development in which children base their predictions of another person's behavior on the persons beliefs, which is called *Belief Reasoning* (BR). In their explanation that children aged five used PAR to solve their verbal TB tasks, they proposed that "children only analyze the current situation to determine whether someone does or does not have perceptual contact with an object in question and consequently whether he will be right or wrong about that object" (Hedger & Fabricius, 2011, p. 5). The PAR argument can also be applied to our nonverbal ToM paradigm. If the children in our sample were using PAR to solve the belief conditions, a clear pattern of correct container choices above chance level under both belief conditions should have been observable, as the children observed that the communicator had or did not have visual access to the displacement of the sticker. Likewise, if the children in our sample were using BR to solve the belief conditions, a clear pattern of correct container choices above chance level under both belief conditions should have been observable. Unfortunately, a distinction between the choice patterns in our hiding paradigm of the children using PAR versus the children using BR would not be possible. Nevertheless, other outcome patterns than correct container choices above chance level under both belief conditions would challenge the hypothesized theory by Hedger and Fabricius (2011). By implementing the TB condition in a novel way in our nonverbal ToM paradigm, we aimed to derive additional information from the task about children's ability to identify a potentially deceiving cue towards them and their proneness to trust our communicator as well as discuss our findings within previously proposed concepts (i.e, PAR and BR).

The current study also serves another purpose. Non-human animals have been subjects of ToM research for some time (Call & Tomasello, 1999; Krachun et al., 2009; Krupenye & Call, 2019; Lonardo et al., 2021; Premack & Woodruff, 1978). A recent study at the VetMed University Vienna implemented an analogous nonverbal ToM paradigm in a canine dog sample. Their results indicate that dogs at least hold an implicit understanding of FBs, but some of the behavioral patterns obtained were unanticipated (Lonardo et al., 2021). With the current study, we wish to provide further insight into the obtained results by investigating how preschoolers that are likely to have a developed ToM behave in an analogous nonverbal ToM hiding paradigm.

In line with previous results (Call & Tomasello, 1999; Krachun et al., 2009) and the finding that children aged five years old perform in a goal-directed manner under a passive deception condition by demonstrating an understanding of deceptive behavior to win rewards (Shultz & Cloghesy, 1981), we assumed that the children would always select the not-cued container holding the reward, regardless of the belief condition. Based on the assumption that ToM ability would be the sole driver of preschoolers' container choices in this nonverbal ToM hiding paradigm, we expected significantly more correct than incorrect container choices under both belief conditions in our sample. To our knowledge, no such TB condition has previously been implemented in a nonverbal ToM paradigm in a preschool sample. Due to the ambigous nature of the task, we had no definite predicitions or strong preconceptions about the children's performance in this task. Other behavioral outcome patterns than significantly correct choices under both belief conditions were also imaginable, due to other possible influences on children's choices in this setting. For example, the cue we used in this paradigm (see Chapter 2) could have influenced the children to select the cued but empty container more often than the not-cued container holding the sticker. A highly salient cue could have led the children to choose the cued but empty container significantly more often under both belief conditions, which would have been in line with other experimental data suggesting that the cue used in such a paradigm can influence children's choices (Couillard & Woodward, 1999; Palmquist & Jaswal, 2012; Palmquist et al., 2012). Another possible outcome was for the children to significantly differentiate between the two conditions and show different choice patterns under each belief condition: The children could have ignored the cue under the FB condition more often, due to their understanding of the communcator's false belief, and chosen randomly under the TB condition, due to the ambiguous nature of the TB condition, as the behavior of the communicator under the TB condition could have been interpreted as deceptive or ambigous. Findings indicating that young children are trusting in general (Jaswal et al., 2010; Palmquist & Jaswal, 2012) as well as children being biased to trust an agent in ambiguous situations (Mascaro et al., 2017) raised the possibility of the children following our communicator's misleading cue under the TB condition more often than under the FB condition. Vice versa, the children could have chosen randomly under the FB condition, but disregarded the cue more often under the TB condition. Lonardo et al. (2021) observed that dogs chose the empty but cued container more often under the FB condition than under the TB condition in their analogous implementation of the nonverbal ToM hiding paradigm. Their interpretation of the dogs following the cue under the FB condition more often than under the TB condition was that the dogs could have interpreted the misleading cue of the previously trustworthy communicator under the FB condition as a "mistake in good faith", while the misleading cue by the previously trustworthy communicator under the TB condition was interpreted as deceiteful and disregarded more often. The children in our sample could have interpreted the cue under the FB condition similarly as the previously trustworthy communicator making an honest mistake and followed the cue more often than under the TB condition, in which they could have interpreted the misleading cue as deceitful and unjustified when taking the communicator's belief into account. The last hypothesized outcome of more cued container choices under the FB condition than under the TB condition seemed least likely based on children's goal-directed performance and ability to chose correctly under FB conditions throughout the literature (Call & Tomasello, 1999; Krachun et al., 2009; Wimmer & Perner, 1983) and their proneness to trust in ambigous situations (Mascaro et al., 2017; Palmquist & Jaswal, 2012; Palmquist et al., 2012), which could have made the TB condition harder to solve and therefore resulted in incorrect container choices.

2. Methods

2.1 Participants

The recruited sample consisted of 38 preschoolers (32 five-year-olds and six six-year-olds, 18 girls) who participated in the nonverbal hiding paradigm. The sample size was calculated based on a power analysis of a previous study (Lonardo et al., 2021). Five children were excluded due to either failing the familiarization phases (n = 3), failing both FB control tasks (n = 1), or because they noticed the manipulation in the experiment due to a mistake by a confederate (n = 1). Therefore, the analysis was run on a final sample of 33 preschoolers (28 five-year-olds and five six-year-olds, age M = 5.55, SD = 0.5, 17 girls). The children were recruited via e-mail contact with their parents, who were registered in the scientific recruiting data base of the Wiener Kinderstudien (the Scientific Department for Child Studies of the Faculty of Psychology at the University of Vienna). The participating families exclusively lived in Vienna and the surrounding suburbs. Further participants were recruited via online flyers and postings on Facebook, Instagram, and Twitter by official accounts of the Wiener Kinderstudien. The Participants received incentives in the form of collected rewards (stickers), a small toy, a certificate, and a travel reimbursement of €6.

2.2 Design

We implemented a nonverbal ToM hiding paradigm in which children could find a sticker in one of two identical containers, after a hider placed them there. Throughout the experiment, children learned to rely on the cue of a communicator regarding the location of the sticker. The design included two first order belief conditions (a FB and TB condition in a randomized and counterbalanced order), in a within-subject design. Under these belief conditions, the communicator cued the wrong container, which was not holding the reward. The dependent variable was the children's first container choice (i.e., the cued versus not-cued container) under either belief condition. The container choice was operationalized as the first container the children touched (and subsequently looked inside). The independent variable was the belief state the communicator held, either informed (TB) or uninformed (FB) about the true location of the sticker, when cueing the wrong container in the belief trials. We randomized the first location of the sticker (i.e., the red or grey container), the familiarization trial sequence (i.e., whether Trial A or B was first per phase), and the first belief condition (i.e., the FB or TB), leading to eight different randomizations.

2.3 Procedure

At the laboratory on the testing day, each child's parent presented a negative COVID-19 test or performed a self-test and signed a written informed consent form confirming their child's health as well as a consent form for videotaping the experiment. Afterwards, both the parent and child were asked to enter the room in which the experiment took place. The parent was asked to sit in a chair in one corner of the room. The child was asked to sit on a pillow directly in front of their parent's chair, facing the room and thus away from the parent. In the middle of the room, a red and a grey container were placed, which the child could see clearly. Figure 1 illustrates the room layout.

The experiment consisted of two identical experimentation blocks (Experimentation Block 1 and Experimentation Block 2). Each experimentation block started with three familiarization phases. Theses phases ensured that the child became acquainted with the experimental setup successively and understood what was asked of them to master the belief trials accurately. After the familiarization, one belief trial

Illustration of the Experimentation Room

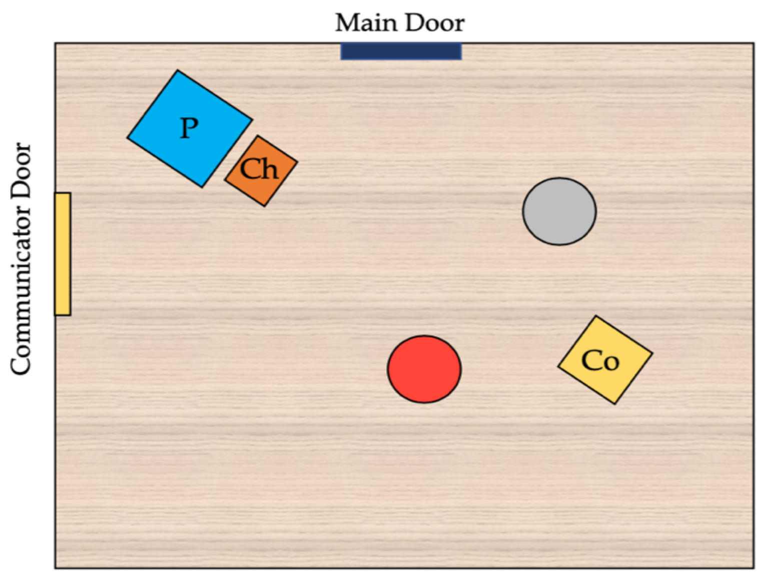

Figure 1. P = Parent, Ch = Child, Co = Communicator, and red and grey circles = red and grey containers.

followed, which concluded Experimentation Block 1. After an optional break, Experimentation Block 2 with identical familiarization trials followed, but a new communicator gave the cues. The change from Communicator 1 to Communicator 2 was necessary to ensure that the child did not hold prejudice due to the communicator's actions in the first belief trial. After mastering the familiarization in Experimentation Block 2, the counterbalanced belief trial was conducted and concluded the experiment. Figure 2 provides an illustration of the overall timeline and test elements.

Each familiarization phase consisted of two trials, Trial A and Trial B. Each participant had to select the correct container holding the sticker in two consecutive familiarization trials per phase; a total of four familiarization trials per phase was allowed. If a child failed to retrieve the sticker in a familiarization trial, the familiarization trial was repeated before moving on. If a child was unable to find the sticker in two consecutive familiarization trials in either familiarization phase, the experiment was aborted, and the child was excluded from the data analysis. Only the main door was used for entering and leaving the experimentation room in the familiarization phases (see Figure 1).

Familiarization Phase 1 (Trial 1A, without displacement and Trial 1B, with displacement) began with the child and parent seated at their assigned places. One of the confederates, the hider, entered the room and hid the sticker in one of the two containers. The same person operated as the hider during all trials for each participant.

Illustration of the Time Series of the Tests Implemented in the Entire Experiment

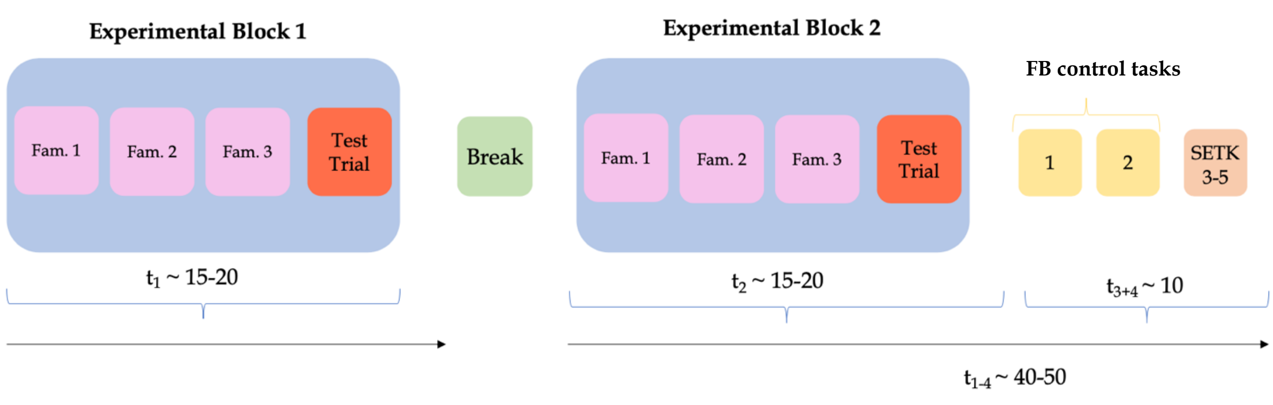

Figure 2. Fam. = familiarization, t_1 = time Experimental Block 1, t_2 = time Experimental Block 2, t_3 = time FB control tasks, and t_4 = time SETK 3-5. All times are in minutes.

She entered the room, placed the sticker in a container, and left (no displacement, Trial 1A) or placed the sticker in a container, waited a moment next to it, removed the sticker, and placed it in the other container (displacement, Trial 1B). The hider performed the hiding of the sticker in an ostensible manner and was watched by the child. After the hiding, the hider gave the keyword "okay" before leaving the room and closing the door behind her. The use of the keyword informed the child that they were allowed to choose a container. If the child successfully chose the correct container holding the sticker in Trial 1A and Trial 1B in consecutive order, Familiarization Phase 2 began.

Familiarization Phase 2 (Trial 2A, without displacement and Trial 2B, with displacement) started with the experimenter entering the experimentation room to ask the child and parent to wait outside for a minute. Then the hider and Communicator 1 entered the room. When the hider finished hiding the sticker in one of the two containers and left the room, the child and parent were asked to take their seats in the experimentation room again. At this point, Communicator 1 was introduced to the paradigm. She remained in the experimentation room at her own spot, which was a pillow placed between the two containers, facing the child from across the room (see Figure 1). From her spot, Communicator 1 approached the container holding the sticker, knelt before it, lifted it, and gave the standardized cue. The cue consisted of saying "Ca c'est bien, c'est très bien" (French for "This is good, this is very good"), while alternately gazing at the container and the child three times in a standardized fashion.[1] Communicator 1 then withdrew to her pillow to give the keyword "okay." Due to the keyword, the child knew they were allowed to approach the containers and make a choice. If the child succeeded in choosing the correct container holding the sticker in Trial 2A and Trial 2B in consecutive order, Familiarization Phase 3 began.

Familiarization Phase 3 (Trial 3A, without displacement and Trial 3B, with displacement) started with the child and parent seated at their assigned spots when the hider and Communicator 1 entered the room. Communicator 1 took her place at

---

[1] Lonardo et al., (2021) used the same cue when implementing the paradigm in a canine dog sample. The cue was pertained to make our study as comparable as possible to theirs. In our case, the cue functioned as a nonverbal cue as the children did not understand what the communicator said.

her pillow, and the hider hid the sticker in one of the containers. The child watched the hiding, and Communicator 1 also obviously closely observed every move the hider made. After the hiding, the hider left the room and Communicator 1 cued the container holding the sticker using the standardized cue. After cueing, Communicator 1 returned to her spot and gave the keyword "okay." The child subsequently chose a container. If the child succeeded in choosing the correct container in Trial 3A and Trial 3B in consecutive order, the first belief trial began.

The FB condition started with the child and parent sitting at their assigned spots when Communicator 1 and the hider entered the room. Communicator 1 took her place at her pillow, and the hider hid the sticker in one of the containers. The hider stayed next to the container while Communicator 1, who closely observed the hiding, left the room through the communicator door (see Figure 1). While Communicator 1 was outside, the hider retrieved the sticker from the container and displaced it into the other container. The hider then left the experimentation room through the main door. The child witnessed the complete hiding and displacement. Communicator 1 returned to her position in the experimentation room. She next gave the child the standardized cue. For the first time, she cued the wrong container, as she was outside when the displacement occurred. She unknowingly suggested the wrong container and returned to her spot to give the child the keyword "okay." The child was then allowed to choose a container.

The TB condition started with the child and parent sitting at their assigned spots when Communicator 1 and the hider entered the room. Communicator 1 took her place at her pillow, and the hider hid the sticker in one of the containers. The hider stayed next to the container while Communicator 1, who closely observed the hiding, left the room through the communicator door. While Communicator 1 was outside, nothing happened in the experimentation room (in contrast to the FB condition). Communicator 1 then returned inside to her position. The hider next retrieved the sticker and displaced it into the other container. The communicator and the child witnessed the displacement. The hider left the experimentation room through the main door, after which the communicator gave her standardized cue. Despite witnessing the displacement, she cued the container that the sticker was placed in before the displacement. The communicator was holding a TB, as she consciously witnessed the displacement and knowingly cued the wrong container. After this, she gave the keyword “okay,” and the child was allowed to choose.

Under both belief conditions, the hider secretly took the sticker with her after she displaced it, so no container held a reward. Therefore, choosing the not-cued container was not rewarded in the first belief trial. This maneuver ensured the child was not biased to choose the same container in the second belief trial, simply because of finding a reward there during the first belief trial.

After Experimentation Block 1, we asked the child and parent whether they would like to take a short break or continue to Experimentation Block 2 immediately. When beginning Experimentation Block 2, the child was told that Communicator 1 had finished playing and that Communicator 2 would play in the forthcoming round. The change of communicator between the experimentation blocks was crucial in ensuring that the child did not hold prejudice due to the ambiguous actions of Communicator 1 in the first belief trial, when she cued the wrong container. The child was told that the game and rules would remain the same. Experimentation Block 2 was structured identically to Experimentation Block 1, containing all three familiarization phases with the same requirements to pass them (see Figure 2). This repetition allowed the child to reacclimatize to the paradigm after the first belief trial and to ensure the child stayed engaged. In Experimentation Block 2, the belief trial was counterbalanced.

During all the trials, the experimenter was seated in the room next door and was able to watch the children’s real-time choices through videocasting to the experimentation room. The children’s choices were written down systematically during the familiarization phases as well as in the belief trials.

After the experiment, the parent was asked to complete the very short form of the Child Behavior Questionnaire (CBQ-VSF; Putnam & Rothbart, 2006), which asked about their child’s temperament (i.e., sociability and effortful control). The CBQ-VSF was included because the temperament trait *effortful control* could have been an influencing factor on the children’s choices during the experiment. Effortful control is the ability that enables children to suppress dominant responses and activate subdominant responses and to detect errors (Posner & Rothbart, 1998, 2000; Rothbart et al., 2003). Effortful control is one of three broad temperamental factors that the CBQ-VSF measures, which is comprised of high positive loadings for inhibitory control, attentional control and perceptual sensitivity scales (Putnam & Rothbart, 2006) and served as our measurement of inhibitory control of executive function.

While the parent completed the questionnaire, the child completed two *verbal FB control tasks* (Wellman & Liu, 2004) in their German translation (Kristen et al., 2006) with the experimenter. These tasks were included to ensure the participants had attained the presupposed level of social-cognitive development. Children who failed both verbal FB control tasks were excluded from the study. The first task was an explicit FB task that concerned the location of an item. The experimenter showed the child a small figure named Paul and a sheet with a picture of a backpack and a cupboard on it. The child judged where Paul would look for his gloves considering his FB about the location. The test question was, "Paul thinks his gloves are in his backpack, but they are in the cupboard. Where do you think Paul will look for the gloves?"). The child scored one point if they correctly answered "backpack."

The second task was a contents FB task concerning the contents of an item. The experimenter presented a Smarties box to the child and asked what they thought the box's contents were. After the child answered "Smarties," the experimenter showed the child that a small toy pig was in the box. The child was then asked to judge Lucas' (another toy figure) FB about the contents of the box. Importantly, the child was told Lucas had never looked inside the box. The test question was, "Lucas has never looked inside the box. What does he think is in the box?" The child scored one point if they correctly answered "Smarties."

Lastly, the experimenter assessed the child's speech comprehension using the SETK 3-5 (Grimm, 2015). A manipulation task was implemented, during which the child had to perform small assignments with tools that were presented, such as in the direction "Before you hand me the small box, put all the red buttons inside." The grammatic complexity of the instructions increased.

Before leaving, the parent received a reimbursement of €6 and the child received their collected stickers, a small toy, and a certificate for participating.

2.4 Instructions

Before the first familiarization phase began, the experimenter gave written instructions to the parent and instructed the child orally. The written instructions told the parent to interfere as little as possible and not to answer any questions the child might ask during the trials with biasing information. However, standardized allowed answers were given, such as "I am sure you are doing well" or "This sounds tricky indeed." The experimenter invited the child to play a game in which they could choose one of two containers visibly placed in front of them. We explained that the parent would sit behind them wearing an eye-covering mask. The parent could therefore not answer questions due to them not seeing what occurred. The experimenter instructed the child that in each trial they could choose a container by standing up, approaching the container, and looking inside (the containers were covered with a thick paper lid that matched the color of the container). The child was told that it could sometimes find something in a chosen container. If they found something, they were allowed to retrieve it and take it back to a box that the parent was holding. When the child was unaccompanied in the experimentation room (n = 3) this box was placed next to the child's seating pillow.

The experimenter then introduced the three confederates who would be playing with the child (two per experimentation block). She explained that the confederates would enter the room to do something with the containers in each trial. The child was told that the confederates did not speak German (to prevent them from asking questions) and instructed to watch what they did closely.

The child was instructed to choose a container when they heard the keyword "okay." This measure ensured that the child did not stand up before the hider and communicator were finished. Lastly, the child and parent were told that the experimenter would enter the room before some of the trials to give specific instructions. For example, Familiarization Phase 2 included the direction "Please both stand up and come outside with me. The next trial will start with you waiting outside."

2.5 Analysis

To test our hypothesis of significantly more correct than incorrect container choices under the TB and under the FB condition, we examined the proportions of correct container choices made under each condition by exact binomial tests. We interpreted the container choice patterns for each belief condition separately to identify whether subtle differences between the belief conditions existed.

For both conditions, we calculated a Wilcoxon signed-rank test for paired samples to investigate whether the median weight of cued container choices was significantly smaller under the FB condition than under the TB condition. This analysis was conducted exploratorily, after the initial analysis of the children's choice proportions under each belief condition with exact binomial tests revealed subtle differences in the correct container choices between the belief conditions.

We also analyzed the children's choice patterns in both belief trials regardless of the belief condition order. The children could have exhibited four different choice patterns across the two belief trials, which analysis provided insight into whether children systematically used choice strategies throughout our paradigm that seemed to not take the communicator's belief into account. The children could have either chosen the not-cued or cued container in both trials or changed their choice from the cued or not-cued container during the first belief trial to the alternate container during the second belief trial. If the children significantly alternated their choice from the not-cued or cued container in the first belief trial to the alternate container in the second belief trial above chance level, the possibility of children choosing systematically without taking the belief of the communicator into account should be discussed as a leading choice strategy in this paradigm.

As we assumed that the belief condition would not significantly influence the children's choices, we modeled two binomial generalized linear mixed models (GLMMs) to examine the explanatory potential of the conditions and other possible influencing variables, such as age, sex, speech comprehension, and effortful control.

In an exploratory analysis, we used a subset consisting of 24 subjects (age M = 5.54, SD = 0.58, 13 girls) to investigate the effect of speech comprehension on the children's choices. Not all the subjects in the study were tested with the SETK 3-5 (Grimm, 2015), as this test was incorporated into the testing battery after piloting and initial tests.

Lastly, we calculated correlations to understand the relations between crucial variables in our nonverbal ToM hiding paradigm. We calculated a correlation between the correct answers during the FB control tasks and the age of the participants to investigate whether we could obtain similar results to previous studies exploring children's performance in verbal FB tasks (Call & Tomasello, 1999; Kristen et al., 2006; Wellman & Liu, 2004). We computed a correlation between the children's correct answers on the FB control tasks and the cued container choices the children made in our nonverbal ToM paradigm to see whether performance on the verbal FB control tasks and performance in the novel, nonverbal ToM paradigm was linked. We calculated a correlation between the CBQ-36 scores parents assigned to their children and the container choices the children made in the nonverbal ToM paradigm to investigate whether inhibitory control abilities and container choices were related to each other. Lastly, we computed a correlation between the total number of familiarization trials a child underwent before the belief trials began and the container choices the children made to investigate whether making more mistakes in the familiarization phases (which is reflected in the total number of familiarization trials per child) was connected to the subsequent container choices the children made.

3. Results

Under both the TB and FB condition, more children chose the container holding the sticker than the empty container the communicator suggested. The percentage of cued but empty container choices was 39%. This percentage resulted from 33% of children who chose the cued but empty container under the FB condition and 45% of children who chose the cued but empty container under the TB condition.

Proportions of the Children's Container Choices Under the Belief Conditions

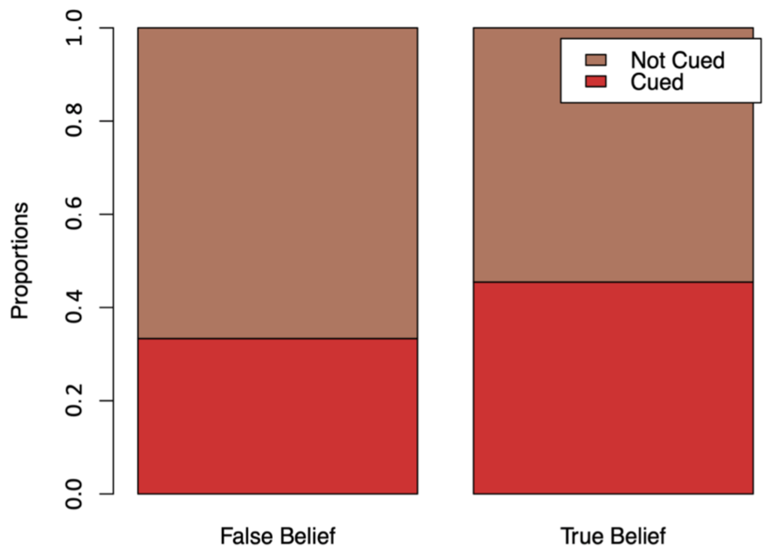

Figure 3. Illustration of the proportions of the container choices children made under the TB and the FB condition.

3.1 Binomial Tests

We analyzed the choice proportions of each condition with exact binomial tests. Under the FB condition, the children chose the cued but empty container significantly less often than expected by chance (p = .04, 95% CI [0.0, 0.49], one-sided). Under the TB condition, the children did not choose the cued but empty container less often than expected by chance (p = .36, 95% CI [0.0, 0.61], one-sided), showing a random choice pattern. Figure 3 plots the proportions of both conditions.

3.2 Wilcoxon Signed-Rank Test for Paired Samples

For both conditions, we exploratorily calculated a Wilcoxon signed-rank test for paired samples. The H_1 can be expressed as a median weight location shift of the cued container choices of less than 0 from the FB condition to the TB condition (one-sided). The test statistic revealed no such difference (V = 26, p = .13), indicating that the median weight of cued container choices under the FB condition was not significantly smaller than under to the TB condition.

3.3 Analysis of the Children's Choice Patterns

The children could have exhibited four different choice patterns across the two belief trials, regardless of the belief trial order. Figure 4 depicts each choice pattern as a colored block on the x-axis and the number of children that chose according to each pattern on the y-axis. Underneath each block on the x-axis the age mean of each resulting group of children is given. The group of children that selected the not-cued container holding the reward in both belief trials was mostly comprised of the older children in our sample, with an age mean of 5.73 years, whereas the group of children

Children's Choice Patterns in Both Belief Trials Regardless of the Belief Condition

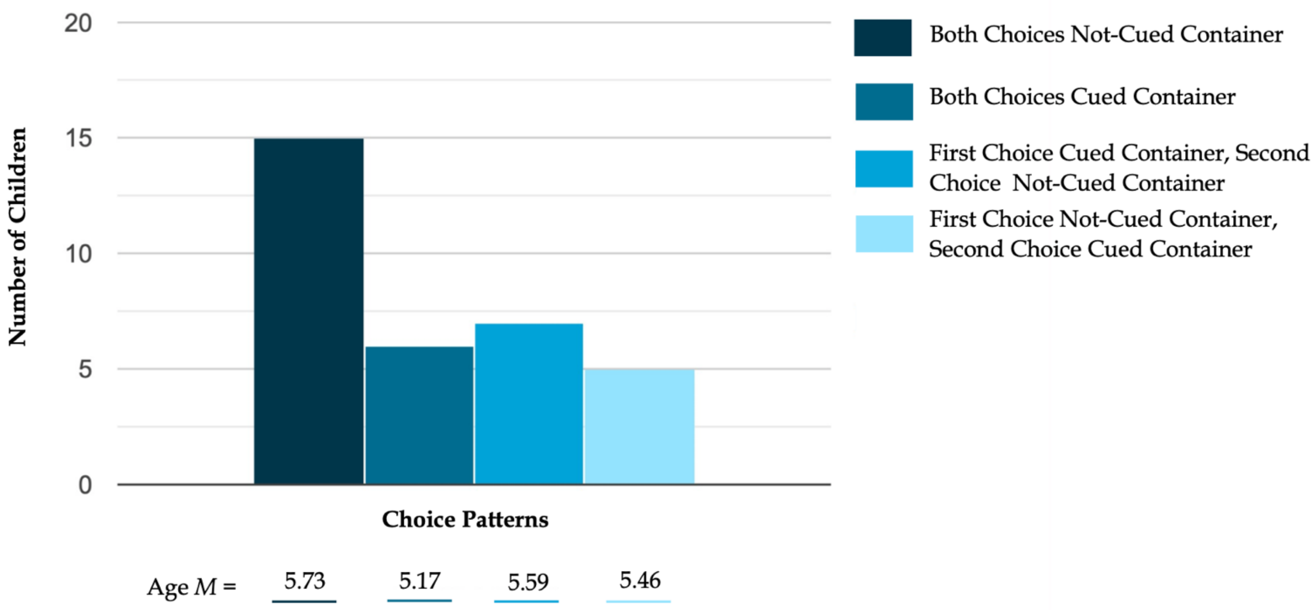

Figure 4. On the x-axis, the different choice patterns are depicted as colored blocks representing the groups of children who exhibited each choice pattern. In the legend on the right, each choice pattern is assigned to the corresponding block in the figure by a different color. Beneath each block on the x-axis and underlined in the corresponding colors of each choice pattern block are the age means of the groups. On the y-axis, the number of children who exhibited each choice pattern is depicted.

that chose the cued container in both belief trials was mostly comprised of the younger children in our sample, with an age mean of 5.17 years. In total, 21 children chose the same container in both belief trials (“Both Choices Not-Cued Container” and “Both Choices Cued Container” in Figure 4) and 12 children alternated their container choice from the first to the second belief trial (“First Choice Cued, Second Choice Not-Cued Container” and “First Choice Not-Cued, Second Choice Cued Container” in figure 4). Neither of the two options (i.e., choosing the same container twice versus choosing alternate containers) was used by children more frequently than expected by chance, as revealed by exact binomial tests (all $p > .08$).

3.4 Generalized Linear Mixed Models

3.4.1 Fixed and Random Effects

We implemented a binomial GLMM to predict the choices the children made using condition as a fixed factor and subject as a random factor. Including age as a fixed factor produced a better model fit as indicated by an analysis of variance (ANOVA; $x^2 = 4.45$, $df = 1$, $p = .04$), although the predictor itself did not reach statistical significance ($z = -1.85$, $p = .07$). During the modelling process, the predictor sex and the scores of the FB control tasks showed no significant influence on choice, as implicated by successive ANOVA testing. Therefore, these predictors were not included in the model. The resulting first GLMM included the predictors condition and age as fixed effects and subject as a random effect. An ANOVA revealed a non-significant result ($x^2 = 5.70$, $df = 2$, $p = .06$) of the first GLMM in comparison to the null model.

We used age as the only fixed effect in the second binomial GLMM. We again included subject as a random factor. Likewise, predictors sex and scores of the FB control tasks showed no significant influence on choice and were not included. The second GLMM showed a significantly better model fit compared to the null model, as estimated by an ANOVA ($x^2 = 4.43$, $df = 1$, $p = .04$). Therefore, we observed a main effect of age, indicating that older children in our sample (six-year-olds) were more likely to choose the not-cued container holding the reward (-1.41 on the log odds scale). The resulting odds ratio for older children choosing the cued container was 0.32. To approximate the explained variance, including both the fixed and random effects of the second binomial GLMM, we calculated a conditional $R^2 = 0.259$.

3.4.2 Interactions and Random Slopes

In the second binomial GLMM, we tested for an interaction between age and condition, but an ANOVA revealed no significant effect (x^2 = 3.81, *df* = 2, *p* = .15).

In the second binomial GLMM, we tested for the inclusion of random slopes for age, but an ANOVA revealed no significant effect (x^2 = 0.29, *df* = 2, *p* = .86).

3.5 Exploratory Analysis of Speech Comprehension

We modeled a binomial GLMM to predict the choices children made, including the fixed effects of age and speech comprehension and the random effect of subject. Speech comprehension did not significantly influence the children's choices as a predictor, and the GLMM that included speech comprehension and age as fixed effects did not indicate a better model fit than the null model, as revealed by an ANOVA against the null model (x^2 = 4.42, df = 2, *p* = .11).

3.6 Correlations

There was a significant negative correlation between correct answers during the FB control tasks and the age of the participants, *r* (31) = -.36, *p* = .04. There was no correlation between correct answers on the FB control tasks and cued container choices, *r* (31) = .24, *p* = .17. The CBQ-36 scores, which were calculated on a subset of children whose parents answered the questions (*n* = 31), did not correlate with the choices the children made, r (29) = -.02, *p* = .92. The number of familiarization trials a child underwent before the belief trials began did not correlate to the choices the children made, *r* (31) = .31, *p* = .07.

4. Discussion

In this discussion, we first summarize our most important findings and interpretations for the TB and the FB condition and place our results within previous literature. We give suggestions on how to synthesize the diverging interpretations of the data we acquired for the TB and the FB condition. Second, we discuss the cue our communicators used in the nonverbal ToM hiding paradigm, the possible link between children's inhibitory control abilities and their container choices and the task demands children faced in our paradigm. Third, we examine the correlation between the children's success in the FB control tasks and the children's age as well as the correlation between the children's success in the FB control tasks and the children's success in our nonverbal ToM hiding paradigm. We discuss how our results regarding the FB control tasks are in opposition to previous results as well as the problems we faced when implementing the FB control tasks. Fourth, we discuss our measurement of the children's speech comprehension, the children's speech comprehension's influence on the nonverbal ToM hiding paradigm and a possible adaptation to lower the language demands in our paradigm even further. Lastly, we make summarizing and concluding remarks about the current implementation of the nonverbal ToM hiding paradigm in a preschool sample.

This study examined whether children would follow the misleading cue of a confederate regarding the location of a sticker under a TB and a FB condition. We hypothesized that the children would be able to disregard the cue and select the container holding the sticker significantly more often than the cued bucket under either condition. This hypothesis was not supported by the current experiment. For both conditions, numerically but not significantly more children chose the not-cued container holding the reward, indicating a random choice pattern overall. In line with our assumption, a binomial GLMM revealed that the belief the communicator held was not a significant predictor of the choices the children made. The GLMM revealed a main effect of age, indicating that older children in our sample (six-year-olds versus five-year-olds) were significantly more likely to disregard the communicator's cue under either condition. The main effect of age was also evident in the different choice strategies the children applied across both trials, regardless of the condition. The largest age difference across the four different choice strategies was between the group that chose the baited, but not cued container twice (age M = 5.73 years) and the group that chose the cued container twice (age M = 5.17 years).

To eliminate the possibility that the children applied a systematic choice strategy regardless of the belief the communicator held throughout the trials (i.e., choosing a container during the first belief trial and always choosing the opposite in the second belief trial), we statistically analyzed the accumulation of different choice patterns in our sample. As the children did not significantly alternate their choice from the not-cued or cued container in the first belief trial to the alternate container in the second belief trial above chance level, we ruled out the possibility of this strategy predominantly leading the children's choices in our paradigm. If a systematic choice strategy that seemed to not take the belief of the communicator into account would have dominated the children's choices in our nonverbal ToM paradigm, we should raise concerns about the paradigms ability to investigate children's understanding of mental states.

When examining each belief condition separately, subtle differences in the children's preferred container choices became evident. The interpretation of the choice proportions for the FB and the TB condition in the following paragraphs is preliminary, as examining the children's choices under each condition separately does not account for the dependency of the children's choices between both belief conditions. Nevertheless, the analysis of the container choice proportions under each belief condition disjointedly enables a more detailed reflection on the results that aligns with previous studies (none of which contain both comparable belief conditions) and allows to propose explanations for the children's behavior under each belief condition.

Under the FB condition, the children were able to retrieve the sticker and disregard the communicator's cue significantly more frequently than expected by chance. Based on their correct choices, the children in our sample who had developed corresponding ToM abilities demonstrated an understanding of the FB the communicator held. In this regard, the main effect of age accords with previous findings (Call & Tomasello, 1999; Krachun et al., 2009; Kristen et al., 2006; Wellman & Liu, 2004), which established that age is a leading explanatory variable in the development of FB understanding. The obtained proportion of children who disregarded the communicator's false cue and chose the not-cued container under the

FB condition (66% correct choices) is comparable to previous findings (Krachun et al., 2009) for this age group.

The TB condition featured new modifications that, to our knowledge, have not been implemented in a nonverbal ToM paradigm before. The TB condition did not only function as a control condition (Krachun et al., 2009) but can also be interpreted as tapping into an understanding of deceptive behavior. As the communicator knew the correct location of the reward but still cued the wrong container, the children could interpret the situation as the communicator trying to deceive them. Under the TB condition, the children in our sample were unable to select the correct container and retrieve the sticker more frequently than expected by chance. They demonstrated a random choice pattern under this condition (55% correct choices). More than one interpretation of these results is possible. First, it has been proposed that children aged five years old have trouble in deceiving others but perform better under a passive condition when they are confronted with deceptive behavior and act in a goal-directed manner, namely to win rewards (Shultz & Cloghesy, 1981). Our TB condition was passive as the children only had to watch the communicator's actions and were allowed to choose the container they perceived as holding the reward. The main effect of age in our data, therefore, agrees with Shultz and Cloghesy's (1981) findings, indicating that the older the child, the more successful they are under such a condition. Second, another interpretation relates to children's disposition to trust in ambiguous situations. In a previous study, children generally trusted an agent's assertions to be reliable, although they knew the agent was being "mean" and did not want them to find a reward (Mascaro et al., 2017). Furthermore, other studies have proposed that children are trusting in general (Jaswal et al., 2010; Palmquist & Jaswal, 2012). The children in our paradigm could have similarly interpreted the communicator's cue as reliable, even though she was informed about the displacement, reflecting children's proneness to trust. Such an interpretation is not probable under the FB condition, when the communicator was indeed unknowing. This evaluation demonstrates how the new implementation of the TB condition can access different fields of ToM development. Mascaro et al. (2017) concluded that sensitivity to deception increases during the fifth year of life, which assertion accords with our detected random choice pattern and main effect of age. The fewer correct choices in our sample under the TB condition compared to the FB condition could reflect the more difficult nature of the TB task due to the deceptive intentions of the communicator and children's bias to trust. Third, the children may not have interpreted the intentions of the communicator under the TB condition as deceptive at all. As she held a TB, the children could have interpreted her intentional wrong cue as her showing something new. This interpretation could also account for the proportionally more cued choices under this condition. Nevertheless, this interpretation is challenged by the fact that the children saw everything that was placed in the containers, and the appearance of a new item was not possible. Moreover, during the familiarization phases, the children often checked both containers in the early trials, which could have hinted to them that the containers did not contain anything of interest besides the sticker. To conclude the interpretation of the TB condition, we want to discuss our results with regards to Hedger and Fabricius (2011), who argued children aged five years old used PAR to solve verbal TB conditions. The five-year-olds in our sample using PAR should have chosen the not-cued container in both belief trials, as they observed that the communicator had or did not have visual access to the displacement of the sticker. As this was not the case, specifically not for the age group Hedger and Fabricius (2011) argued used PAR, it appears that more variables influence children's choices during ambiguous ToM paradigms and that PAR does not account for the variability sufficiently. Additionally, we did not find definite evidence for children in our sample using strict BR to solve the task, which would have also resulted in significantly more correct than incorrect container choices under both belief conditions. Therefore, none of the two theoretical reasoning stages children use to solve TB conditions according to Hedger and Fabricius (2011) could be applied to our data. In the sparse literature about the active implementation of TB trials, Hedger and Fabricius (2011) nevertheless provide important insights into the overarching problems children might experience during TB tasks, although the explanation for these problems is not yet agreed upon. As TB trials have been secondary to the overwhelming focus on FB understanding, the diverging explanations in a yet rather unclear pattern of results is unsurprising. Due to their explanatory potential in the investigation of ToM development, we recommend that the systematic implementation of active TB trials receives careful attention in subsequent verbal and nonverbal paradigms.

To resolve the contrary interpretations of the observed data under the TB condition, a possible adaptation could provide more insight into why the children disregarded or followed the cue under certain conditions. We suggest implementing a short, standardized interview after the nonverbal ToM paradigm, during which the experimenter would ask about the children's thoughts during the belief trials and their subsequent motivation to choose. Additionally, having the children rate certain aspects of the communicators would be revelatory. Possible statements regarding the children's perspective on the trustworthiness (e.g., "She wanted to show me something new" or "She wanted to deceive me"), sympathy (e.g., "I liked her" or "I did not like her"), or intention (e.g., "good" or "bad") of the communicator could be contrasted with the choices the children made when being confronted with her. The ratings of the communicators under the FB and the TB conditions could then be directly compared. This study did not implement such an interview in a systematic way from the beginning of the data collection as the useful information expected to be gained by it was rated as low. However, as we engaged in conversation about the experiment with some of the children after the testing, they spontaneously commented and asked about the hiding paradigm. On the contrary, other children did not state anything about what they thought during the experiment when asked about it in conversation in an unstandardized, chatty way. Therefore, we expect that useful information could possibly be gained by a standardized interview consisting of clear and short questions, especially from older children.

The results of the belief conditions suggest that the paradigm is suitable to investigate ToM development in nonverbal individuals, as previously demonstrated by Lonardo et al. (2021) and the similar implementations by Krachun et al. (2009) and Call and Tomasello (1999), which yielded comparable results. The correct container choices under the FB as well as under the TB condition reflect the children's emerging ability to take the belief of the communicator into account when choosing a container in our paradigm. Likewise, the incorrect container choices indicate that some children aged five to six years old struggle with attributing mental states in our paradigm or used decision strategies that were not based on the belief the communicator had. Under the TB condition, it could have been harder for the children to choose correctly due to their still developing sensitivity to deception in that age span and their disposition to trust in ambiguous situations. We found a main effect of age, indicating that the older the children are, the more successful they are in solving both belief conditions. This main effect of age comes as no surprise considering the magnitude of findings suggesting that the children's ability to understand other's beliefs as well as the children's sensitivity to deceptive behavior develops fundamentally in preschoolers (Call & Tomasello, 2008; Flavell, 1999; Kristen et al., 2006; Mascaro et al., 2017; Shultz & Cloghesy, 1981; Wellman & Liu, 2004; Wellman et al., 2001). Thus, it could be argued that the children in our sample were in a sensitive age window regarding the development of the ToM skills needed to master both belief trials successfully, which is reflected in the obtained proportions of the cued and the not-cued container choices under the belief conditions and the found main effect of age.

Regarding the cue the communicators used in our nonverbal ToM paradigm, the cue could be interpreted as highly salient. We used ostensive multi-modal cueing in the same way that Lonardo et al. (2021) did in their analogous implementation in a canine dog sample. Our multi-modal cue included more than one communicative signal, which other cues in similar experiments were comprised of. Examples are the cue of Call and Tomasello (1999), who used the placing of a wooden block upon the container as a cue and the cue of Krachun et al. (2009), who used the communicator's reaching for the container as a cue. In opposition to Call and Tomasello (1999) and Krachun et al. (2009), our cue did not only consist of marking the container or reaching for the container, but of approaching the container, lifting the container up and giving a verbal, unintelligible signal while repeatedly and alternately gazing at the container and the child in a standardized fashion. Our cue could be described as a cascade of salient communicative signals or a short cueing routine, which the child watched repeatedly throughout the experiment. It is possible that the saliency of the cue, which was correct during the familiarization trials, was too difficult for the children to disregard in the belief trials. That certain cues indeed influence children's choices in such paradigms has been demonstrated previously, especially regarding the pointing gesture (Couillard & Woodward, 1999; Lee et al., 1998; Palmquist & Jaswal, 2012; Palmquist et al., 2012). Nevertheless, no direct comparison can be made between the pointing gesture and our cue, as the latter did not require the inhibition of a socially learned cue for showing something. Couillard and Woodward (1999) proposed that children's experience with using the pointing gesture for showing objects of interest may be the reason for children's biased responses in experiments when pointing is used as a cue. Novel behaviors, like using a marker or multi-modal cueing, do not seem to have this connotation (Couillard & Woodward, 1999). Previous results suggest the pointing gesture is a strong confound influencing the children's choices (Couillard & Woodward, 1999; Lee et al., 1998; Palmquist & Jaswal, 2012; Palmquist et al., 2012), which does not apply to the current experimental data. If our cue biased the children's choices fundamentally, then a clearer influence towards the cued container under both conditions should have been observable. The possibility of the cue influencing the children's choices still presents a possibility for isolated cases, especially for the younger children in our sample.

Closely related to the saliency of the cue is the possibility of the children's low inhibitory control abilities preventing them from disregarding the communicator's cue (Apperly, 2012; Carlson & Moses, 2001; Carlson et al., 1998; Chasiotis et al., 2006). The CBQ-VSF (Putnam & Rothbart, 2006), which we used to measure the temperament factor effortful control (see Chapter 2.3), did not reveal a correlation between the children's choices in the paradigm and the scores parents assigned their children. This result indicates that the differences in effortful control among the children in our sample did not influence their choices. Hence, we found no evidence that an inability to inhibit the communicator's cue primarily led the children to choose the cued container. In opposition to our findings, Chasiotis et al. (2006) proposed a strong correlation between children's inhibitory control abilities, measured by elaborate test batteries which included multiple facets of inhibitory control, and children's success in ToM tasks. In our case, a more objective measure of inhibitory control would be needed, as the short form of the CBQ-VSF does not include sufficient information to make a definite assertion. Implementing a measurement that captures the inhibitory control abilities of the children in an active paradigm (see Chasiotis et al., 2006) instead of asking the parents to fill in a questionnaire could be informative. Nevertheless, the goal of the current study was to disentangle the measurement of the children's ToM abilities from the children's inhibitory control abilities. Therefore, the lack of a connection between the children's inhibitory control abilities and their container choices can be preliminary interpreted as the paradigm being able to measure belief understanding disentangled from the children's inhibitory control abilities.

Next, we want to discuss the task demands children were confronted with in our nonverbal ToM hiding paradigm. We assumed that due to the successive and stepwise familiarization phases, the children would be able to master the task demands before the belief trials were presented. We argued that the task demands should have been lowered in our nonverbal ToM pardigm in comparison to previous studies, as the children did not have to solely rely on the communicator's cue to infer where the sticker must have been hidden (Call & Tomasello, 1999; Krachun et al., 2009), but were able to watch the hiding process directly. Task demand problems, like the child not being able to find the sticker and choosing the wrong container in the familiarization trials, were reflected in the total number of familiarization trials a child underwent before the belief trials. Another example for a task demand problem would be the child interrupting the hiding procedure by standing up to choose a container before hearing the keyword "okay". Such scenarios were also reflected in the total number of familiarization trials a child underwent before the belief trials, as the familiarization trial was aborted and repeated if such an error occurred. We did not observe a correlation between the total number of familiarization trials a child underwent and the child's subsequent container choices. Accordingly, the problems with the task demands some of the children might have experienced during the familiarization phases did not appear to influence the container choices they made in the belief trials. Almost all the children were able to master the belief trials including all task demands after the stepwise familiarization phases, as only three children were excluded due to their erroneous performance in the familiarization phases.

As a next point of discussion, we want to closer examine the relationship between children's success in the FB control tasks (Kristen et al., 2006) and in our nonverbal ToM paradigm. The children's FB control task scores are comparable to previous findings for these age groups (Kristen et al., 2006; Wellman & Liu, 2004). However, in the present study success in FB control tasks was negatively correlated with the age of the participants, indicating that the older children in our sample tended to fail more FB control tasks. Furthermore, our results differed from those of Kristen et al. (2006), who translated the Wellman and Liu ToM task scale into German. In our sample, conversely to theirs, more children failed the contents FB task, which was the Smarties task (n = 24 successes), than the explicit FB task (n = 29 successes). These results are in opposition to the well-established development of ToM with age and could reflect a problem with the selected FB control tasks. Furthermore, no correlation existed between correct answers in the FB control tasks and correct container choices in our nonverbal ToM paradigm. If both the FB control tasks and our nonverbal ToM paradigm measured ToM ability, a relationship between success in both measurements should be observable. Converse results have been found by Call and Tomasello (1999), who used the Sally-Anne FB task (Baron-Cohen et al., 1986) as a FB control task. In their study, children's performance of the verbal FB control task and their nonverbal FB paradigm was highly correlated, providing evidence that both measured ToM ability. Russell et al. (1991) also showed that children's success in their deception task was positively correlated to correct answers in their used FB control task. Possible reasons why this was not the case in our data are discussed subsequently. A major problem with the Smarties task was that the children often did not know how to respond to the setup question "What do you think is in there?" when the experimenter presented the Smarties box. The reactions of the children who failed the task indicated that they were not familiar with the Smarties sweets. The instructions to assist children if they did not know the contents included "Does it look like there are Smarties in there?" "What box is this? What should be in there?" or "Should there be Smarties in there or books?" All these questions required the children to remember if they had seen such a Smarties box before and what the supposed contents of the box should be. The fact that many children did not recognize the sweet nullified the assistance instructions. Confusion about the item, the suggestive helping questions, and the switching of the term "Smarties" to "sweets" make the validity of the test item in this case questionable. The actual FB test determines whether the children can attribute a FB to the protagonist Lucas after seeing for themselves that in fact a toy pig is inside the Smarties box, which Lucas did not see. If a child could not reliably link the original contents to the box just by looking at it, the test question of whether Lucas, who "has never looked inside the box," knows the original content of the box is unintendedly ambiguous.

The problems with the FB control tasks partially reflect the problems with classic FB tasks discussed in the introduction of this paper. The difficulties the children faced during the task illustrate the memory, inhibitory, and linguistic demands placed upon them, which became evident in our sample. A second reason why we did not observe a correlation between success in our nonverbal ToM paradigm and success in the verbal FB control task could be that we implemented a FB and a TB condition in the same paradigm. How verbal FB understanding, and active TB conditions or deceptive tasks correlate is not as well established as the correlation between verbal FB and nonverbal FB paradigms. Some of the authors we referred to did not implement

FB control tasks in their deceptive paradigms (Palmquist & Jaswal, 2012; Shultz & Cloghesy, 1981). Therefore, and due to the challenges we encountered in measuring verbal FB understanding, no final conclusion about the connection between our nonverbal ToM paradigm and verbal FB understanding can be drawn.

Regarding the exploratory findings related to speech comprehension, the SETK 3-5 scores (Grimm, 2015) did not influence the choices the children made in our paradigm. This finding could reveal the nonverbal nature of the task and indicate that the paradigm was able to access ToM abilities disentangled from speech comprehension. This would distinguish our task from classic verbal FB tasks, which have been criticized for their high language demands and the resulting influence on task outcomes. The result of no correlation between our measurement of speech comprehension and success in our nonverbal ToM paradigm provides essential evidence for the usefulness of the nonverbal hiding paradigm in its application in nonverbal populations. A small limitation that should be mentioned is that the SETK 3-5 may not have been able to sufficiently differentiate between speech comprehension abilities in the older preschoolers in our sample with higher skills. The SETK 3-5 is meant to measure speech comprehension in children aged three to five years old, and some of the individuals in our sample exceeded the suggested age. Therefore, the lack of evident speech comprehension influence on our nonverbal ToM paradigm should be interpreted cautiously.

A modification to the paradigm could lower the language demands even further to examine completely nonverbal individuals. As the task is slowly introduced through the familiarization phases, a red light could be used from the beginning of each trial to signal that the participant should sit, wait, and watch. The keyword "okay," which we used to signal that the children were allowed to choose a container, could be replaced with a green light. During the familiarization trials, due to their repetitive nature, the task goal of finding the sticker would become evident without being explicitly stated. However, an auxiliary demonstration trial during which the experimenter attempts to locate the sticker could be acted out. With these adaptations, verbal instructions could be avoided altogether.

To conclude, a revision of the nonverbal ToM paradigm with more suitable and extensive methods to measure concurring variables is implied by the results of this study. We suggest more comprehensive verbal ToM control measurements that access

FB and TB understanding, a standardized interview to investigate children's interpretation of the task, and more sensitive procedures to measure speech comprehension and inhibitory control. To connect success in our nonverbal ToM hiding paradigm and ToM development, investigating how younger and older children (four- to seven-year-olds) resolve our paradigm would be interesting. These age groups should exhibit clearer failures and successes in the tasks. The children in the current sample seemed to be in a sensitive age window regarding the development of the ToM skills needed to master both belief trials successfully, which is indicated by the obtained proportions of cued and not-cued container choices for each belief condition and the finding that the older children were more likely to choose correctly under both belief conditions.

References

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology, 65*(5), 825–839. https://doi.org/10.1080/17470218.2012.676055

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970. https://doi.org/10.1037/a0016923

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*(2), 113–125. https://doi.org/10.1111/j.2044-835x.1986.tb01003.x

Bartsch, K., & Wellman, H. M. (1995). Children talk about the mind. *Psychology in the Schools, 33*(1), 87–87. https://doi.org/10.1002/1520-6807(199601)33:1<87::AID-PITS2310330105 >3.0.CO;2-C

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science, 20*(1), 85–91. https://doi.org/10.1111/j.1467-9280.2008.02246.x

Byom, L. J., & Mutlu, B. (2013). Theory of mind: Mechanisms, methods, and new directions. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00413

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*(2), 381–395. https:// doi.org/10.1111/1467-8624.00028

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*(3), 672. https://doi.org/10.2307/1132197

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-Scale Deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*(6), 1263. https://doi.org/10.2307/1130919

Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development, 30*(3), 249–260. https://doi.org/10.1177/0165025406066759

Couillard, N. L., & Woodward, A. L. (1999). Children's comprehension of deceptive points. *British Journal of Developmental Psychology, 17*(4), 515–521. https://doi.org/10.1348/026151099165447

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Figueras-Costa, B. (2001). Theory of mind development in deaf children: A nonverbal test of false-belief understanding. *Journal of Deaf Studies and Deaf Education, 6*(2), 92–102. https://doi.org/10.1093/deafed/6.2.92

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*(1), 21–45. https://doi.org/10.1146/annurev.psych.50.1.21

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly, 50*(3), 274–290. https://doi.org/10.1353/mpq.2004.0018

Grimm, H. (2015). SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder. Hogrefe.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology, 2*(3), 429–447. https://doi.org/10.1007/s13164-011-0069-9

James, W. (1889). The psychology of belief. *Mind, 14*(55), 321-352.

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science, 12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von "Theory of Mind"-Aufgaben. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(4), 186–195. https://doi.org/10.1026/0049-8637.38.4.186

Krupenye, C., & Call, J. (2019). Theory of mind in animals: Current and future directions. *WIREs Cognitive Science, 10*(6). https://doi.org/10.1002/wcs.1503

Lee, K., Eskritt, M., Symons, L. A., & Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading." *Developmental Psychology, 34*(3), 525–539. https://doi.org/10.1037/0012-1649.34.3.525

Lonardo, L., Völter, C. J., Lamm, C., & Huber, L. (2021). Dogs follow human misleading suggestions more often when the informant has a false belief. *Proceedings of the Royal Society B: Biological Sciences, 288*(1955), 20210906. https://doi.org/10.1098/rspb.2021.0906

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Mascaro, O., Morin, O., & Sperber, D. (2017). Optimistic expectations about communication explain children's difficulties in hiding, lying, and mistrusting liars. *Journal of Child Language, 44*(5), 1041–1064. https://doi.org/10.1017/S0305000916000350

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Palmquist, C. M., Burns, H. E., & Jaswal, V. K. (2012). Pointing disrupts preschoolers' ability to discriminate between knowledgeable and ignorant informants. *Cognitive Development, 27*(1), 54–63. https://doi.org/10.1016/j.cogdev.2011.07.002

Palmquist, C. M., & Jaswal, V. K. (2012). Preschoolers expect pointers (even ignorant ones) to be knowledgeable. *Psychological Science, 23*(3), 230–231. https://doi.org/10.1177/0956797611427043

Porter, M. A., Coltheart, M., & Langdon, R. (2008). Theory of mind in williams syndrome assessed using a nonverbal task. *Journal of Autism and Developmental Disorders, 38*(5), 806–814. https://doi.org/10.1007/s10803-007-0447-4

Posner, M. I., & Rothbart, M. K. (1998). Attention, self–regulation and consciousness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1377), 1915–1927. https://doi.org/10.1098/rstb.1998.0344

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12*(3), 427–441. https://doi.org/10.1017/S0954579400003096

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment, 87*(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09

Reid, V. M., Csibra, G., Belsky, J., & Johnson, M. H. (2007). Neural correlates of the perception of goal-directed action in infants. *Acta Psychologica, 124*(1), 129–138. https://doi.org/10.1016/j.actpsy.2006.09.010

Rothbart, M. K., Ellis, L. K., Rosario Rueda, M., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality, 71*(6), 1113–1144. https://doi.org/10.1111/1467-6494.7106009

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*(2), 331–349. https://doi.org/10.1111/j.2044-835X.1991.tb00881.x

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences, 36*(4), 393–414. https://doi.org/10.1017/ S0140525X12000660

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition, 133*(1), 43–47. https://doi.org/10.1016/ j.cognition.2014.05.016

Shultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology, 17*(4), 465–471. https://doi.org/ 10.1037/0012-1649.17.4.465

Simlesa, S., Hacin, K., Cepanec, M., & Ivsac-Pavlisa, J. (2021). Importance of language for children's theory of mind: Comparison of verbal and nonverbal theory of mind tasks. *Psihologija, 54*(2), 123–135. https://doi.org/10.2298/ PSI190924023S

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

List of Figures

List of Abbreviations

ANOVA .... *Analysis of Variance*
BR .... *Belief Reasoning*
CI .... *Confidence Interval*
df .... *Degrees of Freedom*
FB .... *False Belief*
GLMM .... *Generalized Linear Mixed Model*
M .... *Mean*
PAR .... *Perceptual Access Reasoning*
r .... *Correlation Coefficient*
SD .... *Standard Deviation*
TB .... *True Belief*
ToM .... *Theory of Mind*

Appendix

Abstract in German

In dieser Studie wurde ein nonverbales Theory of Mind Versteckspiel durchgeführt, um zu erforschen, wie Vorschulkinder (N = 33) auf den irreführenden Hinweis eines Komplizen reagieren. Das Kind beobachteten einen Komplizen (Verstecker), als dieser einen Sticker in einem von zwei Kübeln versteckte. Ein weiterer Komplize (Kommunikator) gab dem Kind wiederholt einen korrekten Hinweis darüber, wo sich der Sticker befand (Eingewöhnungsphase). In zwei Testbedingungen gab der Kommunikator dem Kind einen falschen Hinweis über den Versteckort des Stickers, während er eine wahre oder eine falsche Überzeugung über den Versteckort des Stickers hatte. Hatte der Kommunikator eine falsche Überzeugung, war es den Kindern möglich, den falschen Hinweis zu ignorieren und den richtigen Kübel signifikant öfter zu wählen, als es durch zufälliges Wählen zustande kommen würde. Hatte der Kommunikator hingegen eine wahre Überzeugung, war es den Kindern nicht möglich, den falschen Hinweis zu ignorieren und den richtigen Kübel öfter zu wählen, als es durch zufälliges Wählen zustande kommen würde. Die Überzeugung, die der Kommunikator hatte, beeinflusste die Wahl des Kübels der Kinder nicht signifikant. Es zeigte sich ein Haupteffekt des Alters der Kinder, der zeigt, dass ältere Kinder in beiden Bedingungen wahrscheinlicher den richtigen Kübel wählten. Unsere Ergebnisse deuten darauf hin, dass Kinder zwischen fünf und sechs Jahren Probleme damit hatten, die Überzeugung des Kommunikators zu repräsentieren, um richtig zu wählen, oder andere Wahlstrategien anwandten, welche nicht auf der Repräsentation der Überzeugung des Kommunikators basierten. In einer explorativen Analyse zeigte sich, dass das Sprachverständnis und die Fähigkeit zur Impulskontrolle der Kinder deren Entscheidungen nicht beeinflussten. Dies lässt vorläufig schließen, dass das Versteckspiel nonverbal funktioniert und weniger exekutive Ansprüche an die Kinder stellt als verbale Theory of Mind Aufgaben.