Journal of Experimental Child Psychology 231 (2023) 105639

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Toddlers' expressions indicate that they track agent–object interactions but do not detect false object representations

Qianhui Ni *, Jake Shoyer, Zoë Bautista, Alexandra Raport, Henrike Moll

Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

ARTICLE INFO

Article history:
Received 7 August 2022
Revised 10 January 2023

Keywords:
Social cognition
Theory of mind
Perspective taking
Dual systems theory
Expressions
Experiential records

ABSTRACT

In the theory of mind debate, a middle position between nativism and conceptual change theory has gained traction. This position states that children younger than 4 years track agent–object relations (by building "records" of others' experiences) without cognizing *how* agents represent—or misrepresent—the objects they encounter. We tested these claims with 3.5-year-olds using puppet shows geared to evoke suspenseful expressions. In two experiments ($N = 90$), children watched an agent approach an object that looked like her favorite food but was inedible. In Experiment 1, children showed tense expressions when an agent's real food item was, unbeknownst to her, replaced with a fake food item. Children, however, showed no signs of understanding that the agent would mistake the deceptive object for food. Consistent with this, children's expressions in Experiment 2 did not differ when the agent approached a deceptive object compared with when she approached a non-deceptive object. The experiments support the middle position's view that toddlers track agent–object interactions but fail to recognize when agents misrepresent objects.

* Corresponding author.
E-mail address: qni@usc.edu (Q. Ni).

https://doi.org/10.1016/j.jecp.2023.105639
0022-0965/

Introduction

After more than 40 years of intensive research, the debate about when and how children construct a theory of mind continues to flourish (Krupenye & Call, 2019; Premack & Woodruff, 1978). Conceptual change theory, which stated that children first come to understand beliefs when they acquire the concept of a belief at around 4 or 5 years of age (Wellman et al., 2001; Wellman & Woolley, 1990), was challenged about 20 years ago by nativists who supported their claim of an innately given theory of mind with reports that even infants can detect false beliefs (e.g., Baillargeon et al., 2010; Onishi & Baillargeon, 2005). But there are both conceptual and empirical problems with the nativist picture. Conceptually, it can be questioned whether something as basic as a look can signal something as complex as belief understanding (see Aslin, 2007; Perner & Ruffman, 2005); empirically, the looking-time data that nativists cite have resisted replication (Dörrenberg et al., 2018; Kulke & Hinrichs, 2021).

A middle position between nativism and classic conceptual change theory has been proposed, stating that infants have *some* theory of mind skills but lack a full grasp of beliefs and other representational states. Apperly and Butterfill's (2009; see also Butterfill & Apperly, 2013) two-systems account and Perner and Roessler's (2010, 2012) teleology-in-perspective account, despite somewhat different aspirations and terminologies, are variants of such a position. The position argues that toddlers track others' perspectives by building and drawing on so-called "experiential records." As Perner and Roessler (2012) wrote, "Infants keep track of what agents perceive (experiential record); in particular, the state of the world last seen by the agent. Focus on the agent activates this record and induces them to construe the agent's actions or to anticipate future actions on the basis of this record" (p. 524). According to the middle position, infants do this with no awareness of the specific way in which agents represent objects and thus without detecting when agents misidentify or misrepresent objects. To give an example, 3.5-year-olds would, in this view, grasp that Lucinda left her shoes by the front door and anticipate her returning there to fetch them even if her shoes have meanwhile been moved. Children at this age would not, however, think of Lucinda as falsely representing her shoes as still being by the front door or as mistaking a similar pair of shoes in the same location for her own.

The middle position embraces an eclectic methodical approach because only a combination of tests with varying affordances can establish toddlers' theory of mind abilities while also exposing their limits. Contingent on successful replication, infants pass "indirect" tests measuring looking time (He et al., 2011; Onishi & Baillargeon, 2005; Surian & Geraci, 2012), anticipatory looking (Clements & Perner, 1994), helping behavior (Buttelmann et al., 2009), and other behaviors (Rubio-Fernández & Geurts, 2013) because these tests only require the child to connect an agent's ongoing activity back to her prior object interactions—by relying on her experiential record. These tests do not, however, force the child to contemplate how the agent views or construes the object (Moll et al., 2022). By contrast, tests exceeding toddlers' theory of mind abilities, including classic location change (Wimmer & Perner, 1983), unexpected content (Hogrefe et al., 1986), appearance–reality tasks (Flavell et al., 1983), and identity relations tests (Perner et al., 2011), demand awareness of the aspectual shape or mode of presentation in which the object is presented to an agent.

Importantly, not only direct tests like the ones just mentioned but also indirect tests can expose the limits of toddlers' theory of mind. This is a problem for the nativists' claim that toddlers fail direct tests not because their mental state reasoning is immature but because eliciting answers overwhelms toddlers' processing capacities (Scott & Baillargeon, 2017; Setoh et al., 2016). Low and colleagues (2014; see also Low & Watts, 2013) displayed limits in 3-year-olds' belief reasoning with anticipatory looking tests, which do not elicit responses. In line with the middle position's predictions, toddlers did not foresee that an agent, due to her perspective, would perceive an ambiguous figure as showing a specific image (a rabbit or a duck; see Low et al., 2014) or mistake an object for a different one she witnessed earlier (Low & Watts, 2013). The negative results correspond with those from direct tests of children's understanding of identity relations, which likewise suggest that children younger than 4 or 5 years do not comprehend that someone might, for example, falsely assume that the morning star and the evening star are two different stars (Perner et al., 2011; Sprung et al., 2007).

In this study, we further tested the scope and limits of toddlers' theory of mind skills by measuring their expressions. Expressions are a useful window onto toddlers' grasp of the mind because they

reveal in a direct, non-symbolic way what an agent makes of a situation (Bar-On, 2013, 2017). Because expressions are prior to language (Bar-On, 2013; Darwin, 2002), they help to uncover early "empractical" awareness (Bühler, 1934/1990) of other minds that manifests in interaction and requires neither reflective distancing nor articulation. Moll et al. (2016, 2017) showed that 2- and 3-year-olds express tension by, for example, lip biting and brow furrowing when watching others act on false beliefs. The expression paradigm also uniquely captures the underappreciated *affective* dimension of erring. Learning that one is wrong often has affective repercussions (in the form of surprise and possibly disappointment). This is certainly true of children's theory of mind tests, whose protagonists typically fail to find desired objects.

We deployed the expression paradigm with 3.5-year-olds to measure the extent of theory of mind skills before the critical threshold of 4 or 5 years. In two experiments using puppet shows, we recorded children's expressive reactions as they watched an agent approach an object that looked like food but was known by children to be unpalatable. In Experiment 1, the agent approached the fake food either after she, along with the children, witnessed (Witnessed condition) or failed to witness (Unwitnessed condition) a demonstration of the object's inedibility. One group of children (Experiential Record group, hereafter the ER group) had previously seen how the agent interacted with and left behind a real food item that was then replaced with the indistinguishable lookalike. The other group saw the agent encounter the fake food "cold" without prior engagement with a real object that the fake object mimicked (No Experiential Record group, hereafter the No-ER group).

The middle position predicts that children will react with more expressions in the Unwitnessed condition compared with the Witnessed condition in the ER group but not in the No-ER group. This is because, according to this position, children will only realize that the agent cannot reconnect with her original object (in the ER group) without understanding that the agent (in both groups) mistakes the deceptive object for food. Because these predictions were confirmed, we extended the investigation in Experiment 2 by comparing children's expressions when an agent approached a deceptive object with when she approached a non-deceptive object.

Experiment 1

The experiment was approved by the university's institutional review board and was preregistered on the Open Science Framework (https://osf.io/te4va). Two groups of 3.5-year-olds watched puppet shows in which Puppet A (hereafter "A") approached a delicious-looking but inedible fake food item after she witnessed (Witnessed condition) or did not witness (Unwitnessed condition) Puppet B's (hereafter "B") demonstration that the object was artificial and inedible. One group had the chance to build experiential records by watching A leave her favorite food on stage before B replaced it with the inedible lookalike (ER group). The other group could not build experiential records because A did not interact with any object prior to approaching the fake food (No-ER group). Children in both groups watched B's demonstration and thus were aware that the object on stage was inedible. Children also received the Standard Appearance–Reality Task, allowing us to assess the relation between children's implicit knowledge (Expression Task) and explicit knowledge (Standard Appearance–Reality Task). Based on reports that children younger than 4 or 5 years have yet to acquire the ability to contrast appearance and reality in explicit judgments (e.g., Flavell et al., 1983, 1986), low performance was predicted in the Standard Appearance–Reality Task.

Method

Participants

A power analysis with alpha set at .05 and power of .80 (G*Power Version 3.1; *F* tests–linear multiple regression) yielded that a sample of $N = 52$ was necessary to detect an effect size $f^2 = .20$. For counterbalancing reasons, we decided on $N = 60$. The final sample included 60 English-speaking 3.5-year-olds (30 girls; $M = 42.29$ months, range = 36.77–47.30). An additional 6 children (4 from the ER group) were tested but excluded because they were uncooperative ($n = 3$) or there was indication that the children's identity and age were misreported ($n = 3$).

Children were recruited from social media, research mailing lists, and websites advertising childhood cognition studies (e.g., childrenhelpingscience.com). The ethnoracial composition of the sample was 23% multiracial, 5% African American, 5% Asian, 65% White, and 2% "other," with 15% of the children being Latinx. Children came from diverse socioeconomic backgrounds, as shown by widely varying household incomes (from <$20,000 to >$120,000).

Materials

Expression Task. Children watched puppet shows on computers or tablets with a minimum screen size of 11 inches. Puppet shows were displayed on a puppet theatre (71 × 74 × 14 cm) with a curtain and dowel mounted in the left corner. Each of four stories involved a different deceptive object and a different pair of puppets (all measuring 25–35 cm in height). The deceptive objects, shown in Fig. 1A, were a fake chocolate bar (5 × 3 × 0.6 cm), a fake cupcake (7.6 cm in diameter), fake cherries (each cherry 2.3 cm in diameter), and a fake ice cream bar (12 × 4 × 1.8 cm). A validation study (n = 20) with 3- and 4-year-olds confirmed that these objects were deceptive; all children thought that the objects were food and, depending on the story, 75% to 85% believed that the objects were the specific items they purported to be (e.g., cherries). A clear plastic cup (8.5 × 8.5 × 7 cm) was used to contain the chocolate, cherries, and ice cream bar in the response phase.

Standard Appearance–Reality Task. As depicted in Fig. 1B, the deceptive objects were a candle that looked like a crayon ("crayon–candle"; 7.4 × 0.8 cm) and a box that looked like a book ("book–box"; 20.3 × 14 × 2.5 cm). A small cake and a lighter served to demonstrate that the "crayon–candle" was a candle.

Design

Expression Task. A 2 × 2 experimental design was used with condition (Unwitnessed vs. Witnessed) as a within-participants variable and group (ER vs. No-ER) as a between-participants variable. Each child received four stories: chocolate story, cupcake story, cherry story, and ice cream story. Every story came in two versions: one in which A *did not* witness B's demonstration of the object's fakeness (Unwitnessed condition) and one in which A *did* witness it (Witnessed condition). Each child received

Fig. 1. Deceptive objects used in the Expression Task (A) and the Standard Appearance–Reality Task (B) in Experiment 1.

the Unwitnessed condition for two stories and the Witnessed condition for the other two stories. Story order and condition order were counterbalanced.

Standard Appearance–Reality Task. Children received an appearance question (AQ) and a reality question (RQ) for each of the two deceptive objects. Order of objects (crayon–candle first vs. book–box first), order of questions (AQ first vs. RQ first), and order in which the alternative labels were mentioned in the questions (e.g., "crayon" first vs. "candle" first) were counterbalanced.

Procedure

Children completed the Expression Task followed by the Standard Appearance–Reality Task in a single session on the video platform Zoom. Parents had previously given written consent, reported demographic information, and received instructions. The study was conducted by a single female experimenter (E). The child sat approximately 45 cm in front of the screen. The parent was positioned behind the child and was instructed not to interfere. E shared her screen and presented the videos in the order prescribed by the counterbalancing schedule. The session was recorded.

Expression Task. Each participant watched four prerecorded puppet show stories with the following structure (for individual stories, see Appendix A). A bell rang and the curtain opened. In the ER group, A entered with a food item (e.g., cupcake) from the left. After introducing herself, A tasted the food (e.g., by taking a bite from the cream of the cupcake) and said how much she liked it (e.g., "It tastes so good!"). A placed the food in the right corner of the stage, excused herself ("I have to go wash my hands now. [...] I will be back in a bit. Bye!") and exited left. In the Unwitnessed condition, B entered from the right with an inedible lookalike of A's food item. B demonstrated that the object was inedible ("Look at what I brought. This is not tasty at all. It is hard like wood. Look! You cannot eat this!"), replaced A's original object with the fake object ("I will take this one away and put this one here instead"), and exited right. In the Witnessed condition, the procedure was identical with the difference that after leaving for 3 s, A returned *prior to B's arrival* and positioned herself on the dowel. A witnessed the object's replacement, vocalizing "aha" when B demonstrated the deceptive object's inedibility and again when B placed the deceptive object where A's original object had been. The response phase started when A returned (Unwitnessed condition) or stood up from the dowel (Witnessed condition). A exclaimed "Oooh, al right —let me go get that!" and traversed the stage left to right humming. When A touched the object, the bell rang and the curtain shut, which terminated the response phase (lasting 20–23 s, depending on the story).

In the No-ER group, children watched the same puppet shows with the difference that A entered without food. After introducing herself, A declared, for example, "I love cupcakes! Do you like cupcakes? I eat cupcakes all the time!" and then left the stage again. The subsequent parts of the story were as in the ER group, with B demonstrating the lookalike's inedibility and placing it on stage ("Now let me put this here"). Fig. 2 shows the procedural steps by group and condition.

Standard Appearance–Reality Task. A female adult held up the first deceptive object and said to the child, "Look what I have here. What do you think this is?" After the child answered, the adult said "OK. But look!" and demonstrated the object's identity by opening the book–box or lighting the crayon–candle. The adult then asked the following test questions in the order prescribed by the counterbalancing schedule: "When you look at this with your eyes right now, what does it look like? Does it look like a [e.g., box] or does it look like a [e.g., book]" (AQ) and "What is it really? Is it really a [e.g., box] or really a [e.g., book]?" (RQ). If a child gave no answer for 5 s, the adult moved on to the next question. The procedure was repeated with the second deceptive object.

The entire session lasted about 20 min and concluded with parents receiving an electronic gift card.

Scoring and reliability

Expression Task. A rater who was unaware of group and condition (her vision of the left part of the stage was blocked and the audio was turned off) judged based on video-recordings whether tense expressions were absent or present (0 = expressions absent, 1 = expressions present) in each response phase. If expressions were present, she judged how many expressions occurred (e.g., 2 = two

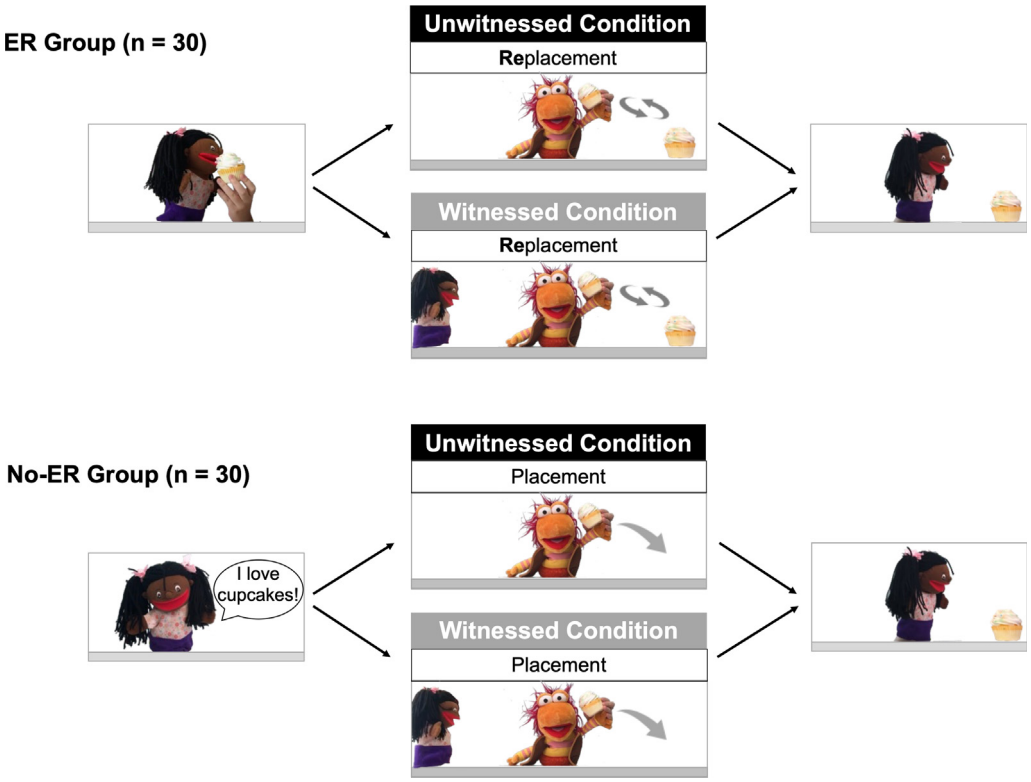

Fig. 2. Procedural steps of the Expression Task in Experiment 1. Children in the Experiential Record (ER) group witness Puppet A (A) interacting with a real cupcake. Puppet B (B) enters and demonstrates the fakeness of a lookalike, which she then substitutes for the real cupcake in A's presence (Witnessed condition) or absence (Unwitnessed condition). Finally, A approaches the fake cupcake. Children in the No Experiential Record (No-ER) group do not witness A interacting with a real cupcake. Instead, they hear A expressing her fondness of cupcakes. The remainder of the procedure is the same as in the ER group with the difference that B places the fake cupcake on stage without replacing a real cupcake.

expressions). A coding scheme developed by Moll et al. (2016, 2017) listing expressions such as smirking, brow furrowing, lip biting, and making a fist was used and treated as changeable and non-exhaustive, with the rater changing and adding expressions based on her observations (see Appendix B for the final coding scheme). Expressions were coded only if they were absent during baseline, defined as the initial 7 s of the trial. A second rater, also unaware of group and condition, made the same judgments for 25% (12) of the children. Inter-rater reliability was excellent for the absence/presence of expressions (95% agreement, κ = .90) and very good for the number of expressions (83% agreement, κ = .76). Disagreements were resolved by discussion.

Standard Appearance–Reality Task. An independent rater listened to children's answers without knowing what children had been asked. First, the rater determined whether the child said "book," "box," "crayon," or "candle" (or near synonyms, e.g., "pen" for "crayon"). Second, it was revealed to the rater what children had been asked, allowing her to classify the answer as correct (1) or incorrect (0). No answers (39 trials) and unintelligible answers (10 trials) were scored as 0. If a child answered "both" (4 trials), a score of 0.5 was given. To assess inter-rater reliability, a second rater scored answers from > 25% of the children (8 from each group) with the same procedure. Inter-rater reliability was very good (κ = .90). Disagreements were resolved by discussion.

Results

Expression Task

A total of 175 expressions were observed, the most common of which were smirking (49), lip biting (19), brow raising (18), and brow furrowing (16). Generalized linear models (GLMs) found no effects of age, gender, story, or trial on the presence or number of expressions, respectively (*ps* > .12).

To investigate whether the absence or presence (binary outcome) or the number (continuous outcome) of children's expressions differed between groups and conditions, a logit generalized linear mixed-effects model (GLMM) with group as a between-participants predictor and condition as a within-participants predictor was run. For the *presence* of expressions, the interaction between group and condition approached significance (β = .99, p = .07). Children who had the opportunity to build experiential records expressed more suspense in the Unwitnessed condition than in the Witnessed condition (β = .92, p < .05), whereas children who could not build such records did not distinguish between conditions (β = −.08, p = .85) (see Fig. 3). There was no significant main effect of group (β = .11, p = .73) or condition (β = .41, p = .13).

For the *number* of expressions, the interaction between group and condition was significant (β = .77, p < .05). As shown in Fig. 4, children in the ER group showed more expressions in the Unwitnessed condition than in the Witnessed condition (β = .69, p < .01). In the No-ER group, children showed an equal number of expressions between conditions (β = .08, p = .73). There was no significant effect of group (β = .22, p = .29), but there was a significant effect of condition (β = .34, p < .03).

Standard Appearance–Reality Task

When first asked what the objects were, most children responded "book" (68%) and "crayon" (72%), respectively, indicating that the objects were deceptive. A GLM predicting children's scores showed no effects of age, gender, trial, or the order in which the labels were mentioned in the questions (e.g., "book" vs. "box" first) (*ps* > .19). Table 1 shows how many children gave a certain number of correct answers in response to the two types of question (AQ and RQ).

Three scores were calculated for each child: (a) a sum score across all four test questions (0–4), (b) a "reality score" by adding scores from the two reality questions (0–2), and (c) an "appearance score" by adding scores from the two appearance questions (0–2). On average, children reached a sum score of 1.47 (*SE* = 0.13), a reality score of 0.78 (*SE* = 0.09), and an appearance score of 0.68 (*SE* = 0.09). Wilcoxon signed rank tests (which were chosen over one-sample *t* tests due to skewedness, kurtosis = −.70) showed that children's correct answers did not exceed chance (set at 2 for the sum score and at

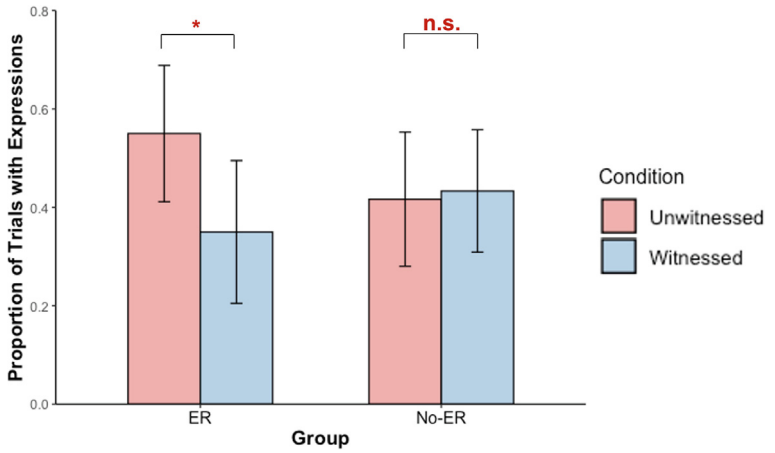

Fig. 3. Presence of expressions broken down by condition and group. ER, Experiential Record; No-ER, No Experiential Record; n.s., not significant. *p < .05.

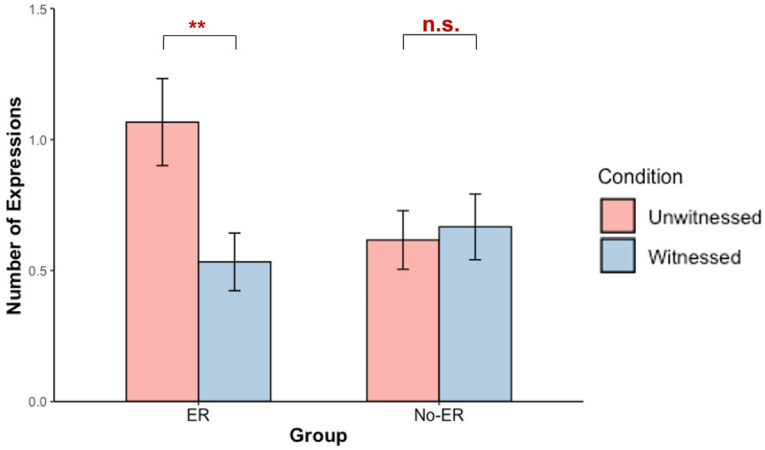

Fig. 4. Number of Expressions as a Function of group and condition in Experiment 1. ER, Experiential Record; No-ER, No Experiential Record; n.s., not significant. $^{**}p < .01$.

Table 1
n (and %) of children with a given combination of scores on the two types of test question.

Appearance question	Reality question			
	0	0.5	1	2
0	11 (18.33%)	1 (1.67%)	8 (13.33%)	7 (11.67%)
0.5	–	1 (1.67%)	–	1 (1.67%)
1	9 (15.00%)	–	11 (18.33%)	2 (3.33%)
2	3 (5.00%)	–	5 (8.33%)	1 (1.67%)

Note. If children responded "both," they received a score of 0.5.

1 for the reality and appearance scores, respectively) (Bonferroni-adjusted *ps* > .99). This did not change when unintelligible and missing answers were not counted as incorrect but instead were removed from the analyses (Bonferroni-adjusted *ps* > .31).

Relation between Expression Task and Standard Appearance–Reality Task

To test whether children's expressions and their answers in the Standard Appearance–Reality task correlated, we derived an "ability score" for children in the ER group of the Expression Task. This score was calculated by summing up the number of expressions from both trials of the Unwitnessed condition, summing up the number of expressions from both trials of the Witnessed condition, and then subtracting the latter sum from the former sum (Σ expressions in the Unwitnessed condition $-$ Σ expressions in the Witnessed condition). There was no association between children's responses in the Expression Task and those in the Standard Appearance–Reality Task (Spearman's $\rho = .14$, $p = .47$).

Discussion

In this experiment, 3.5-year-olds expressed more tension when an agent approached a scene not knowing that her favorite food item had been replaced with a fake lookalike than when she had witnessed the exchange. No difference between witnessed and unwitnessed demonstrations was observed when children watched an agent approach a deceptive object without having previously engaged with a real analog that was no longer accessible.

The findings support the middle position's claim that toddlers rely on records of others' experiences (experiential records), which allow them to recognize when agents fail to reconnect with their

objects (Perner & Roessler, 2012; Roessler & Perner, 2013). The method of measuring expressions also reveals that toddlers anticipate the affective impact of having one's expectations violated. The experiment supports the middle position's claim about the distinctness of basic early-emerging versus complex later-developing theory of mind skills, which the two-systems account stresses (Apperly & Butterfill, 2009; Low & Watts, 2013). Although the 3.5-year-olds in this experiment skillfully tracked agent–object relations (expressing "Level 1" perspective knowledge), they lacked awareness that the agent was deceived by the fake object's appearance. This is reflected in the negative data from the No-ER group as well as children's low performance in the Standard Appearance–Reality Task. The absence of a positive relation between the Expression Task and the Standard Appearance–Reality Task echoes earlier reports of a dissociation between indirect tests and classic direct test measures (Clements & Perner, 1994; Moll et al., 2016); it also supports the middle position's hypothesis that toddlers' theory of mind involves different cognitive processes than the later-developing deeper understanding of the mind (Apperly & Butterfill, 2009).

Taken as a whole, the data suggest that children in the ER group expressed more tension in the Unwitnessed condition than in the Witnessed condition not because they realized that the agent was fooled by the fake food but more simply because they knew that the agent would be disappointed to find her object gone and replaced. In Experiment 2, we sought further support for this interpretation by comparing children's expressions when an agent returning to fetch her food encountered either a fake lookalike (Deceptiveness condition) or an altogether different non-deceptive object (Non-Deceptiveness condition).

Experiment 2

This experiment, which was preregistered on the Open Science Framework (https://osf.io/te4va), investigated whether 3.5-year-olds' expressions differ when an agent approaches a deceptive object versus a non-deceptive object. There were two conditions. In the Deceptiveness condition, the procedure was identical to that in the Unwitnessed condition of the ER group from Experiment 1. In the Non-Deceptiveness condition, the agent's food item was replaced with an object that looked nothing like food (e.g., a toy).

Expressions were expected in both conditions because, as Experiment 1 and prior work have shown, children respond with expressions when an agent returns to where she left but will no longer find her object. Thus, no binary measure of expressions (absence vs. presence) was taken; instead, the *number* and *timing* of expressions were measured. If, as Experiment 1 suggests, 3.5-year-olds do not acknowledge how an object's (deceptive) appearance affects an agent's representation of it, then children's expressions should be identical in number and time regardless of whether an agent's food was replaced by a deceptive or non-deceptive object. If, by contrast and against Experiment 1's findings, children aged 3.5 years detect deceptiveness, then they might show more frequent and temporally extended expressions in the deceptive case. This is because, unlike a non-deceptive object, a deceptive object *sustains* the agent's belief that the object she is approaching is the same (edible) object she had left there earlier.

Method

Participants

To match statistical power to that of Experiment 1, a sample size of $N = 30$ 3.5-year-olds (15 girls; $M = 42.49$ months, range = 36.03–48.00) was chosen. One additional child was tested but did not finish the procedure due to uncooperativeness.

The same recruitment method as in Experiment 1 was used. All participants were English speakers. The ethnoracial breakdown of the final sample was 17% multiracial, 3% African American, 7% Asian, 70% White, and 3% "other," with 20% of the children being Latinx. The families' household incomes, again ranging from <$20,000 to >$120,000, indicated widely varying socioeconomic backgrounds.

Materials

The same deceptive objects from Experiment 1's Expression Task were used. In the Non-Deceptiveness condition, four additional non-deceptive objects were an orange rubber toy ($8 \times 8 \times 1.5$ cm), a yellow wooden block ($9 \times 2.8 \times 1.5$ cm), a pile of pink foam cubes (each one's side length was 3.8 cm), and a blue spiky rubber ball (5 cm in diameter). Each non-deceptive object was matched with a deceptive object and appeared in the same story (toy = chocolate story, block = cupcake story, cubes = cherry story, and ball = ice cream story).

Design

Children watched the same four puppet show stories as in Experiment 1. Each story came in two versions. In the Deceptiveness condition, B replaced A's food item with an inedible lookalike, as in the ER group of Experiment 1. The Non-Deceptiveness condition differed in that B replaced A's food item not with a lookalike but instead with a non-deceptive object. In neither condition did A witness the replacement, with the effect that A returned with a false expectation of encountering her food item. Each child received the Deceptiveness condition of two stories and the Non-Deceptiveness condition of the other two stories. Story order and condition order were counterbalanced.

Procedure

The procedure of the Deceptiveness condition was the same as in the Unwitnessed condition of Experiment 1's ER group, with the only difference that A, when seeing the new object upon her return, said "Oooh, look!", paused 2 s, and stated "All right, let me go get that." The reason for this change was that the same expression needed to suit both conditions ("Oooh, look" expresses joyful anticipation in the Deceptiveness condition and recognition of the presence of a different object in the Non-Deceptiveness condition). The procedure of the Non-Deceptiveness condition was the same as in the Deceptiveness condition with the difference that B exchanged the food not with an inedible lookalike (e.g., fake cupcake) but with an entirely different object (e.g., yellow rectangular block). In both conditions, B demonstrated the new object's inedibility as in Experiment 1.

Scoring and reliability

The same coding scheme and scoring rules were applied as in Experiment 1. The rater added a time stamp to each expression, indicating when in the response phase the expression occurred. A second rater made the same judgments for 25% of children from each condition. Inter-rater reliability was good for the number of expressions (agreement on 81% of trials, $\kappa = .66$) and was excellent for the time stamps ($\kappa = .88$). Disagreements were resolved by discussion.

Results

A total of 83 expressions was observed, the most common of which were smirking (20), raising the brow (15), lip biting (13), and furrowing the brow (6). GLMs examining potential effects of age, gender, story, and trial on the number or timing of expressions found no such effects ($ps > .44$).

A GLMM with the number of expressions as the dependent variable and condition as a predictor showed no difference in the number of expressions when children watched an agent approach a deceptive object compared with a non-deceptive object ($\beta = .02$, $p = .91$), as shown in Fig. 5.

We compared the mean latency of children's expressions using a one-way analysis of variance (ANOVA) with condition as the independent variable and time of expression as the dependent variable. As Fig. 6 shows, the timing of children's expressions did not differ between conditions, $F(1, 81) = 2.31$, $p = .13$.

Discussion

In this experiment, we looked for signs that 3.5-year-olds detect when an agent is misled by deceptive appearances. Consistent with Experiment 1, no such signs were found. Children's expressions were the same in number and timing regardless of whether an agent approached a deceptive or non-deceptive object. If children had realized that the agent would misidentify the object in the

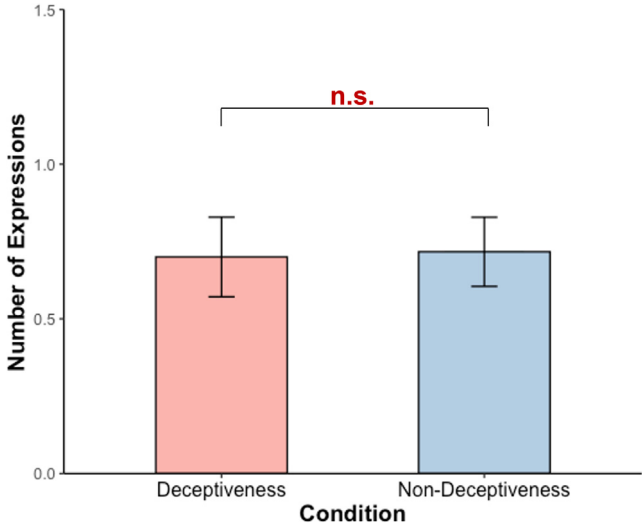

Fig. 5. Mean number of expressions as a function of condition in Experiment 2. n.s., not significant.

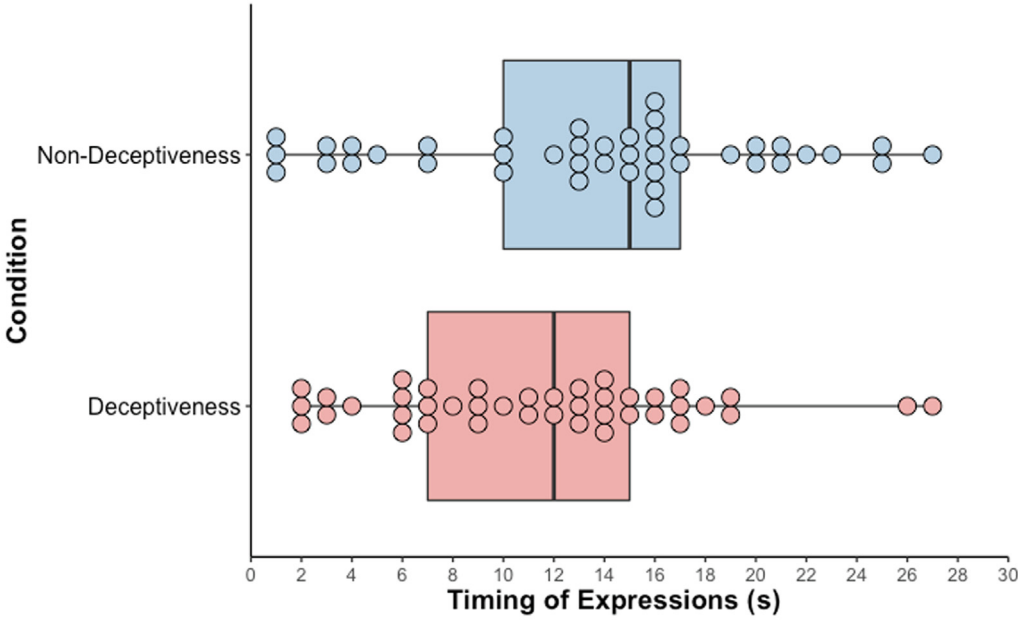

Fig. 6. Mean latency (in seconds) of expressions in Experiment 2 broken down by condition.

Deceptiveness condition, they might have expressed such understanding through more frequent and delayed expressions compared with the Non-Deceptiveness condition. The agent's perception of an object that is visually indistinguishable from the food she left there moments ago would support her assumption that she is approaching her food item—evoking in children more expressions as the agent comes closer to the object. By contrast, seeing a toy that looks nothing like food (in the

Non-Deceptiveness condition) would make the agent's expectation of food fall instantly to the floor, leading to fewer expressions as the agent nears the noticeably different object.

Ideally, a study with older children who understand how deceptive objects affect object representations would be conducted to confirm the predictions about the frequency and timing of expressions. A pilot study we ran suggested that the pace and perhaps content of the stories would need to be adjusted for older children's attention to be adequately engaged. Although this experiment provides no contrast to older children, its null findings, in conjunction with the data from Experiment 1, are consistent with the view that toddlers anticipate an agent's surprise of finding her object missing without being aware that the agent misidentifies the new object with the old object.

General discussion

The current study offers new evidence for the view that children younger than 4 or 5 years have theory of mind skills that are limited in systematic and predictable ways. We tested children just below this age threshold with a sensitive expression-based measure to determine the degree to which these children understand how changes in the world affect an agent's perspective on a situation.

In Experiment 1, 3.5-year-olds expressed tension when an agent approached a scene in which her food item was replaced with something inedible. The ersatz object was visually indistinguishable from the original object, with the effect that the agent would mistake the new object for the original food item. The children, however, only seemed to understand that the agent could not reconnect with her object without acknowledging that the agent would mistake the new object for the old object. This is suggested by the negative findings from the No-ER group, in which children saw the agent approach the deceptive object without prior interaction with another object. In this scenario, children did not show more expressions than in a control condition where the agent knew of the object's fakeness—again suggesting that children were unaware of the object's deceptive quality. The conclusion that 3.5-year-olds do not recognize that appearances can mislead was further substantiated in Experiment 2, where 3.5-year-olds responded in the same way when an agent approached a deceptive object as when she approached a non-deceptive object.

The pattern of findings supports the middle position. It suggests that, as Perner and Roessler (2010) argued, children under 4 or 5 years of age rely on records of others' experiences (experiential records) as they make sense of others' reactions to situations. With these records, toddlers can anticipate the return of an agent where she left an object or, in our study, foresee her emotional response to a salient change of reality. However, what experiential records do not provide, and thus what children under 4 or 5 years of age do not possess, is an understanding that appearances can induce false assumptions.

The insight that toddlers' theory of mind stands and falls with records of others' experiences helps to explain why numerous indirect tests use the location-change procedure but almost no indirect tests use the unexpected-content or appearance–reality procedure. Most looking time (He et al., 2011, Onishi & Baillargeon, 2005; Surian & Geraci, 2012), anticipatory looking (Clements & Perner, 1994), helping (Buttelmann et al., 2009), and other paradigms (Rubio-Fernández & Geurts, 2013) rely on the location-change procedure because it enables children to draw on records of the agent's prior interactions when anticipating her next action steps. In their original versions, the unexpected-content and appearance–reality tasks provide no opportunity for experiential record building. They require the child to understand that a deceptive container or object, seen for the first time, induces false expectations of its content or identity. Even if experiential records are made available by incorporating object exchanges or relocations, the data remain negative. This was observed in our Experiment 2 and in Low and Watts' (2013) study, in which 3- and 4-year-olds did not anticipate that an agent would erroneously believe that an object she was encountering was a different entity than an object she saw earlier from a different vantage point. The results match those from direct tests that likewise indicate that children understand identity relations no earlier than they solve classic theory of mind problems (Perner et al., 2011; Sprung et al., 2007).

Three published studies reported positive looking time data with infants using identity tasks. One involved a container designed to make an agent falsely assume that it contained her plush toy (with plush sticking out from underneath the container's lid, Song & Baillargeon, 2008). But the results can

be accounted for by records of where the agent last experienced the plush toy, rendering any conceptual understanding of the relation between appearance and object identification superfluous. A similar reductive explanation was given (Apperly & Butterfill, 2009) for the second study, in which toddlers tracked an agent's interaction with two penguin toys (Scott & Baillargeon, 2009). In the third study, infants allegedly realized that an agent would mistake a silent and visually identical toy (Deceptive condition), but not a silent and visually distinct toy (Non-Deceptive condition), for her rattling toy (Scott, Richman, & Baillargeon, 2015). Problematically, however, the authors inferred such deep understanding from longer looks in response to the agent's encounter of the non-deceptive toy compared with the deceptive toy. The longer looks can, however, be equally well explained by infants' encoding a new encounter between an agent and a new distinct object. Inferring that infants recognized the agent's misidentification of the object in the deceptive case does not seem warranted.

Our study is a reminder that studies presented as evidence of full belief understanding during infancy might in actuality have tested for something simpler. Priewasser et al. (2018) explained how infants can pass Buttelmann et al.'s (2009) helping study by inferring different action goals—not beliefs—in the two conditions. Likewise, infants' pointing gestures in Knudsen and Liszkowski's (2012a, 2012b) "false belief condition" are more parsimoniously viewed as attempts to update the recipient's information rather than to correct a prior belief. This does not mean that the word "belief" needs to be banned altogether from the description of what toddlers reason about in these studies. The problem with classic conceptual development theory is that it cannot account very well for the belief-related capacities toddlers displayed in the ER group of our Experiment 1 and in other indirect tests. Different variants of the middle position offer different solutions to indicate that toddlers are sensitive to reality-incongruent beliefs while failing to fully acknowledge beliefs. Apperly and Butterfill (2009) proposed to speak of toddlers as tracking beliefs by representing not beliefs but simpler relational analogs called "registrations." Perner and Roessler (2010) argued that toddlers have implicit knowledge of beliefs as part of their teleological understanding of human action. Moll et al. (2022) spoke of an early practical grasp of beliefs that manifests in skilled (inter)action rather than in correct judgment. What these variants of the middle position share is the assumption that an understanding of beliefs neither is present in mature form during infancy nor emerges through the acquisition of a belief concept at 4 or 5 years of age. Instead, it develops in steps, and one of the later steps, taken at 4 or 5 years of age, is the realization that appearances can induce false beliefs about objects and their identities.

Future studies should further delineate the extent of toddlers' theory of mind abilities. One unanswered question concerns the kinds of interaction toddlers need to witness to detect others' informational states. Butterfill and Apperly (2013) postulated that toddlers need to witness an agent *physically encountering* (seeing and/or touching) an object to recognize the agent as having registered the object. Perner and Roessler's (2010, 2012) term "experiential record" likewise implies that direct perceptual experience is necessary for toddlers to recognize the agent's epistemic relation to an object. To test this prediction, one could compare situations in which an agent gains firsthand perceptual knowledge of where or what an object is with situations in which the agent acquires the same knowledge secondhand through another's testimony. Would the latter case allow toddlers to recognize the agent's new informational state? And can children inferentially "piece together" records of experience, for example, when seeing an agent looking inside a box whose content is only later revealed to the children themselves? The middle position hypothesizes that unless toddlers witness a direct perceptual encounter between an agent and an object, they do not grasp that an agent holds information or expectations about the object. More research is needed to establish whether these are in fact conditions that need to be met for toddlers to detect others' informational states.

To conclude, the current findings favor a middle position between nativism and classic conceptual change theory. This position rejects the all-or-nothing character of the emergence of a theory of mind that the two alternatives defend, each in their own way. In line with the middle position, our study supports the idea that toddlers have *some* grasp of beliefs before recognizing the possibility of misidentification. This is revealed by the expressions children show when tracking an agent whose expectations (recently built up through perceptual experiences) are about to be violated. As Roessler and Perner (2013) argued, toddlers skillfully track others' interactions with objects across time, building records of experience that let them anticipate how agents will respond to relevant changes in the world (such as object removals or replacements). Classic conceptual development

theory has not sufficiently acknowledged children's early responsiveness to these belief-involving cases (but see Perner et al., 1994). At the same time—and this is the crucial piece that nativism misses—toddlers' theory of mind abilities are significantly limited. The limits become obvious when toddlers do not have experiential records available or when these records are of no use because general knowledge of the links among perception, belief formation, and action is required. When watching an agent naïvely heading toward a deceptive object, children under 4 years of age seem unaware that the agent is misled by the object's deceptive quality. They have yet to become aware that appearances shape one's perception and how that perception informs one's beliefs and actions.

Acknowledgments

This research was supported by USC's Department of Psychology Doctoral Research Grant Award for 2019-2021. We would like to express our gratitude to the children and families who participated in this research. We are also grateful to Bella Fascendini for her invaluable assistance with coding and scoring. Additionally, we would like to thank the members of the USC 'Minds in Development' lab for their insightful ideas and valuable feedback throughout the course of the research.

Appendix A. Puppet show stories in Experiment 1

Chocolate story

Maxi (A) entered the stage with a real chocolate bar (ER group) or empty-handed (No-ER group), saying "Hi, I'm Maxi. I love chocolate! Do you like chocolate? I eat chocolate all the time!" In the ER group, Maxi held the chocolate, bit into it, and ate a piece of the chocolate piece, stating "It's so yummy!" He then placed the chocolate bar in the right corner of the stage and exited. Depending on the condition, Maxi then either returned after about 3 s and positioned himself on the dowel (Witnessed condition) or did not return until later (Unwitnessed condition). In both conditions, Bob (B) entered with a fake chocolate bar that looked exactly like the chocolate bar. Bob demonstrated that the fake chocolate is inedible by bending it back and forth, stating "This is not tasty at all. It is rubbery, look! You cannot eat this!" Next, Bob replaced the real chocolate with the fake one (ER group) or placed the fake chocolate in the right corner of the stage (No-ER group) and left. After 3 s, Maxi stood up (Witnessed condition) or came back (Unwitnessed condition). Looking at the fake chocolate, he exclaimed "Oooh, alright – let me go get that!" and started slowly walking toward it. When Maxi touched it, the bell rang and the curtain closed.

Cupcake story

Cindy (A) entered the stage with a real cupcake (ER group) or empty-handed (No-ER group), saying "Hi, I'm Cindy. I love cupcakes! Do you like cupcakes? I eat cupcakes all the time!" In the ER group, Cindy held the cupcake and licked the cream on top, stating "It tastes so good!" She placed the cupcake in the right corner of the stage and exited. Cindy then either returned 3 s later and positioned herself on the dowel (Witnessed condition) or did not return until later (Unwitnessed condition). In both conditions, Daisy (B) brought a deceptive cupcake that looked the same as the real one. Turning the deceptive cupcake upside down, shaking it, and knocking it against the stage, Daisy exclaimed "This is not yummy. It is hard like wood. See? You cannot eat this!" Daisy then replaced the real cupcake with the fake one (ER group) or placed the fake cupcake in the right corner of the stage (No-ER group) and exited. After 3 s, Cindy got up from the dowel (Witnessed condition) or returned to the stage (Unwitnessed condition). Focusing on the fake cupcake, Cindy said "Oooh, alright – let me get this now!" and walked slowly toward it. Lastly, Cindy fetched the fake cupcake, the bell rang, and the curtain closed.

Cherry story

Emma (A) entered the stage with a handful of real cherries (ER group) or nothing (No-ER group), stating "Hi, I'm Emma. I love cherries! Do you like cherries? I eat cherries all the time!" The ER group watched Emma eating a cherry, saying "Oh, so juicy!". Emma placed the cherries in the right corner of the stage and left. Depending on the condition, Emma returned 3 s later and sat down on the dowel in the left corner (Witnessed condition) or did not return until later (Unwitnessed condition). Next, Fiona

(B) entered with a handful of fake cherries that looked exactly like the real ones. Fiona opened a fake cherry and showed its foamy inside, exclaiming "These things aren't tasty and sweet. Look, they're foamy! You cannot eat these!" She then replaced the real cherries with the fake ones (ER group) or placed the fake cherries in the right corner of the stage (No-ER group) and left. Emma stood up from the dowel (Witnessed condition) or came back to the stage (Unwitnessed condition) 3 s later. She looked at the fake cherries and said "Oooh, alright – now I'm going to get those!" She slowly approached the fake cherries. When she touched the fake cherries, the bell rang and the curtain was closed.

Ice cream story

Grace (A) entered the stage with a real ice cream bar (ER group) or nothing (No-ER group), stating "Hi, I'm Grace. I love ice cream! Do you like ice-cream? I eat ice cream all the time!" In the ER group, Grace licked the ice-cream bar, stating "This is delicious!" before leaving the ice-cream bar in the right corner of the stage. She excused herself and left. Grace then either returned 3 s later and sat down on the dowel in the left corner (Witnessed condition) or did not return until later (Unwitnessed condition). Next, Hank (B) appeared with a fake ice cream bar that looked identical to the real one. He demonstrated its inedibility by squeezing it and said "It's not tasty at all, it's squishy! See? You cannot eat this thing!" Hank then switched the fake and real ice cream bars (ER group) or placed the fake ice cream bar in the right corner (No-ER group). Next, Grace stood up (Witnessed condition) or returned (Unwitnessed condition). Looking at the fake ice cream, she stated "Oooh, alright – now I'm going to get that!" She walked with a slow pace toward the fake ice cream. When she made contact with it, the bell rang and the curtain closed.

Note. A, Puppet A; B, Puppet B; ER, Experiential Record; No-ER, No Experiential Record.

Appendix B. Coding scheme

Facial expressions

Lip bite: bites on the lower lip, with teeth visible or not.
Lip pout: lower lip curls down, with or without "orange peel chin" appearing.
Lip curl/Lip tension: upper and/or lower lip tightens, or upper lip curls up.
Furrowed brow: eyebrows lower creating vertical lines and other wrinkles on the forehead.
Raised brow: eyebrows move upward, creating horizontal lines across the forehead.
Mouth opens or closes: mouth suddenly opens or (if the baseline is an open mouth) suddenly closes.
Mouth tightening: muscles around mouth stretch and tighten, with lips slightly thinning.
"Fish mouth": pulls lips together to create a fish mouth.
Smirk: smiles with a smirk.
Hand(s) over/inside mouth: hands are brought to or inside the mouth.
Hand(s) over ears: hands are brought to ears.
Hand(s) out of mouth: hand moves out of mouth area suddenly and mouth appears (if hand covered mouth at baseline).
Foot inside mouth: foot is placed inside mouth.
Head shake: moves head left and right quickly.

Bodily expressions

Body tightening: bringing shoulders near chest while rising in chair.
Pointing: gesturing excitedly toward the stage.
Raised hand: raises one/two hands, similar to a hand gesture of stop.
Bouncing: moves body up and down or in a rocking motion, arms/legs may move as well.
Shrugging shoulders: shoulders quickly move up and down.
Neck tension: shows sudden neck muscle tightness with head slightly moving forward or backward.
Standing up: child suddenly stands up from chair.
Sitting up: changes from slouched position to sitting up straight.
Pushing back into chair: child pushes body back into the chair.

References

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.

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006.

Bar-On, D. (2013). Expressive communication and continuity skepticism. *Journal of Philosophy, 110*(6), 293–330. https://doi.org/10.5840/jphil2013110624.

Bar-On, D. (2017). Communicative intentions, expressive communication, and origins of meaning. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds.* Routledge. https://doi.org/10.4324/9781315742250.

Bühler, K. (1990). *The Theory of Language: The Representational Function of Language (Sprachtheorie).* Amsterdam: John Benjamin's Publishing Company. (Original work published 1934).

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*(5), 606–637. https://doi.org/10.1111/mila.12036.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development, 46*, 12–30. https://doi.org/10.1016/j.cogdev.2018.01.001.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance–reality distinction. *Cognitive Psychology, 15*(1), 95–120. https://doi.org/10.1016/0010-0285(83)90005-1.

Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of knowledge about the appearance–reality distinction. *Monographs of the Society for Research in Child Development, 51*(1), 1–87. https://doi.org/10.2307/1165866.

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation-of-expectation, change-of-location and unexpected-contents tasks. *Developmental Science, 14*(2), 292–305. https://doi.org/10.1111/j.1467-7687.2010.00980.x.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567–582. https://doi.org/10.2307/1130337.

Knudsen, B., & Liszkowski, U. (2012a). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy, 17*(6), 672–691. https://doi.org/10.1111/j.1532-7078.2011.00105.x.

Knudsen, B., & Liszkowski, U. (2012b). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science, 15*(1), 113–122. https://doi.org/10.1111/j.1467-7687.2011.01098.x.

Krupenye, C., & Call, J. (2019). Theory of mind in animals: Current and future directions. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(6), 1–25. https://doi.org/10.1002/wcs.1503.

Darwin, C. (2002). *The expression of the emotions in man and animals.* Oxford University Press. (Original work published 1872).

Kulke, L., & Hinrichs, M. A. B. (2021). Implicit theory of mind under realistic social circumstances measured with mobile eye-tracking. *Scientific Reports, 11*(1), Article 1215. https://doi.org/10.1038/s41598-020-80614-5.

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child Development, 85*(4), 1519–1534. https://doi.org/10.1111/cdev.12224.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science, 24*(3), 305–311. https://doi.org/10.1177/0956797612451469.

Moll, H., Kane, S., & McGowan, L. (2016). Three-year-olds express suspense when an agent approaches a scene with a false belief. *Developmental Science, 19*(2), 208–220. https://doi.org/10.1111/desc.12310.

Moll, H., Khalulyan, A., & Moffett, L. (2017). 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development, 88*(1), 114–122. https://doi.org/10.1111/cdev.12581.

Moll, H., Ni, Q., & Stekeler-Weithofer, P. (2022). Ontogenetic steps of understanding beliefs: From practical to theoretical. *Philosophical Psychology.* Advance online publication. https://doi.org/10.1080/09515089.2022.2073211.

Onishi, K. K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. https://doi.org/10.1126/science.1107621.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 261–286). Lawrence Erlbaum.

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science, 333*(6041), 474–477. https://doi.org/10.1126/science.1201216.

Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children's theory of mind. In J. H. Aguilar & A. A. Buckareff (Eds.), *Causing human actions: New perspectives on the causal theory of action* (pp. 199–228). MIT Press. https://doi.org/ https://doi.org/10.1080/09515089.2012.681869.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*(5719), 214–216. https://doi.org/10.1126/science.1111656.

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.

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or teleology? *Cognitive Development, 46*, 69–78. https://doi.org/10.1016/j.cogdev.2017.08.002.

Roessler, J., & Perner, J. (2013). Teleology: Belief as perspective. In S. Baron-Cohen, S. H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 35–51). Oxford University Press.

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*(1), 27–33. https://doi.org/10.1177/0956797612447819.
Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x.
Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012.
Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology, 82*, 32–56. https://doi.org/10.1016/j.cogpsych.2015.08.003.
Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences of the United States of America, 113*(47), 13360–13365. https://doi.org/10.1073/pnas.1609203113.
Song, H.-J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*(6), 1789–1795. https://doi.org/10.1037/a0013774.
Sprung, M., Perner, J., & Mitchell, P. (2007). Opacity and discourse referents: Object identity and object properties. *Mind & Language, 22*(3), 215–245. https://doi.org/10.1111/j.1468-0017.2007.00307.x.
Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x.
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., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E.
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/10.1002/app.25226.