

It takes one to know one: Theory of mind helps children to detect lies that are revealed by semantic leakage

Özlem Yeter¹ (O.Yeter@rug.nl)
Barteld Kooi¹ (B.P.Kooi@rug.nl)
Rineke Verbrugge² (L.C.Verbrugge@rug.nl)
Petra Hendriks³ (P.Hendriks@rug.nl)

¹Department of Theoretical Philosophy, University of Groningen, The Netherlands

²Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, The Netherlands

³Center for Language and Cognition Groningen, University of Groningen, The Netherlands

Abstract

Can children detect a lie when the liar unintentionally reveals essential information (displaying so-called semantic leakage)? Furthermore, because previous research found that theory of mind (ToM) is a factor in children's lie production ability, what role does ToM play in their lie detection ability? An experiment was carried out with 128 Dutch-speaking children (4 - 14 years old). Children's lie detection ability was assessed using a story in which one of the characters produced a lie signalled by semantic leakage. A false-belief task was used to test children's first-order ToM (Kevin thinks that...) and second-order ToM (Marieke thinks that Kevin thinks that...). Children were 73% accurate in identifying the liar by referring to semantic leakage. Their performance improved with age, with 12-year-olds showing ceiling performance. Finally, first-order ToM was a significant predictor, even after controlling for age, suggesting that lie production and lie detection partly rely on shared cognitive functions.

Keywords: lie detection; semantic leakage; theory of mind; children; cognitive development

Introduction

For navigating society, it is essential to be able to distinguish lies from truths. This holds especially for today's society, where social media often serve as a source of unreliable information. However, several studies have shown that adults are not very good at detecting lies (for a review, see Bond & DePaulo, 2008). Detecting lies seems particularly challenging for children. First of all, their limited cognitive maturity and experience makes them more vulnerable to lies than adults. Furthermore, they learn about themselves and the world around them partly on the basis of what others tell them (Clement, Koenig, & Harris, 2004), and this requires an assumption that what they are told is true. However, even parents – typically trustworthy sources – lie to their kids occasionally, for example as an instrument to educate them (Heyman et al., 2013). Thus, a relevant question is how children learn to distinguish lies from truths and thus learn to detect lies.

Because lies are not merely untruthful statements but also involve the intentions and beliefs of others, the cognitive function of theory of mind (ToM) is thought to play a crucial role in both lie telling and lie detection (Lee, 2013). Theory of mind is the ability to understand that other people's beliefs can differ from one's own beliefs and from reality (Premack

& Woodruff, 1978; Wimmer & Perner, 1983). Lying involves the manipulation of others' mental states. In order to manipulate a lie receiver's mental state, a lie teller should evaluate what the lie receiver knows, believes and feels. Similarly, a lie receiver should evaluate the intentions of the lie teller and relate these to what is said.

This study examines the relationship between the ability to detect lies and theory of mind understanding in children. Before presenting our study, we first review the literature on children's ability to tell lies and its relation to theory of mind, and then on children's ability to detect lies.

Children as lie tellers

Children already tell lies from a young age. To explain children's lie-telling behaviour, Talwar and Lee (2008) proposed a developmental model consisting of three levels. According to this model, children start producing "primary lies" around the age of two (Ahern, Lyon, & Quas, 2011). That is, some children deny their wrongdoings (e.g., peeking when told not to), but without the intention of deceiving the listener. Some researchers suggest that children's denial of their wrongdoing could be a reflection of their desires and emotions (Hummer, Wimmer, & Antes, 1993; Sodian, 1991). So, when a child says "no" to the transgression question "Did you peek?", they might mean "I wish I had not peeked" (Ahern et al., 2011).

Children start telling "secondary lies" between the ages of three and four. These children deny their wrongdoings and try to control their non-verbal behaviour in order to appear honest (Polak & Harris, 1999; Talwar & Lee, 2002). However, at this stage, they still have difficulty maintaining their lies. For example, most children at this stage are unable to control *semantic leakage* (Talwar & Lee, 2008). Semantic leakage is the phenomenon of unintentionally revealing ('leaking') information that one could not have known if one were truthful (Yeter et al., 2024). For example, in Talwar and Lee's (2002) experiment that uses a temptation resistance paradigm, children are shown a toy and are asked to guess what the next toy will be. The experimenter then leaves the room and tells the child not to peek at the next toy. However, while the experimenter is away, most children peek at the toy. When the experimenter returns and asks them whether they have peeked, many children lie and say they did not peek.

Among these children, those younger than five often fail to control semantic leakage. They correctly name the toy when asked next to guess what the toy is, thus revealing that they had been lying (see also Polak & Harris, 1999).

Children between the ages of six and eight begin producing “tertiary lies” that are more sophisticated and are tailored to the receiver’s mental state. At this stage, children are better able to maintain their lies by controlling semantic leakage (Talwar & Lee, 2008). This ability improves with age (Lavoie & Talwar, 2020; O’Connor, Dykstra, & Evans, 2020; Williams et al., 2016), and has been found to depend on cognitive factors such as executive function (EF) and theory of mind (ToM) (Arslan, Hohenberger, & Verbrugge, 2017).

In this study, we focus on the relation between lies and ToM. ToM has a recursive nature, since one may not only reason about another person’s mental state, but also about what that person thinks about someone else’s mental state, and so on. ToM reasoning starts at *zeroth order*: Until around the age of four, children cannot yet infer others’ mental states and are limited to their own cognitive perspective (Wellman, Cross, & Watson, 2001). *First-order* false-belief understanding emerges between the ages of four and five. It allows children to attribute mental states to others and understand that other people’s beliefs can differ from their own and from reality (Wimmer & Perner, 1983). For instance, at this stage, they can understand that someone can falsely believe that pigs can fly. Children’s denial of wrongdoings is related to their first-order false-belief understanding (Polak & Harris, 1999; Talwar & Lee, 2008). *Second-order* false-belief understanding emerges between the ages of five and nine (Perner, 1988). At this stage, children can attribute first-order reasoning to another person: They can reason that person B believes that person A believes that pigs can fly. Children’s ability to maintain their lies over multiple question-answer sequences by controlling semantic leakage is related to second-order ToM (Talwar, Gordon, & Lee, 2007).

So, different levels of lie-telling behaviour are related to different stages of ToM reasoning. ToM itself is related to yet other factors. For example, having siblings and being bi- or multilingual may positively influence ToM development in childhood (Goetz, 2003; Zhang et al., 2021).

Children as lie detectors

Three-year old children still have a strong tendency to believe what they are told, and therefore have difficulty detecting lies (e.g., Jaswal et al., 2010; Palmquist & Kondrad, 2024). At age four, children can detect logical inconsistencies, such as the claim that a ball is big and small at the same time (Doebel, Rowell, & Koenig, 2016). When presented with an implausible statement that contradicts the real world (e.g., a ghost jumping out of a book and breaking a glass), most five- and six-year-old children label it as a lie, while most younger children believe the statement to be true (Lee et al., 2002).

However, lies are not only characterized by being inconsistent with other statements or with reality, but crucially involve the liar’s intention to deceive (Mascaro & Sperber, 2009; van Ditmarsch, Hendriks, & Verbrugge,

2020). That is why the ability to detect inconsistencies does not guarantee the ability to detect lies. In support of this, Wimmer, Gruber and Perner (1984) showed that children between the ages of 4 and 6 still incorrectly evaluate a well-intended false statement as a lie, instead of a mistake (see also Bussey, 1999).

Maas (2010) tested children’s understanding of lies with short stories, where the inner thoughts of the main character were shared with the children. For instance, a sincere character wants to play but refuses a game invitation from her friend because she has to help her mom at home, and she promises to play with her friend another time. An insincere character refuses the same game invitation with the same excuse but in reality, she does not want to play with her friend. Four-year-old children who pass first-order false belief tasks correctly label sincere characters’ statements as promises and insincere characters’ statements as lies, showing sensitivity to the characters’ intentions underlying their statements (see also Nancarrow et al., 2018).

Although the above suggests a relation between lie detection and ToM, findings in the literature are mixed. For example, Palmquist and Kondrad (2024) do not find such a relation in three-year-olds. Thus, this warrants further investigation.

Present Study

In research on children’s lie-telling behaviour, the role of semantic leakage control during lie telling has been extensively studied. However, not much is known about children’s use of semantic leakage for lie detection. For example, we know that children cannot control semantic leakage until about the age of 6, but we do not know whether and at what age children can make use of semantic leakage to detect other people’s lies. Given that even adults fail to exceed chance-level accuracy if the lie is not signalled by semantic leakage, and that adults use semantic leakage as a cue when it is available (Yeter et al., 2024), in the present study we aim to test children’s ability to detect lies signalled by semantic leakage.

Additionally, first-order ToM has been shown to play a role in children’s lie production abilities. However, it is not known whether ToM plays a role in children’s lie detection abilities too.

As far as we know, the present study is the first to examine children’s ability to detect lies signalled by semantic leakage, and whether this ability is linked to ToM. In light of the literature on lie telling and lie detection in children, we hypothesise that first-order false belief understanding predicts children’s lie detection performance. Furthermore, we expect children’s performance at lie detection to improve with age.

Method

In this study, an experiment was carried out consisting of three parts: (1) a short questionnaire, (2) the Chocolate Bar Story testing children’s first- and second-order ToM abilities, and (3) the Apple Story testing children’s ability to detect lies

using semantic leakage as a cue. Both stories were illustrated by Avik Kumar Maitra.

Participants

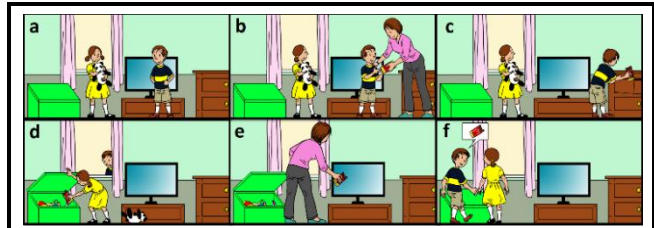
The data was collected in 2023 and 2024 at science festivals held at two different locations in the Netherlands, namely Zpandend Zernike in Groningen and the Science Weekend at the Royal Netherlands Academy of Arts and Sciences (KNAW) in Amsterdam. In Groningen, data was collected both in 2023 ($N = 57$) and 2024 ($N = 40$). In Amsterdam, data was collected only in 2024 ($N = 44$). Tests were conducted in Dutch. Data of 13 children who were not native speakers of Dutch (12 from the test session in Groningen in 2023, and 1 from the test session in Groningen in 2024) were excluded. Thus, this study presents data from 128 monolingual and bi-/multilingual Dutch-speaking children (62 female), whose age ranges between 4;5 and 14;1 ($M_{age} = 9;0$, $SD = 1;11$). Of the 128 children, 80 were bi-/multilingual, and 109 had at least one sibling.

Prior to the experiment, children and their accompanying caregivers gave their oral and written informed consent, respectively. Each child received a sticker after their participation. The study was reviewed by the Ethical Review Committee CETO of the University of Groningen.

Design and Materials

Questionnaire A short questionnaire was presented to caregivers to identify their child's individual characteristics: age, gender, language status, sibling status. In the 2024 testing sessions, but not those of 2023, we also asked them whether their child had been diagnosed with any neuro-cognitive disorders such as autism, because the ToM development of neurodivergent children and neurotypical children may differ. Regardless of their answer to this question, all children who met the language requirement were included in the analyses. Since data for diagnostic information was unavailable for the 2023 sample, we did not include diagnosis as a variable in our analysis. However, because of neurodiversity in our sample (autism: $N = 3$, ADHD: $N = 2$, dyslexia: $N = 4$, gifted: $N = 8$) and a wide age distribution, we expected to see variation in our ToM results.

Chocolate Bar Story To test children's first- and second-order false belief understanding, we used the Chocolate Bar Story by Arslan et al. (2020), a task based on a three locations task (Flobbe et al., 2008), adapted from Hogrefe, Wimmer and Perner (1986) and translated to Dutch (see Figure 1 for an English version). In this task, six pictures were presented sequentially on a laptop screen, while a story was told by the experimenter. At the end of the story, all pictures were visible on the screen. Interspersed throughout the story, three control questions, a first-order false-belief question and a second-order false-belief question were asked (correct answers are given in parentheses in Figure 1). For the two false-belief questions, children were also asked to give a justification.



a) Kevin and Marieke are brother and sister. They are in the living room. b) Their mother bought a chocolate bar and gives it to Kevin. Marieke doesn't get any chocolate, because she has been naughty. c) Kevin eats some of his chocolate and puts the remainder into the drawer. He doesn't give any chocolate to Marieke. Marieke is upset that she does not get any chocolate. d) After that, Kevin goes to help his mother in the kitchen. Marieke is alone in the room. Because she is upset, she takes the chocolate from the drawer and puts it into the toy box. While she is putting the chocolate into the toy box, Kevin is passing by the window. He sees how Marieke takes the chocolate out of the drawer and puts it into the toy box. Marieke does not see Kevin.

Control question 1: "Does Kevin know that Marieke put the chocolate into the toy box?" (yes)

Control question 2: "Does Marieke know that Kevin saw her put the chocolate into the toy box?" (no)

e) After that, Kevin goes back to the kitchen and Marieke goes to the kitchen, as well. While Kevin and Marieke are in the kitchen, their mother goes to the living room to watch TV. While she is searching for the remote control, she sees the chocolate in the toy box. The mother is surprised that the chocolate is in the toy box. She takes the chocolate from the toy box and puts it into the TV stand. She watches TV for a while and goes to her room.

Control question 3 (reality question): "Where is the chocolate now?" (in the TV stand)

f) Now, Kevin and Marieke go back to the living room. Kevin wants to eat some of his chocolate. He says 'Hmm I would like to eat some chocolate'.

First-order false belief question: "Where will Kevin look for the chocolate?" (in the toy box)

Justification question: "Why does he look there?"

Second-order false belief question: "Where does Marieke think that Kevin will look for the chocolate?" (in the drawer)

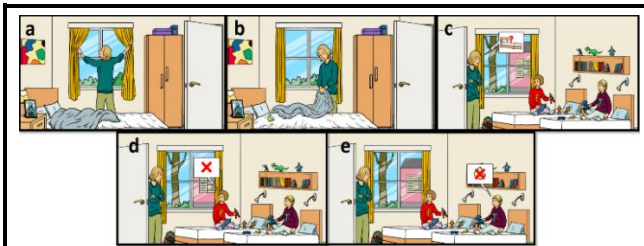
Justification question: "Why does she think that?"

Figure 1 | Chocolate Bar Story for testing first- and second-order theory of mind abilities (Arslan et al., 2020).

Scoring the Chocolate Bar Story task took place at two levels: judgments and justifications. Judgment and justification responses were scored separately for first- and second-order false belief questions. For judgments, children received a score of 1 for the first- or second-order false belief question if they gave a correct answer, and 0 if they gave an incorrect answer. Additionally, if they correctly answered the judgment question, their justification was scored as 1 if

correct and as 0 if incorrect. Justifications were classified as correct if they referred to the mental state of the related story character or included information about a character’s history of exposure to relevant information (see Arslan et al., 2020).

Apple Story The Apple Story is a story accompanied by five pictures (Figure 2). The pictures were presented sequentially on a laptop screen, while the Dutch version of the story was told by the experimenter. At the end of the story, all pictures were visible on the screen and the children received a lie detection question followed by a justification question.



a) Here is Lisa. Lisa is the mother of two boys. She is tidying up her own bedroom. b) The mother goes to her bed to tidy it. She pulls back the blanket. And then she finds something strange in her bed. It is a half-eaten apple! The mother is very surprised. c) The mother leaves the apple in the bed and walks to her sons’ bedroom. There, the two boys, Peter and Tom, are playing with their toys. Peter and Tom see their mother come into the room. The mother asks them: “Who put something in my bed?” d) Peter says: “I did not put anything in your bed.” e) Tom says: “I did not put an apple in your bed.”

Lie detection question: “You are the detective! What do you think? Who put an apple in the mother’s bed?” (Tom)

Justification question: “Why do you think so?”

Figure 2 | Apple Story (version 1) for testing lie detection.

In this task, the question “Who is lying?” was avoided because the literature shows that young children do not fully understand the concept of lying. Instead, the children were asked to indicate who had transgressed in the story. They could base their response on semantic leakage, because Tom referred to the apple in his answer to his mother’s question, although his mother did not mention the apple. For a correct answer, a child should therefore deduce that Tom could not have known what the object in the bed is, unless he had put the apple in the bed himself (or saw who did it). To control for the effect of order of question presentation, we created a second version in which we swapped pictures *d* and *e*: In version 2, Tom answers first and Peter second. These two versions were counterbalanced across participants. In total, 65 children received version 1 and 63 children version 2.

Scoring was conducted at two levels for the Apple Story. Children received a score of 1 for a correct judgment and 0 for an incorrect judgment. Additionally, if children provided

a correct judgment on the lie detection question, they received a score of 1 for a correct justification if this justification referred to semantic leakage (e.g., “Tom knows it’s an apple”, “the mother did not mention an apple”). Otherwise, they received a score of 0 for their justification.

Analysis

Children might make correct judgments but fail to justify them correctly due to immature language skills. However, we would not capture the whole picture if we focused solely on the scores for judgment questions, since children may simply guess when selecting an answer, without knowing the correct answer. To provide a full picture, analyses were done in two steps: by looking at the judgment scores (0 or 1) and, for the children who gave a correct judgment, at their justification scores (0 or 1).

First, we present the descriptive and inferential statistics for the characteristics of the individual participants, their accuracy rates on ToM assessed by the Chocolate Bar Story, and their accuracy rates on lie detection assessed by the Apple Story. Next, we present the results of our logistic regression analysis, with the scores on the Apple Story as the dependent variable. The latter analysis allows us to determine which factors influence children’s ability to detect lies.

Results

Theory of Mind

Children had an average accuracy rate of 97% on the three control questions in the Chocolate Bar Story task, indicating ceiling performance. As expected, children had higher accuracy scores for the first-order false belief questions than for the second-order false belief questions. Table 1 summarises children’s performance on the first- and second-order false-belief questions in the Chocolate Bar Story, distinguishing between performance on the judgment questions and performance on the justification questions.

For the first-order false-belief question (i.e., *Where will Kevin look for the chocolate?*), 88% ($N = 112$) of the children gave the correct answer, *toy box*, and 84% ($N = 107$) of the children also gave a correct justification. For the second-order false belief question (*Where does Marieke think that Kevin will look for the chocolate?*), 81% ($N = 104$) of the children gave the correct answer, *drawer*, while 73% ($N = 94$) of the children also correctly justified their answers.

Table 1 | Percentages of correct answers (standard errors) for false belief questions.

Question Type	Accuracy
First-order false belief judgment	88% (.029)
First-order false belief justification	84% (.033)
Second-order false belief judgment	81% (.035)
Second-order false belief justification	73% (.039)

Children’s mean accuracy on first- and second-order false-belief (henceforth ToM1 and ToM2, respectively) judgment questions was similar, $p > .05$. However, their accuracy on justifications was higher for the ToM1 than for the ToM2 question, $t(127) = 2.377, p = .019, 95\% \text{ CI } [-.017, .186]$.

We further investigated the effect of individual differences on children’s scores on the false belief questions. First, a binary logistic regression analysis was run, with ToM1 judgment scores as the dependent variable and participant characteristics (i.e., age, gender, language status, and sibling status) as predictor variables. This analysis was repeated three times, each with a different dependent variable: ToM1 justification scores, ToM2 judgment scores, and ToM2 justification scores. Preliminary analyses suggested that the assumption of multicollinearity was met in all four models (Tolerance $> .93$), and the results of the Hosmer and Lemeshow test indicated a good model fit, $p > .05$.

For *ToM1 judgment scores*, the overall model was significant, $\chi^2(4) = 21.228, p < .001$. Higher age, $B = 0.055, SE = 0.017, Wald = 10.83, p = .001, \text{Exp}(B) = 1.056, 95\% \text{ CI } [1.022, 1.092]$ and being a girl predicted higher accuracy, $B = -1.87, SE = 0.72, Wald = 6.77, p = .009, \text{Exp}(B) = 0.15, 95\% \text{ CI } [0.038, 0.630]$. For *ToM1 justification scores*, the overall model was also significant, $\chi^2(4) = 21.914, p < .001$. Again, higher age $B = 0.053, SE = 0.014, Wald(1) = 13.68, p < .001, \text{Exp}(B) = 1.05, 95\% \text{ CI } [1.03, 1.08]$ and being a girl predicted higher accuracy, $B = -1.27, SE = 0.59, Wald(1) = 4.63, p = .031, \text{Exp}(B) = 0.28, 95\% \text{ CI } [0.09, 0.89]$. For *ToM2 judgment scores*, the model significantly explained the variance in children’s performance scores, $\chi^2(4) = 22.841, p < .001$, and age was the only significant predictor, $B = 0.053, SE = 0.013, Wald \chi^2(1) = 15.623, p < .001, \text{Exp}(B) = 1.054, 95\% \text{ CI } [1.027, 1.083]$. For *ToM2 justification scores*, the model was significant too, $\chi^2(4) = 28.393, p < .001$, with age as the only significant predictor, $B = 0.047, SE = 0.011, Wald = 17.86, p < .001, \text{Exp}(B) = 1.048, 95\% \text{ CI } [1.025, 1.071]$.

So, children’s ToM performance was influenced by age and (only for ToM1) gender. Whether the child was mono- or bilingual and had siblings did not have an effect.

Lie Detection Accuracy

In this section, we present the results on the Apple Story task. First, we checked whether children’s performance on the Apple Story varied between version 1 and 2. T-test results revealed no significant difference between the two versions, neither for judgment scores [$t(126) = .327, p > .05$] nor for justification scores [$t(126) = .106, p > .05$].

Table 2 | Percentages of correct answers (standard errors) for lie detection questions.

Question Type	Accuracy
Lie detection judgment	87% (.030)
Lie detection justification	73% (.039)

Table 2 presents the accuracy rates of children on the lie detection questions in the Apple Story. For the lie detection judgment question (i.e., *Who do you think put an apple in mom’s bed?*), the overall accuracy was 87% ($N = 111$), while it was 73% ($N = 94$) for the justification question (*Why do you think so?*).

Factors Influencing Lie Detection

To investigate the factors influencing lie detection, we first explored participant characteristics that might influence children’s lie detection ability. When lie detection judgment scores were used as the dependent variable, the model failed to explain the variance, $p = .158$, and even age was not a significant predictor. An explanation for the absence of effects of participant characteristics on lie detection judgment accuracy may be that a two-choice task is not very sensitive and children may have accidentally guessed correctly. Indeed, several children selected the correct answer (Tom) but provided an irrelevant justification such as “because he responded first”, “because the apple is green and his shirt is green”, or “because he looks fat”. For this reason, in the following analyses we use justification scores as the dependent variable for lie detection ability.

A second regression analysis was run with lie detection justification scores as the dependent variable and individual characteristics as predictor variables. This model was significant, $\chi^2(4) = 17.994, p = .001$, and the Hosmer and Lemeshow test indicated a good fit between the model and the data, $\chi^2(8) = 2.767, p = .948$. Age was the only significant predictor that could explain higher lie detection justification scores, $\text{Exp}(B) = 1.036, 95\% \text{ CI } [1.015, 1.057], Wald = 11.781, p = .001$.

To illustrate the effect of age on lie detection justification: The earliest instance of an accurate justification was observed in one of the four 5-year-olds. Mean accuracy at this age was only 25%. Children’s performance improved gradually with age, and was 70% accurate at age 6, 90% accurate at age 9, and 100% accurate from age 12 onwards.

A Pearson’s chi-squared test with Yates’ continuity correction showed that lie detection justification scores were associated with ToM1 justification scores [$\chi^2(1, N = 128) = 10.24, p = .001$] and ToM2 justification scores [$\chi^2(1, N = 128) = 7.31, p = .007$] (Figure 3). The association remained significant when the ToM justification scores were replaced by ToM judgment scores.

Next, a regression analysis was run to investigate the relationship between ToM and lie detection justification ability. First, ToM1 and ToM2 judgment scores were used as independent variables, while lie detection justification scores were used as the dependent variable. The assumption of multicollinearity was met (tolerance = .941, VIF = 1.062) and the Hosmer-Lemeshow test indicated good model fit, $\chi^2(1) = 0.45, p = .831$. This model was significant $\chi^2(2) = 7.952, p = .019$, indicating that the predictors collectively explain variance in the outcome. However, neither of the individual predictors reached statistical significance (both $p > .05$). Possibly, this is because ToM judgment scores are not

sensitive enough to differentiate among these relatively older children (with a mean age of 9).

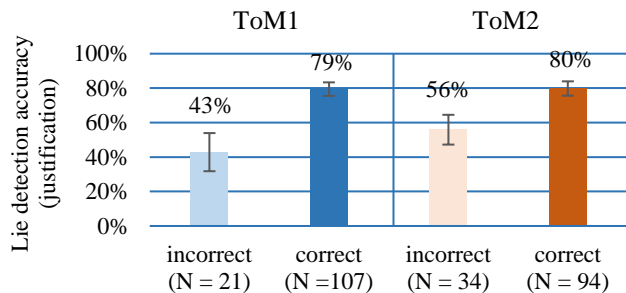


Figure 3 | Lie detection justification accuracy for children who gave an incorrect versus correct justification for their answer to the first-order false belief (ToM1, left panel) and second-order false belief (ToM2, right panel) question.

We ran the same analysis again, but now with justification instead of judgment scores for ToM tasks. This model was significant, $\chi^2(2) = 13.788, p = .001$, with only ToM1 predicting higher scores on lie detection justification, $B = 1.383, SE = 0.525, Wald = 6.931, p = .008, Exp(B) = 3.987, 95\% CI [1.424, 11.165]$. ToM1 justification scores remained a significant predictor even after controlling for age, $B = 1.107, SE = 0.548, Wald = 4.079, p = .043, Exp(B) = 3.024, 95\% CI [1.033, 8.853]$. These findings suggest that children’s ability to detect lies that are signalled by semantic leakage relies on first-order, but not on second-order, ToM justification ability.

Discussion

To the best of our knowledge, this study is the first to examine children’s ability to use semantic leakage as a cue for lie detection and the role of ToM in this process. In total, 73% of the children both identified the liar and provided correct justifications by referring to the semantic leakage. This ability improved with age, and better first-order ToM justifications significantly predicted children’s lie detection performance. These findings align with the results of Yeter et al. (2024) on adults. They found that adults reached 90% accuracy when lies were signalled by semantic leakage. This lends support to Levine (2010), who argues that humans are not as bad at lie detection as sometimes described in the literature (e.g., Bond & DePaulo, 2006) and suggests that individuals perform much higher than chance levels when the lie teller ‘leaks’ information.

Our results further show that ToM1 understanding as measured by justifications was related to lie detection as measured by justifications even after controlling for age. This suggests that the contribution of ToM is not simply a reflection of general cognitive maturation. Instead, there seems to be a specific role for ToM in detecting lies, independent of overall cognitive development as reflected by age. This relationship with lie detection was not observed for

ToM2. One explanation is that a mature level of ToM1 could already be sufficient for using semantic leakage as a cue to detect lies. Alternatively, ToM2 may be involved as well but, to justify their ToM2 answer, children also need good verbal and EF skills (Hollebrandse, van Hout, & Hendriks, 2014; cf. Arslan et al., 2020). Future studies could control for these skills while testing the influence of ToM2.

Our analysis showed different results with ToM judgment scores than ToM justification scores as the predictor variable. This could be due to ceiling performance on first-order ToM judgments, since most children in our study were older than the age at which first-order ToM emerges (i.e., first-order false-belief tasks are passed from age 4 or 5). Because giving ToM justifications seems to be more challenging than giving ToM judgments, the justification score could provide a more useful measure of ToM understanding in older children.

Past research has shown that ToM is a crucial factor in lie-telling (e.g., Talwar & Lee, 2008). Our study shows that ToM is crucial for lie detection, as well. Both the detection and the production of lies depend on first-order ToM, which suggests that lie production and lie detection share the same cognitive mechanisms. Future research could explore whether children who are skilled lie-tellers are also effective lie detectors, that is, whether it really “takes one to know one”. While our study compares findings from earlier lie production research to our own results, follow-up studies should directly investigate this by testing the same participants on both lie production and lie detection.

Gender, specifically being a girl, was observed to play a positive role in making first-order ToM judgments and justifying them, even when controlling for age. Similar observations were made in earlier studies (e.g., Charman, Ruffman, & Clements, 2002). A relevant factor may be that parents use more emotion words with their daughters and respond more to their emotions, giving girls an advantage. (Adams et al., 1995; Simpson & Stevenson-Hinde, 1985). There is also evidence of a neurological advantage for girls (e.g., Adenzato et al., 2017). However, we did not find an effect of gender on second-order ToM understanding or lie detection. Further research should investigate the role of gender in children’s lie detection abilities.

This study also has some limitations. Because our data was collected at science festivals, our study does not have evenly distributed age groups. Hence, we are unable to provide precise cut-off points for the development of children’s ability to detect lies. Additionally, diagnosis information was not available for all children, and where it was available it resulted in small sample sizes. Therefore, we could not draw solid conclusions about the influence of neurodiversity on lie detection. However, note that this was not an aim of our study.

In conclusion, our study shows that school-aged children are able to detect lies signalled by semantic leakage, and that first-order ToM contributes to their lie detection ability.

References

- Adams, S., Kuebli, J., Boyle, P. A., & Fivush, R. (1995). Gender differences in parent-child conversations about past emotions: A longitudinal investigation. *Sex Roles, 33*(5–6), 309–323.
- Adenzato, M., Brambilla, M., Manenti, R., De Lucia, L., Trojano, L., Garofalo, S., Enrici, I., & Cotelli, M. (2017). Gender differences in cognitive Theory of Mind revealed by transcranial direct current stimulation on medial prefrontal cortex. *Scientific Reports, 7*(1), 41219.
- Ahern, E. C., Lyon, T. D., & Quas, J. A. (2011). Young children's emerging ability to make false statements. *Developmental Psychology, 47*(1), 61–66.
- Arslan, B., Hohenberger, A., & Verbrugge, R. (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLOS ONE, 12*(1), e0169510.
- Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2020). Accelerating the development of second-order false belief reasoning: A training study with different feedback methods. *Child Development, 91*(1), 249–270.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234.
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*(4), 477–492.
- Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development, 70*(6), 1338–1347.
- Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development, 11*(1), 1–10.
- Ditmarsch, H., Hendriks, P., & Verbrugge, R. (2020). Editors' review and introduction: Lying in logic, language, and cognition. *Topics in Cognitive Science, 12*(2), 466–484.
- Doebel, S., Rowell, S. F., & Koenig, M. A. (2016). Young children detect and avoid logically inconsistent sources: The importance of communicative context and executive function. *Child Development, 87*(6), 1956–1970.
- Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information, 17*(4), 417–442.
- Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6*(1), 1–15.
- Heyman, G. D., Hsu, A. S., Fu, G., & Lee, K. (2013). Instrumental lying by parents in the US and China. *International Journal of Psychology, 48*(6), 1176–1184. <https://doi.org/10.1080/00207594.2012.746463>
- Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567.
- Hollebrandse, B., van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese, 191*(3), 321–333.
- Hummer, P., Wimmer, H., & Antes, G. (1993). On the origins of denial negation. *Journal of Child Language, 20*(3), 607–618.
- 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.
- Lavoie, J., & Talwar, V. (2020). Care to share? Children's cognitive skills and concealing responses to a parent. *Topics in Cognitive Science, 12*(2), 485–503.
- Lee, K. (2013). Little liars: Development of verbal deception in children. *Child Development Perspectives, 7*(2), 91–96.
- Lee, K., Cameron, C. A., Doucette, J., & Talwar, V. (2002). Phantoms and fabrications: Young children's detection of implausible lies. *Child Development, 73*(6), 1688–1702.
- Levine, T. R. (2010). A few transparent liars explaining 54% accuracy in deception detection experiments. *Annals of the International Communication Association, 34*(1), 41–61.
- Maas, F. K. (2010). Children's understanding of promising, lying, and false belief. *The Journal of General Psychology, 135*(3), 301–322.
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*(3), 367–380.
- Nancarrow, A. F., Gilpin, A. T., Thibodeau, R. B., & Farrell, C. B. (2018). Knowing what others know: Linking deception detection, emotion knowledge, and theory of mind in preschool. *Infant and Child Development, 27*(5), 1–9.
- O'Connor, A. M., Dykstra, V. W., & Evans, A. D. (2020). Executive functions and young children's lie-telling and lie maintenance. *Developmental Psychology, 56*(7), 1278–1289.
- Palmquist, C., & Kondrad, R. (2024). Knowledge and source type influence children's skepticism of misinformation. *Journal of Cognition and Development, 25*(3), 437–460.
- Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*(2), 561–568.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.
- Simpson, A. E., & Stevenson-Hinde, J. (1985). Temperamental characteristics of three- to four-year-old boys and girls and child-family interactions. *Journal of Child Psychology and Psychiatry, 26*(1), 43–53.
- Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9*(1), 173–188.
- Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43*(3), 804–810.
- Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal*

- of *Behavioral Development*, 26(5), 436–444.
- Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, 79(4), 866–881.
- 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.
- Williams, S., Moore, K., Crossman, A. M., & Talwar, V. (2016). The role of executive functions and theory of mind in children's prosocial lie-telling. *Journal of Experimental Child Psychology*, 141, 256–266.
- Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism—moral subjectivism. *Journal of Experimental Child Psychology*, 37(1), 1–30.
- 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.
- Yeter, Ö., Kooi, B., Weerd, H. de, Verbrugge, R., & Hendriks, P. (2024). Semantic leakage enables lie detection, but first-person pronouns and verbosity can get in the way of detection. In L. Samuelson, S. F. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, pp. 2768-2775.