DOI: 10.1111/cdev.14073

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

The cognitive underpinnings and early development of children's selective trust

Benjamin Schmid[1,2] | **Natalie Bleijlevens**[1,2] | **Nivedita Mani**[2,3] | **Tanya Behne**[1,2]

[1]Developmental Psychology, University of Göttingen, Göttingen, Germany
[2]Leibniz Science Campus "Primate Cognition", Göttingen, Germany
[3]Psychology of Language, University of Göttingen, Göttingen, Germany

Correspondence
Tanya Behne, Department of Developmental Psychology, University of Goettingen, Waldweg 26, 37073 Goettingen, Germany.
Email: tanya.behne@psych.uni-goettingen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 254142454/GRK 2070

Abstract
Young children learn selectively from reliable over unreliable sources. However, the cognitive underpinnings of their selectivity (attentional biases or trait ascriptions) and its early ontogeny are unclear. Thus, across three studies ($N = 139$, monolingual German speakers, 67 female), selective-trust tasks were adapted to test both preschoolers (5-year-olds) and toddlers (24-month-olds), using eye-tracking and interactive measures. These data show that preschoolers' selectivity is not based on attentional biases, but on person-specific trait ascriptions. In contrast, toddlers showed no selective trust, even in the eye-tracking tasks. They succeeded, however, in eye-tracking tasks with the same word-learning demands, if no ascriptions of reliability were required. Thus, these findings suggest that preschoolers, but not toddlers, use trait-like ascriptions of reliability to guide their selective learning.

Most of what we know we have learned through the testimony of others. Given that not all potential sources of testimony are equally reliable, there is a strong need to determine whom to learn from. This is especially true in the case of the acquisition of conventional knowledge, where the novice learner cannot determine the reliability of the content by herself. Here, children and other novice learners need to be able to determine the reliability of the source in order to evaluate the novel content provided. Thus, in the case of word learning, for example, learners need to be selective in terms of whom to trust.

Much recent research has demonstrated that young children are indeed selective in whom they learn novel words from. For example, preschoolers learn novel labels selectively from a confident rather than an uncertain speaker, a knowledgeable rather than an ignorant one, and a previously accurate rather than a previously inaccurate one (see, e.g., Harris, 2012; Mills, 2013; Robinson & Einav, 2014). When presented with a choice of two speakers who provide conflicting information, children as young as 3–5 years selectively learn from the more trustworthy source (Koenig & Harris, 2005). While the robustness of preschoolers' selective trust is well established, unresolved debates concern (i) the cognitive underpinnings of children's selective trust and (ii) its early ontogeny.

Different proposals have been put forward regarding the cognitive underpinnings of children's selective trust (see, e.g., Birch et al., 2008; Heyes, 2017; Sabbagh & Shafman, 2009). Attentional bias accounts (e.g., Heyes, 2017) suggest that children's selective trust is based on low level, attentional differences. Children may develop an attentional bias toward the more reliable speaker, because her labeling of familiar objects matches children's visual input, whereas the unreliable speaker's utterances cause prediction errors. This may lead children to attend more to the reliable speaker and, as a consequence, they would be more likely to acquire the novel information provided by her.

In contrast, other accounts propose that children's selective trust builds on their conceptual knowledge (Sobel & Kushnir, 2013) including their trait reasoning (Hermes, Behne, & Rakoczy, 2018). This is in line with research showing that children's ability to explicitly

Abbreviations: AOI, area of interest; GLMM, generalized linear mixed model; PTL, proportion of target looking.

identify the more reliable sources correlates with their selective learning from such sources (Hermes et al., 2015; Koenig & Harris, 2005). According to these accounts, children's selectivity is not simply based on an attentional bias toward the more reliable source, but on their selective endorsement of the information provided by this source.

Both the attentional bias accounts and the selective endorsement accounts predict children's preferential learning from the more reliable speaker, raising the question of how the two can be tested for empirically. Research by Sabbagh and Shafman (2009) speaks to this question. In their study, 4- to 5-year-olds recalled the novel information provided by an ignorant speaker, but did not endorse it (see Sabbagh & Shafman, 2009). This would speak against attentional bias accounts and in favor of the selective endorsement of information provided by reliable sources. It is an open question, however, to what extent this finding is due to the specifics of their ignorance condition (with the speaker explicitly marking her ignorance, at the same time as providing the information) or whether it can be generalized more widely. Another test of the attentional bias account would be to examine its premise that children attend more to reliable speakers directly. To the extent that children's looking behavior can be considered a potential index of their attentional focus, testing for differences in children's looking behavior toward reliable and unreliable speakers may help to evaluate the attentional bias account. Research to date, however, has not yet assessed whether or not children look more toward reliable relative to unreliable speakers providing novel information and how this relates to their selective trust judgments.

The second debate concerns the early ontogeny of children's selective trust. Much research has demonstrated that preschoolers, from around age 4 on, show a robust preference for learning from reliable sources. That is, when presented with two speakers who present conflicting novel information, 4- to 5-year-olds will selectively endorse the information provided by the more reliable speaker. But what about younger children? Do infants and toddlers also selectively endorse the testimony from reliable speakers? A limited number of studies has explored this question and the findings of these studies are mixed.

Some research suggests that toddlers, or even infants, are sensitive to speakers' prior inaccuracy (Brooker & Poulin-Dubois, 2013; Crivello et al., 2018; Koenig & Woodward, 2010; Krogh-Jespersen & Echols, 2012; Kuzyk et al., 2020; see also Begus & Southgate, 2012 for related findings). In these one-informant studies, children aged 18–24 months were presented with *one* speaker who consistently labeled familiar objects either accurately or inaccurately. Subsequently, the same speaker taught the child a novel label. The toddlers mapped this novel label onto its referent at above chance levels in both conditions, but toddlers taught by the reliable speaker tended to outperform those taught by the unreliable speaker. Note that toddlers were tested with a one-informant design (i.e., each participant interacted either with a reliable or an unreliable speaker who provided novel information), rather than with the two-informant design (i.e., each participant interacted with a reliable and an unreliable speaker who provided conflicting novel information) that has been used to document preschoolers' selective trust.

In contrast to the standard two-informant design, it is less clear how to interpret the findings from the one-informant design. One possible interpretation is that participants adjust their epistemic trust to the reliability or unreliability of the specific speaker. Such an interpretation of children's responses as person-specific attributions of trustworthiness would imply that this task assesses the same competences as the standard two-informant design. However, there is an alternative interpretation of the findings from the one-informant design: Improved performance when encountering the reliable speaker may be moderated by the familiarity or predictability of the situation rather than by attributions of epistemic trust to a specific person. Children in the reliable-speaker condition experienced a familiar, everyday situation (with an adult labeling familiar objects correctly)—a situation that presumably ought not to affect their performance. Children in the unreliable-speaker condition, however, experienced quite an unusual, bizarre situation (with an adult consistently mislabeling objects)—a situation that might have affected their performance.

Thus, in case of the one-informant design we cannot distinguish whether children simply perform worse in bizarre rather than nonbizarre circumstances or whether they make person-specific attributions of trustworthiness. In contrast, due to the within-subject nature of the two-informant design, this ambiguity does not arise: All participants experience the bizarreness of the unreliable speaker mislabeling familiar objects. However, as their selective trust is assessed by testing *whom* they trust, it is possible to distinguish a distrust in the information from an unreliable speaker from an overall decrease in performance when encountering bizarre circumstances. Hence, the two-informant design may be considered a clearer test of children's selective trust (see also Hermes et al., 2019 for a discussion of the benefits and performance demands of the standard two informant design).

A couple of studies on young children's selective trust have employed a methodological variation that is of interest here. Children were first introduced to two speakers, a reliable and an unreliable one, and then, in a between-subject design, one of these speakers taught the child novel labels (e.g., Kim et al., 2017; Luchkina et al., 2018; Vanderbilt et al., 2014). So, just as in the one-informant studies, children were taught novel uncontested information by just one informant, but all participants had encountered both a reliable and an unreliable speaker beforehand. The findings obtained across studies present a puzzling pattern. Two studies with preschoolers suggest that in the absence of conflicting testimony, children endorse and generalize novel labels from reliable and unreliable speakers alike (Kim et al., 2017; Vanderbilt et al., 2014). In contrast, a preferential-looking study with infants suggests that 18-month-olds acquire labels for novel objects if taught by the reliable, not the unreliable, speaker (Luchkina et al., 2018, but see Krogh-Jespersen & Echols, 2012). It is unclear, whether this different pattern of results reflects genuine age changes—with infants demonstrating less trust in uncontested testimony from unreliable sources than preschoolers—or whether it reflects methodological differences (e.g., object choice vs. looking time measure). Given that the two-informant design has robustly shown that, when presented with conflicting information, preschoolers selectively trust the more reliable source, this seems to be an informative methodological approach to use with younger children, too.

Two studies have, in fact, adapted the two-informant to test children aged 2 and 3 years. Both these studies engaged participants in a hiding game and two speakers, a reliable and an unreliable one, provided conflicting information regarding the bait's hiding location. This research showed that 3-year-olds selectively endorsed the information provided by the more reliable speaker, but 2-year-olds did not (Ganea et al., 2011; Hermes et al., 2019). It is important to note, however, that the information the speakers provided in these games was episodic in nature (e.g., where the object had been hidden) rather than semantic (what an object is called). Conceptually, episodic errors are less informative than semantic ones, and research with older children has found that children are less likely to differentiate between speakers on the basis of episodic rather than semantic information (Stephens & Koenig, 2015). Thus, it is unclear what factors affected 2-year olds' poor performance in these hiding games. Did 2-year-olds have problems because children this age do not take speaker reliability into account? Or did they have problems because the task required them to make judgments based on episodic rather than semantic information? Given these methodological issues—and the problems of interpretation that arise—the early ontogeny of children's selective trust remains unclear. A new methodological approach is needed to help explore this issue: a version of the standard two-informant paradigm that is adapted for younger children and that explores participants' selective trust judgments on the basis of semantic, rather than episodic information.

The studies we present here aim to shed some light on both these issues, the cognitive underpinnings underlying children's selective trust judgments and the question whether children as young as 2 years selectively trust reliable over unreliable speakers in learning the labels of novel objects. Thus, 2- and 5-year-olds were presented with the standard two-informant selective trust design, implemented as an eye-tracking study. Using gaze data as the dependent variable lowered the task demands (e.g., Clements & Perner, 1994), helping us to adapt the selective trust task for the younger age group. Furthermore, in addition to the eye-tracking tasks, children also participated in interactive versions of the same tasks. The comparison between the two tasks allowed us to assess whether children would show an early implicit understanding of selective trust (as indicated by their looking behavior) that they may not yet be able to express in their more explicit action choices (as indicated by their responses in the interactive task; see, e.g., Hood et al., 2003 for a similar pattern of findings in other areas of cognitive development). Additionally, the eye-tracking set-up allowed us to assess differences in children's looking behavior to the reliable and unreliable speakers—as a potential index of differences in their attention to the two speakers.

Three studies were run. Study 1 explored the selective trust responses of both 2- and 5-year-olds. Two follow-up studies then focused on the 2-year-olds alone and explored whether the amount of input participants receive (Study 2) or the word learning demands involved (Study 3) might explain the pattern of 2-year-olds' performance.

STUDY 1

The novel eye-tracking method we developed was based on the interactive method used by Birch et al. (2008). In Birch et al. (2008), two speakers, a reliable and an unreliable one, used the same novel label (e.g., "toma") to refer to two different novel objects. Children's selective trust was then assessed using two tasks. In the *endorsement* task, the two novel objects that the speakers had labeled were presented and the experimenter asked children to select the "toma." Preschoolers demonstrated their selective trust by selecting the object that had been labeled by the reliable speaker. In the *contrast* task, the same pair of novel objects was presented, but this time the experimenter used another completely novel word and asked children to select the, say, "blicket." Following the logic of a disjunctive syllogism, preschoolers selected the object labeled by the unreliable speaker—indicating that they consider this object—not the one labeled by the reliable speaker—as the one without a known label (see, e.g., Halberda, 2006, for further discussion of the mechanism involved in this process of disambiguation). We adapted this method to assess 2- and 5-year-olds' selective trust, using eye-tracking and interactive versions of both tasks.

Method

Participants

Thirty-two monolingual 2-year-olds (M_{age} = 24.5 months, range = 23.2–25.8 months, 16 female) and thirty-six 5-year-olds (M_{age} = 65.5 months, range = 60.8–71.5 months, 18 female) were included in the final sample. Four additional children (two at each age) were tested but excluded from the final sample due to technical problems with the eye-tracker (n = 3) or the child choosing not to complete the session (n = 1). Participants were recruited from a database of German-speaking children whose parents had expressed interest in participating in child development studies. The participating children lived in a German university city and its surroundings. Further demographic data (race or ethnicity, income, education etc.) were not collected due to the local data protection rules. Parents gave informed consent for their child's participation. The Ethics committee of the University's Psychology Department approved this study (as well as Study 2 and 3, reported below, data for all studies were collected in 2017–2018). The sample size per age group was set to 32 based on recommendations for looking time data (see Csibra et al., 2016). We oversampled slightly when inviting participants, based on expectations of cancellations etc., and included all children tested. We tested 2-year-olds given the mixed findings regarding selective trust at this age and 5-year-olds given their competent selective trust decisions (Hermes, Behne, Bich, et al., 2018; Schütte et al., 2020).

Material and design

Children's gaze was recorded with a SMI Red250 eye-tracker attached to a 24-inch TFT flat-screen monitor with a display resolution of 1920 × 1080 pixels. The stimuli were presented using E-Prime 2.0. Prior to testing, we calibrated the participant's gaze using a 5-point calibration procedure, followed by a 4-point validation procedure.

Each child participated in the same sequence of events: a warm-up phase, a presentation phase (consisting of accuracy exposure, a novel label training and a brief reminder phase), and a test phase with first eye-tracking and then interactive tests. The stimuli for the presentation phase and the eye-tracking test phase were presented onscreen, with the eye-tracker providing data of participants' eye movements across the screen during these phases. For the presentation phase, video clips were prepared displaying two female actors on the left and the right side of the screen (Figure 1). Each clip showed one of the two speakers labeling an object that was presented as a photo at the center of the bottom half of the screen. For the eye-tracking test phase, two objects were shown on the left and right side of the screen together with the audio clip of a male voice asking children for the referent of the chosen labels. As an attention getter, a video of balloons to piano music was included at the beginning of the accuracy exposure, the novel label training and the eye-tracking test phase. For the interactive test phase, 3D objects were presented on a tray. The novel pseudo word "toma" was used by both speakers during the novel label training, to refer to one of two novel objects, either a green or a blue plush toy (each representing a microbe, see Figure 1). In addition, another two-syllable pseudo word, "blicket," was used in the test phase.

For each age group, the assignment of speaker (reliable or unreliable) to actor (actor A or B) was counterbalanced across participants. The same identical recordings were used for the two speaker roles. In the accuracy exposure phase, the difference in reliability arose by merging these recordings with the presentation of an object that either did, or did not, match the label used (see Figure 1).

Accuracy exposure

Children saw both speakers introducing themselves one after the other by saying hello and waving their hands. Then each speaker labeled the same four familiar items (baby, car, apple, dog), with the reliable speaker labeling each item correctly (saying, e.g., "Oh, look! This is an apple! An apple!"), and the unreliable speaker labeling each item incorrectly. The label the unreliable speaker used did not refer to the object shown on screen, but was one of the other familiar labels used by the reliable speaker during the study. Additionally, we made sure that the label she used did not correspond to the object presented to the child directly before or after the current object. To establish the speakers' consistent difference in labeling accuracy and to highlight the contrast between them, the first speaker labeled three familiar objects in turn, then the other speaker labeled those same three objects (in reversed order), followed by each speaker labeling the fourth object, one after the other (e.g., Schütte et al., 2020; see also Birch et al., 2008). For each age group, the order of speakers (reliable or unreliable first) and their location on the screen (left or right) were counterbalanced across participants, while both these factors (order and location of speakers) were fixed for each participant across the different experimental phases (e.g., accuracy exposure, label training etc.). The duration of each labeling event was 9 s.

Novel label training

Both speakers introduced the same novel label, "toma," but each referred to a different novel object (either the blue or the green plush toy). Each child saw two labeling events, one with each speaker (in blocked order). At the start of each labeling event, the novel object that the speaker would label appeared in the middle of the bottom half of the screen. Next, the speaker appeared and addressed the child saying, "Oh, look!", and then pointed

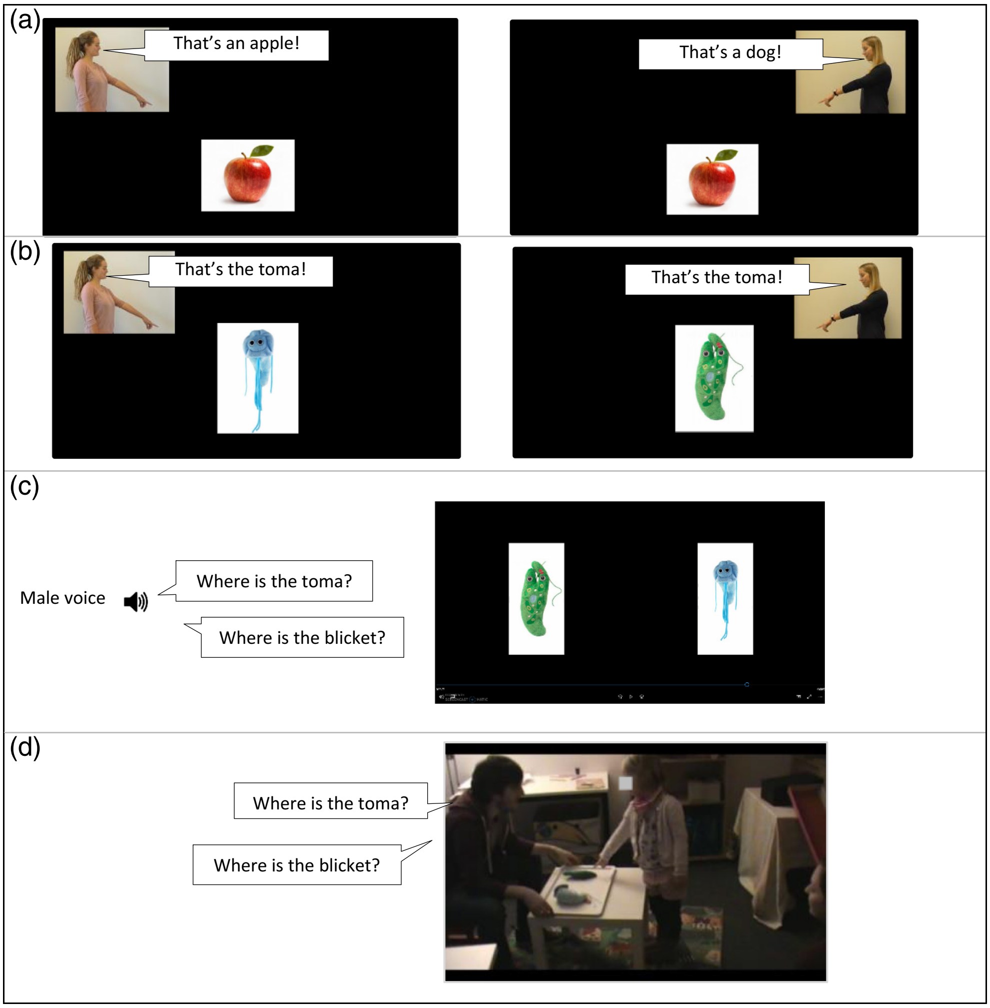

FIGURE 1 Schematic representation of the procedure. (a) Accuracy exposure; (b) novel label training; (c) eye-tracking test phase; (d) interactive test phase. (The speech presented is a translation from the German version used.)

at the novel object, saying, "That's the toma, the toma!" This sequence was then repeated for the second labeling event of the same speaker. This was followed by the two labeling events of the other speaker. Thus, each speaker used the novel label four times. The assignment of novel object (green or blue) to speaker (reliable or unreliable) was counterbalanced across participants within each age group. Duration of each labeling event was 8.5 s.

Reminder

To reduce demands on participants' memory, we then provided further information about the speakers' accuracy, with both speakers labeling one familiar entity (the picture of a baby) in accordance with their respective reliability. This was followed by a reminder of the novel label training, with one labeling event per speaker (i.e., each speaker labeling their respective novel object saying "That's the toma, the toma!"). Thus, in total during the novel label training and the reminder, children heard each speaker use the novel label six times. The reminder gave children the chance to evaluate the speakers' respective reliability after they had encountered the conflicting information the two speakers provided. Thus, even if participants' interest in speakers' labeling accuracy only arose at this stage, they would still be able to gather all the information needed to selectively endorse the more reliable source.

Eye-tracking test phase

Children participated in the Endorsement task and the Contrast task, with a block of two trials each per task (with the order of tasks counterbalanced across children within each age group). Each trial started with a fixation-cross to guide the child's gaze to the center of the screen. Immediately after that, the two novel objects, the blue and the green one, appeared on screen and then a male voice asked a test question. In the Endorsement task, the male speaker asked, "Where is the toma?" (i.e., the novel label that both speakers had introduced), whereas in the Contrast task, he used a different, completely novel word, asking "Where is the blicket?" The location of the objects was counterbalanced across the two trials of each task. The recording of the male voice was timed such that the objects were visible on screen for a preview time of 2.5 s before label onset. Total duration of these clips was 5 s. In addition to these two types of Novel object tasks, children also participated in a Familiar object task, consisting of four trials, one before each of the Novel object trials. For each Familiar object trial, a pair of familiar items (e.g., a boat and a baby) was presented and the male voice asked for one of them (e.g., "Where is the boat?"). Across the four trials, two pairs of items (either boat and baby, or house and car) were presented on two trials and the male voice asked for each of the four items on one of the four trials.

Interactive test phase

The experimenters and child then moved to a table in an adjacent room where E1 introduced the selection game. On each round, E1 presented two toy objects on a tray and specified which one of them should be placed in the chute. He demonstrated the game by presenting a pair of familiar objects and telling E2 which one to place in the chute. Children first engaged in two Familiar object trials. In each of these trials, a pair of familiar objects (a ball and rabbit on trial 1, a flower and a boat on trial 2) was presented and the child was told which one to place in the chute, "Where is the [ball/boat]?" and, if the child did not respond, "Could you put the [ball/boat] in the chute." For the endorsement task, E1 presented the blue and green novel objects and asked the child to place the "toma" in the chute (one trial). Following another Familiar object trial (a car and cow, with E asking for the former), children engaged in the contrast task (one trial): Again, the two novel objects were presented, but this time children were asked to place the "blicket" in the chute. For each child, the positions of the two novel objects on the tray were counterbalanced across the endorsement and contrast task, and across participants their positions were counterbalanced for each type of task. Finally, as an explicit competence judgment, E1 showed the child a print-out of a screenshot of the two speakers and asked the child who they thought was good at naming things. This task was only conducted with the 5-year-olds, as piloting revealed that the explicit competence judgment was linguistically too demanding for 2-year-olds. Among the 5-year-olds, more than 90% of children identified the reliable speaker as the one good at labeling objects (i.e., 32 responded correctly, 3 incorrectly, and 1 was not asked due to an E error).

Analyses

Accuracy exposure phase and novel label training

We defined the location of the speaker and the object as separate areas of interest (AOIs). The AOIs for each speaker measured 570 × 440 pixels in both phases, the object AOI for the accuracy exposure measured 760 × 450 pixels, the object AOI for the novel label training 470 × 575 pixels. None of the AOIs overlapped. For each phase, we assessed the duration of time children spent looking to each of these AOIs for the labeling events by the reliable speaker and by the unreliable speaker, respectively. We then tested whether children's gaze behavior during the Accuracy exposure and the Novel label training differed between the labeling events by the reliable and unreliable speaker (for details see Appendix).

Eye-tracking test phase

We defined AOIs for the objects' locations on the left and right side of the screen where the images appeared during the test phase. Each image measured 344 × 696 pixels; AOIs were defined as 440 × 750 pixels around the center of the image, with the two AOIs separated by 540 pixels. We excluded trials in which the child did not attend to the presentation, operationalized as looking away from the screen for at least half of the trial's duration. For the 2-year-olds, this led to the exclusion of 12 trials (7 in the Familiar object task, 2 in the Endorsement and 3 in the Contrast task) and for the 5-year-olds, five trials were excluded (2 in the Familiar object task, 2 in the Endorsement and 1 in the Contrast task).

To assess children's looking responses, we compared their proportion of target looking (PTL) before and after label onset. The target in the Familiar object task was the labeled object. The target in the two Novel object tasks was the object labeled by the reliable speaker. The prenaming window was defined as the period 2000 ms before label onset (i.e., 500–2500 ms following video onset). The postnaming window also lasted 2000 ms, starting 240 ms after label onset (i.e., 2740–4740 ms following video onset) to ensure that only fixations that could be reliably construed as a response to the auditory stimulus were included in the postnaming target window. The proportion of time spent looking at the target was calculated by dividing the time spent looking at the target by the sum of time looking at either object. The PTL was determined for each time window, averaging across trials in the same task (Endorsement task, Contrast task, and Familiar object task).

For the Familiar object task, we then ran a generalized linear mixed model (GLMM) with a beta error distribution and logit link function using the R package glmmTMB (version 1.1.7; Brooks et al., 2017). We included PTL time as dependent variable, age group (2 vs. 5 years), time window (pre- vs. postnaming) and their interaction as predictors and random intercepts for participants. For the Novel label tasks, we used the same analysis and added type of task (Endorsement vs. Contrast) as well as all its possible interactions as predictors.

Interactive tasks

For the Familiar object task, we coded whether children selected the labeled object, the distractor, both, or none. We conservatively treated trials on which children participated but then chose neither object or both of them, as incorrect rather than excluding them. To test their performance against chance, we ran a GLMM with binomial error distribution using the R package lme4 (version 1.1-32; Bates et al., 2015). We included the selection of the labeled object as dependent variable, age group (2 vs. 5 years) as predictor and random intercepts for participants. We suppressed the intercept to test performance against 0.5 (chance level) and obtained separate values for each age group.

Similarly, for the two Novel object tasks, we coded whether children selected the object labeled by the reliable speaker, the one labeled by the unreliable speaker, both, or none. Again, selections of both or neither object were treated as incorrect. To assess whether children's object choices differed between tasks (Endorsement vs. Contrast) and age groups (2 vs. 5 years), we ran a binomial mixed effects regression with choice of the object labeled by the reliable speaker as dependent variable, task, age group and their interaction as predictors, and random intercepts for participants.

Relation between the eye-tracking and interactive tasks

To explore the relation between children's performance in the eye-tracking version and the interactive version of each task, we fitted a GLMM with binomial error distribution. We predicted children's choices in the interactive tasks by age group, task, the PTLs in the eye-tracking tasks (post labeling, and aggregated across trials) and all their possible interactions, and added random intercepts for participants.

Success in the endorsement task predicted by looking times during the novel label training

Finally, we investigated whether children's looking times during the novel label training would predict their success at test (the eye-tracking or interactive endorsement task). One prediction derived from the attentional account would be that those children who look more when the reliable rather than the unreliable speaker taught the novel label would be more likely, upon hearing that novel label, to choose the object that the reliable speaker had referred to. Thus, we first calculated the proportion of time children looked at the training provided by the reliable speaker versus the reliable and unreliable speaker combined (aggregated across trials). We then assessed whether children's success in the endorsement tasks (eye tracking and interactive) was predicted by the proportion of looking to the reliable speaker during the label training by age group and their interaction (for details see Appendix).

The analyses of Study 1 to 3 were not preregistered. The data, analysis scripts, and detailed model results for all three studies are accessible on OSF (https://osf.io/qsyu9/?view_only=5c5bdaa82666487c94ea894d6c98d11a). If not stated otherwise, the model assumptions were met.

Results

We first present the data for the Familiar object tasks, followed by the data for the Novel object tasks, and in the end the analyses of gaze behavior during the accuracy exposure phase and the novel label training.

Familiar object tasks

We predicted for both the eye-tracking and the interactive task that, upon labeling, children would look at/select the target object (i.e., the requested familiar object) more (often) than the distractor object. In fact, in the eye-tracking task, both 2-year-olds ($b=0.70$, $SE=0.15$, $p<.001$) and 5-year-olds ($b=1.21$, $SE=0.14$, $p<.001$) looked significantly more at the target in the postnaming compared to the prenaming window (Figure 2). A significant interaction between age group and time window indicated that this effect was stronger for the 5-year-olds than the 2-year-olds ($b=0.51$, $SE=0.21$, $p=.014$). Similarly, in the interactive task, 2-year-olds selected the familiar target object significantly above chance (88% correct trials; $b=2.79$, $SE=0.51$, $p<.001$) and 5-year-olds even performed at ceiling (100% correct trials; test results thus uninformative).

Novel object tasks

If participants were influenced by the speakers' previous accuracy, they should show the following response pattern: In the Endorsement task, when they heard the trained novel label "toma," they should look at/ select the object labeled by the reliable speaker and in the Contrast task, when they heard the completely novel label "blicket," they should look at/ select the object labeled by the unreliable speaker.

Our analysis revealed a three-way interaction of age group, time window and type of task ($b=-1.04$, $SE=0.41$, $p=.012$): For the 5-year-olds, there was a significant interaction between time window and type of task ($b=-0.84$, $SE=0.28$, $p=.003$), but not for the 2-year-olds ($b=0.20$, $SE=0.30$, $p=.497$). As expected, 5-year-olds significantly increased their looking to the object labeled by the reliable speaker after label onset in the Endorsement task

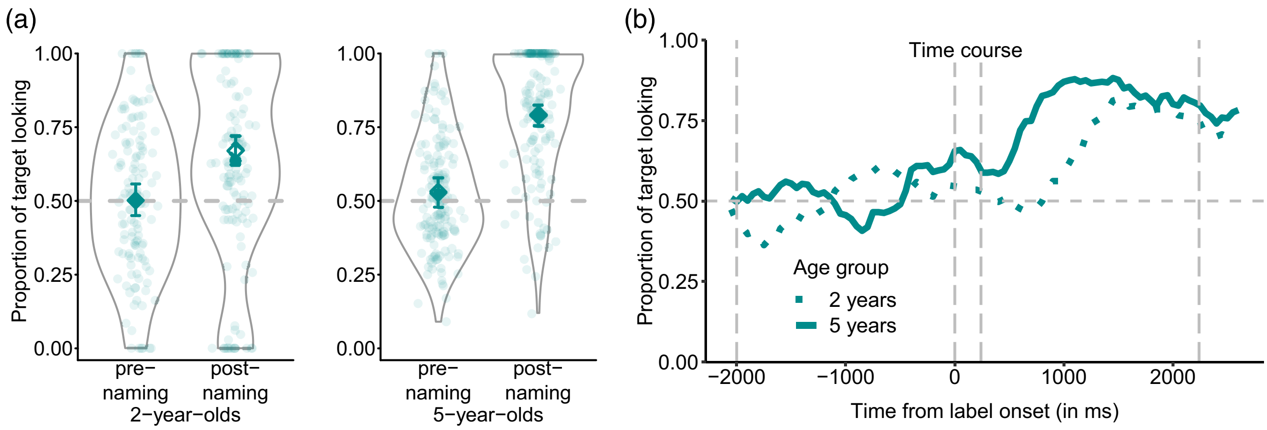

FIGURE 2 Eye-tracking results for the familiar object task. (a) Proportion of target looking (PTL) in both age groups in the pre- and postnaming phase. Transparent dots show the individual PTLs per trial and filled dots the average PTL per age group and time window. Diamonds represent the fitted values and vertical lines their 95% confidence intervals, both obtained by the generalized linear mixed model via bootstrapping with 1000 boots. (b) Time course visualization.

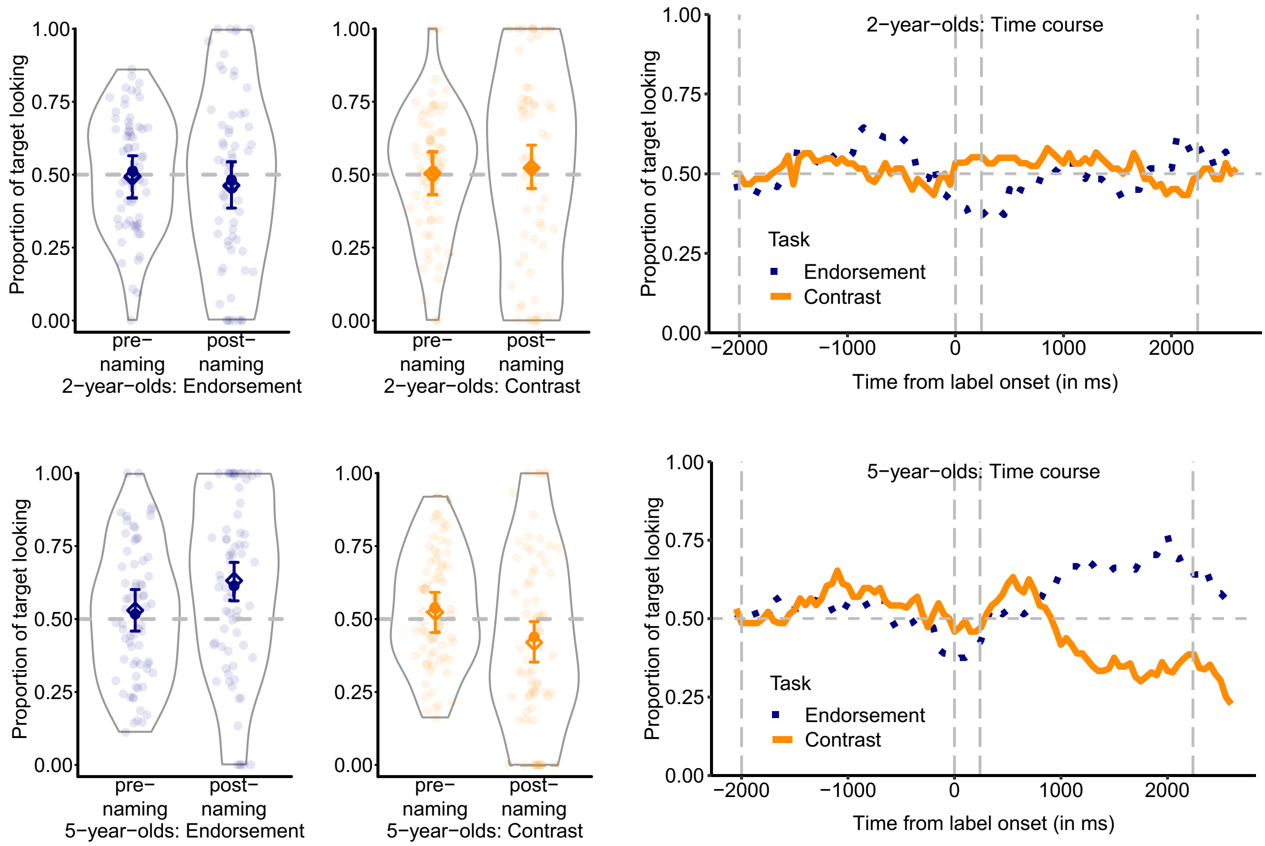

FIGURE 3 Eye-tracking results for the two Novel object tasks. Left: Proportion of looking at the object labeled by the reliable speaker in the pre- and postnaming phase in both tasks, for 2-year-olds (upper) and 5-year-olds (lower). Transparent dots show the individual proportions per trial and filled dots the average proportion per age group and time window. Diamonds represent the fitted values and vertical lines their 95% confidence intervals, both obtained by the generalized linear mixed model via bootstrapping with 1000 boots. Right: Time course visualization for both tasks, for 2-year-olds (upper) and 5-year-olds (lower).

(b=0.42, SE=0.20, p=.037), and significantly decreased it in the Contrast task (b=−0.42, SE=0.20, p=.038). In contrast, there was no significant difference in the PTL across the pre- and postnaming phase for 2-year-olds, neither for the Endorsement (b=−0.13, SE=0.21, p=.552) nor the Contrast task (b=0.08, SE=0.21, p=.714) (Figure 3).

Children's responses in the two interactive novel object tasks mirror these findings. The 5-year-olds chose the object labeled by the reliable speaker significantly more

TABLE 1 Object choices in the interactive tasks in Study 1 and Study 2.

	Object labeled by the reliable speaker	Object labeled by the unreliable speaker	Both or none
5-year-olds Study 1			
Endorsement task	26	10	0
Contrast task	10	25	1
2-year-olds Study 1			
Endorsement task	12	17	3
Contrast task	7	18	7
2-year-olds Study 2[a]			
Endorsement task	20	14	1
Contrast task	16	17	2

[a] $n=35$, as one child did not take part in the novel interactive trials.

often in the Endorsement than the Contrast task ($b=1.91$, SE$=0.53$, $p<.001$; see Table 1). For the 2-year-olds, however, object choices were not significantly predicted by the task ($b=0.76$, SE$=0.56$, $p=.175$). There was no significant interaction of age group and task ($b=1.15$, SE$=0.77$, $p=.136$). Accordingly, 5-year-olds selected the object labeled by the reliable speaker significantly more often than 2-year-olds in the Endorsement task ($b=1.47$, SE$=0.52$, $p=.005$), but not in the Contrast task ($b=0.32$, SE$=0.57$, $p=.575$).

We also explored the relation between children's performance in the eye-tracking and the interactive version of each task. For 5-year-olds, the postlabeling PTLs positively predicted their success in both the interactive endorsement ($b=6.08$, SE$=2.10$, $p=.004$) and contrast task ($b=5.35$, SE$=1.94$, $p=.006$). However, this was not the case for 2-year-olds, (endorsement: $b=1.37$, SE$=1.53$, $p=.371$; contrast: $b=2.70$, SE$=2.57$, $p=.293$).

Gaze behavior during the accuracy exposure phase and the novel label training

During the accuracy exposure, the gaze behavior of neither age group differed depending on whether they watched the labeling events by the reliable speaker or by the unreliable speaker, neither for their looking at the speaker nor the object (for details see Appendix). During the novel label training, no differences in gaze behavior were observed, either. And the proportion of looking to the reliable speaker during label training did not predict children's success in the subsequent endorsement tasks, neither on the eye tracking nor the interactive version (for details see Appendix).

Discussion

In Study 1, children were told by a previously accurate speaker that one object was called a "toma" while a second, previously inaccurate speaker claimed a different object was called a "toma." In eye-tracking and interactive tasks, we investigated children's selective endorsement of this conflicting testimony and the inferences they based on it.

At age five, children looked at and selected the object labeled by the reliable (vs. unreliable) speaker when they were asked for the "toma." Furthermore, when they heard another completely novel label ("blicket") they looked at and selected the object labeled by the unreliable speaker rather than the one labeled by the reliable speaker. These findings suggest that 5-year-olds selectively endorsed the testimony offered by the speaker with the more reliable track record. Regarding the cognitive underpinning of their selectivity, children's looking behavior during the label training (and its relation to their subsequent performance in the test phases) contradicts the predictions derived from the attentional account, but, instead, is in line with the selective endorsement account (for further elaboration see the "General Discussion" section).

At age two, however, children performed at chance level in both the Endorsement and Contrast task. In neither task did they show a clear preference for either the object labeled by the reliable or unreliable speaker. In fact, their responses did not differ systematically depending on which label they encountered: the trained label introduced incompatibly by the two speakers or a completely novel label. They showed at-chance performance in both the eye-tracking and interactive tasks. Thus, in contrast to the 5-year-olds, they showed no evidence of selectively endorsing the testimony offered by the speaker with the more reliable track record. Two possible interpretations come to mind. The first is that 2-year-olds are as yet unable to use speakers' past accuracy to evaluate whose testimony to endorse (Ganea et al., 2011; Hermes et al., 2019). Another possibility is that the present study masked 2-year-olds' ability to selectively endorse reliable over unreliable sources of novel information. Their successful performance in the familiar object trials suggests that the general structure of the test trials was not too demanding for 2-year-olds. Nevertheless, some aspects of our design may have presented performance demands masking 2-year-olds' potential competence. For example, children received little input regarding which novel object each speaker had been labeling, given that they were presented with only three labeling events in total per speaker. Furthermore, it is possible that children only attended in detail to the speakers' respective trustworthiness once the two speakers provided conflicting information regarding the labels of the novel objects. At this stage, however, the speakers only provided very limited information on which children could judge their reliability (labeling one familiar entity) and limited input regarding who referred to which object (one labeling event each). In Study 2, we modified these aspects of our design to assess whether 2-year-olds selectively endorse novel information from reliable rather than unreliable speakers given less demanding circumstances.

STUDY 2

Method

Participants

Thirty-six monolingual German children (M_{age}=24.0 months, range=23.1–25.9 months, 17 female) were included in the final sample of this experiment. Three additional children were tested but could not be included in the final sample due to technical problems with the eye-tracker (n=1) or their not completing the experiment (n=2). The recruitment and parental consent were the same as in Study 1.

Materials, design and procedure

The materials used and the design of the study were the same as described for Study 1. The general procedure was also the same, except for the following changes that were introduced to make the procedure less demanding for the 2-year-olds. First, we increased the input provided in the Novel label training and the input provided in the reminder (see details below). Second, there were slight changes to the test phase, both regarding the eye-tracking part and the set-up of the interactive task.

Novel label training and reminder

After the accuracy exposure phase that followed the same script as was presented in Study 1, children were presented with two sessions of Novel label training, interspersed by the accuracy reminder phase. During each Novel label training session, each speaker engaged in three labeling events (in blocked order). Specifically, the novel object to be labeled by the first speaker appeared on the screen (in the bottom corner of the screen on the opposite side to the speaker's location). Then the first speaker appeared, addressed the child saying, “Oh, look!”, and turned to point at the object, saying, “This is the toma, the toma!” Then the object moved to another corner of the screen and the speaker commented saying “Oh, look! Now the toma is here, the toma!” and then the object moved to the center of the screen, she again said “Oh, look! And now the toma is here, the toma!” This same sequence was presented with the second speaker labeling a different object a “toma.” The movement of the novel objects across the screen was included to keep children interested during the longer input sequence (38.5 s per speaker). Subsequently, during the accuracy reminder phase, children saw each speaker label two familiar entities (a baby and a house) either accurately or inaccurately according to their speaker role (in alternating order). Then the Novel label training session was repeated as a whole (i.e., three labeling events per speaker in blocked order) starting again with the speaker who spoke first in the earlier Novel training phase. Thus, in Study 2, children heard each speaker use the label a total of 12 times (during six labeling events), compared to six times (during three labeling events) in Study 1.

Test phase

For both the eye-tracking and the interactive version of the tasks, the endorsement task always preceded the contrast task. Tasks were presented in this fixed order, as the introduction of a completely novel label in the contrast task may confuse the younger children and may mask any fragile competence, they would otherwise have been able to demonstrate on the endorsement task. The novel label used in the contrast task in Study 2 was “shoofie” (as this presented a better match to the sound pattern of the participants' native language). The test session started with two familiar trials (with the familiar target appearing once on each side of the screen), to give children more experience with the general procedure before the start of the novel label trials. In the interactive phase, children were asked to place the target in a box that E2 presented to them (so that children did not have to get up and walk somewhere in between trials).

Analyses

The analyses were conducted the same way as in Study 1, although without age group and any interaction with age as predictors. Again, we excluded trials in which the child looked away from the screen for at least half of the trial's duration. This led to the exclusion of three

Familiar object trials, one Endorsement trial, and four Contrast trials.

Results

Familiar object task

In the eye-tracking task, the PTL was significantly higher in the postnaming phase (M=0.72, SD=0.24) than the prenaming phase (M=0.50, SD=0.25; b=0.96, SE=0.16, p<.001). Similarly, in the interactive task, they selected the target object significantly above chance (86% correct trials; b=2.30, SE=0.60, p<.001). Thus, 2-year-olds looked at and selected the labeled object more often than chance, showing that they could cope with the general demands of these tasks when they involved familiar objects.

Novel object tasks

In the eye-tracking task, the proportion of looking to the object labeled by the reliable speaker did not differ between the pre- and postnaming window, neither for the Endorsement (b=0.07, SE=0.20, p=.721), nor the Contrast task (b=0.30, SE=0.20, p=.139). Furthermore, our analysis revealed no significant effect of type of task, neither for the prenaming (b=−0.09, SE=0.20, p=.666) nor the postnaming time window (b=0.14, SE=0.20, p=.476), and no interaction (b=0.23, SE=0.28, p=.419) (Figure 4). Similarly, in the interactive task, 2-year-olds' choices of the object labeled by the reliable speaker were not predicted by the type of task (Endorsement vs. Contrast task; b=−0.35, SE=0.48, p=.473, see Table 1).

Discussion

In Study 2, we doubled the amount of exposure to the specific label-speaker-object pairings as well as provided children with a longer reminder of the accuracy of the different speakers prior to assessing their endorsement of the information provided by the two speakers. Nevertheless, we found no evidence that 2-year-olds selectively endorse the testimony by the more reliable source, despite the increased input. Instead, they performed at chance level both when they needed to selectively endorse the information provided by either the reliable or unreliable speaker (Endorsement task) and when they needed to choose which object a completely novel label referred to (Contrast task). Such a chance level performance was found both in the eye-tracking and in the interactive version of these tasks.

Children's successful performance when they were presented with familiar objects (Familiar object tasks) highlights again that the general task demands were not too challenging for this age group. So what is the reason that they failed at the same kind of task when the two novel objects were presented? It could be that despite the adjustments we made in Study 2 the linguistic demands involved may still have been too challenging for the 24-month-olds. One issue might have been the type and amount of input children experienced. For example, a number of studies have shown that 24-month-olds sometimes struggle to learn novel information when the input is provided on screen (esp. in a non-interactive setting such as watching a film) rather than in a real-life interactive scenario (Strouse & Samson, 2021). The second reason for children's chance-level performance may be that at this age they lack the requisite competence: selective trust in the information provided by the more reliable source. Children may have learned the novel label, but because this label was associated with two novel objects, if they did not take the speakers' previous accuracy into account

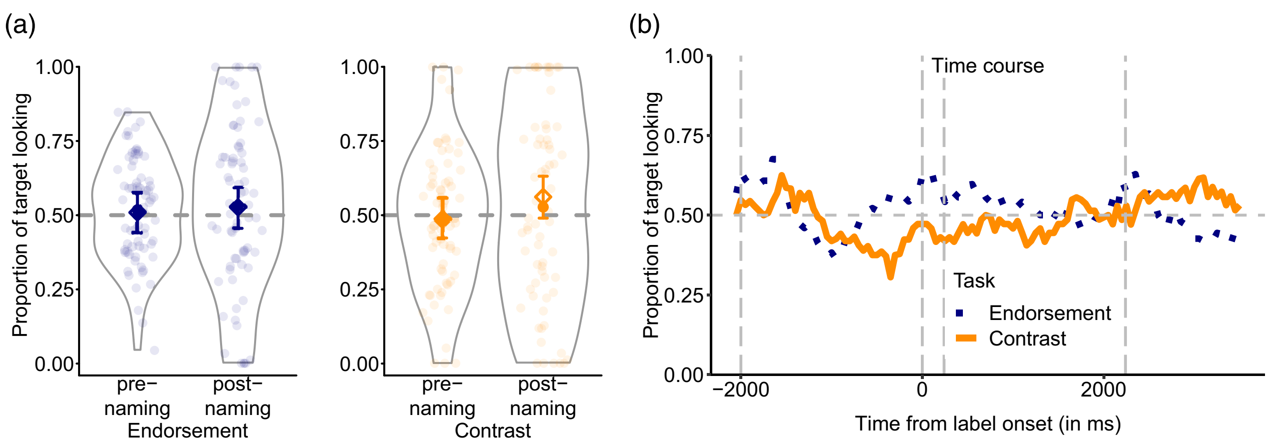

FIGURE 4 Eye-tracking results for the two Novel object tasks in Study 2. (a) Left: 2-year-olds' proportion of looking at the object labeled by the reliable speaker in the pre- and postnaming phase in both tasks. Transparent dots show the individual proportions per trial and filled dots the average proportions per age group and time window. Diamonds represent the fitted values and vertical lines their 95% confidence intervals, both obtained by the generalized linear mixed model via bootstrapping with 1000 boots. (b) Right: Time course visualization for both tasks.

and selectively endorsed the more reliable source, they would have to choose randomly between the two novel objects.

To distinguish between these two possible explanations, in Study 3, the two speakers, who were first established as reliable or unreliable in the accuracy phase, then labeled two different objects using two different labels in the novel label training phase. Thus, the information provided by these speakers did not conflict with one another. If 2-year-olds' difficulties in Study 2 were due to performance factors, such as the linguistic demands involved, then they would again show chance-level performance in Study 3 when, upon hearing one of the two novel labels, they were presented with the two novel objects. If anything, the word-learning demands were increased, rather than lowered, in Study 3 compared to Study 2, as children were taught two novel labels instead of just one. If, however, 2-year-olds' difficulties in Study 2 were not due to the linguistic performance demands involved, but reflected a lack of competence in monitoring and trusting the more reliable source, then they may succeed in a word-learning task in which the two speakers provide nonconflicting information.

STUDY 3

Study 3 examined how 2-year-olds respond to novel information provided by a reliable and an unreliable speaker, when the information provided by the two does not conflict with one another. Thus, in contrast to the previous studies, in Study 3, the two speakers provided compatible rather than conflicting information. As in Study 1 and 2, following exposure to the accuracy of each of the speakers, each speaker labeled a different novel object. In Study 3, however, they used different, rather than the same novel label to refer the two different objects. One speaker called one novel object a "toma" and the other speaker called a different object a "shoofie." This study thus presented the same, or actually higher, word learning demands than the previous ones, given that children were presented with two, rather than one, novel label. However, in contrast to the eye-tracking task used in Study 1 and 2, children here were not forced to choose between the two speakers' testimony. Thus, this design allows us to assess whether 2-year-olds are capable of meeting the linguistic demands posed in this kind of eye-tracking tasks. This enables us to disentangle whether 2-year-olds' poor performance in Studies 1 and 2 reflects their difficulties with the linguistic task demands or with the requirement to make person-specific attributions of trustworthiness.

Method

Participants

Thirty-five monolingual German children (M_{age} = 23.9 months, range = 23.3–24.9 months, 16 female) were included in the final sample of this experiment. Five additional children were tested but not included in the final sample, because of technical problems (n = 4) or because the child did not complete the experiment (n = 1).

Materials and design

The materials and design were identical to Study 2, with the following exceptions. The two speakers did not use the same label in the novel training phase, instead one labeled one of the novel objects, "toma," and the other labeled the other object "shoofie." The assignment of labels ("toma" or "shoofie") to speakers (reliable vs. unreliable speaker) and objects (green novel object vs. blue novel object) was counterbalanced across children. Except for this, the procedure of the Novel label training was identical to that of Study 2. The test phase comprised two types of Novel object tests: the "reliable-speaker test" assessed the label-object link taught by the reliable speaker, and the "unreliable-speaker test" examined this for the label taught by the unreliable speaker. The eye-tracking test phase consisted of four such Novel object trials (two per type of test), and the interactive task comprised two Novel object trials (one per type of test).

Procedure

Following an identical accuracy exposure phase as was presented in Study 2, children were presented with the Novel label training and reminder as in Study 2, followed by an eye-tracking test phase. For the Novel object trials, again the two novel objects that had previously been labeled by either speaker were presented on screen. However, in Study 3, the male voice asked for either the referent of the label used by the reliable or the unreliable speaker, that is, either the "toma" or the "shoofie" across trials. That is, in the "reliable speaker test," children were asked for the referent of the label introduced by the reliable speaker and in the "unreliable speaker test" for the referent of the label provided by the unreliable speaker. In the interactive task, children participated in the same two types of tests. In addition, both the eye-tracking and the interactive tasks presented children with the same (number of) Familiar object trials as in Study 2.

Analyses

The analyses were the same as in Study 2, except that in the Novel object tasks, depending on the type of test, the target was either the object labeled by reliable or the object labeled by the unreliable speaker. In addition, we tested children's interactive selections of the target object in the "reliable speaker" and "unreliable speaker" test (separately) against chance by running a GLMM with binomial error distribution. We predicted children's target selections by type of test, added random intercepts, and suppressed the intercept to test against the chance level of 0.5. We excluded trials in which the child looked away from the screen for at least half of the trial's duration. This led to the exclusion of 31 trials in the Familiar object task and 32 trials in the Novel object tasks (14 trials in the tests relating to the reliable speaker and 18 trials relating to the unreliable speaker). Consequently, 6 children did not provide data for both types of tests of the Novel object task.

Results

Familiar object task

As in both previous studies, the 2-year-olds showed a significant increase in the proportion of looking to the labeled familiar objects in the postnaming phase ($M=0.66$, $SE=0.28$) relative to the prenaming phase ($M=0.49$, $SD=0.24$; $b=0.90$, $SE=0.19$, $p<.001$). Similarly, in the interactive tasks, they selected the target object significantly more often than predicted by chance (88% correct trials; $b=1.96$, $SE=0.29$, $p<.001$).

Novel object task

We assessed whether 2-year-olds were able to identify the correct object upon hearing the labels introduced by the reliable and the unreliable speaker, respectively. For the eye-tracking data, the analysis revealed a significant effect of time window, both for the "reliable speaker" ($b=0.55$, $SE=0.23$, $p=.019$) and the "unreliable speaker" tests ($b=0.56$, $SE=0.24$, $p=.021$). There was no effect of type of test in either time window (prenaming: $b=0.24$, $SE=0.24$, $p=.321$; postnaming: $b=0.24$, $SE=0.24$, $p=.301$) and no significant interaction of time window and type of test ($b=0.01$, $SE=0.33$, $p=.977$). In other words, upon hearing the novel labels, the 2-year-olds looked at the respective target object, and they did so irrespective of which speaker had taught them the novel label (Figure 5).

In the interactive task, children's performance did not differ significantly depending on whether the novel label had been taught by the reliable or the unreliable speaker ($b=-0.85$, $SE=0.52$, $p=.099$). Furthermore, their selections did not differ from chance level, neither in the reliable speaker ($b=0.67$, $SE=0.38$, $p=.077$) nor in the unreliable speaker tests ($b=-0.18$, $SE=0.35$, $p=.613$; see Table 2).

Discussion

Upon hearing the two novel labels trained by the two speakers, 2-year-olds looked at the respective target object (i.e., the novel object that the speaker had referred to during the Novel label training). This change in their PTL in response to hearing the novel labels clearly indicates that the 2-year-olds recognized the novel words and linked them to the respective novel objects. Thus, children this age mastered the demands of this word-learning task, at least in the eye-tracking version, even though successful performance involved linguistic demands that were as high (or perhaps even higher) than those in Study 1 and 2. In contrast to those studies, however, successful performance in Study 3 did not necessitate making person-specific attributions of trustworthiness on the basis of past reliability. Whereas in studies 1 and 2, the two speakers had provided conflicting information which children needed to choose between (with successful performance requiring the selective endorsement of the information provided by the more reliable speaker), in Study 3 the two speakers provided compatible information. Thus, children could

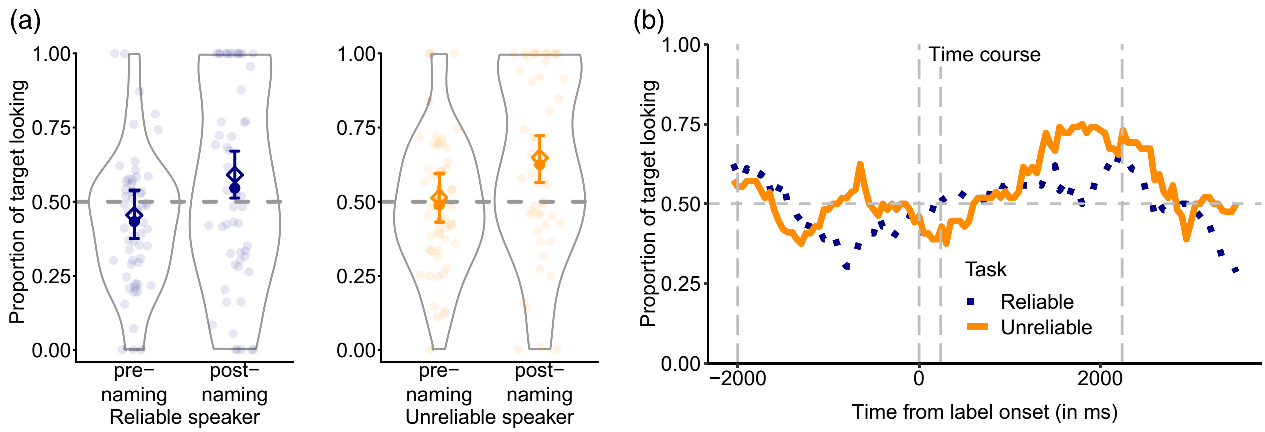

FIGURE 5 Eye-tracking results for the Novel object task in Study 3. (a) Two-year-olds' proportion of target looking by type of test (label introduced by the reliable or unreliable speaker) and time-window. Transparent dots show the individual proportions per trial and filled dots the average proportions per age group and time window. Diamonds represent the fitted values and vertical lines their 95% confidence intervals, both obtained by the generalized linear mixed model via bootstrapping with 1000 boots. (b) Time course visualization.

TABLE 2 Object choices in the interactive task by type of test.

	Choice of target object	Choice of distractor object	Both or none
Reliable speaker test	23	10	2
Unreliable speaker test	16	19	0

demonstrate their recognition of the novel words, without having to track the speakers' previous reliability. Our pattern of results suggests that the 2-year-olds' chance performance in Study 1 and 2 may not reflect difficulties with the linguistic task demands (given that these were no higher than in Study 3), but instead seem to reflect a lack of competence with respect to evaluating a speaker's testimony on the basis of that person's past reliability.

In the interactive task, the 2-year-olds again showed no robust difference in performance depending on which speaker had introduced the novel label. However, their overall performance did not differ significantly from chance suggesting that the 2-year-olds may have had difficulty retaining several novel word-object links, that their concentration may have decreased at the end of the session, or that their performance may diverge depending on the response formats (looking time vs. interactive responses, see, e.g., Bannard & Tomasello, 2012; Clements & Perner, 1994; Hood et al., 2003).

GENERAL DISCUSSION

The aim of this set of studies was to investigate the cognitive underpinnings and early development of young children's selective trust in reliable over unreliable sources. Using a two-informant design, children aged 2 and 5 years were presented with an unreliable and a reliable speaker who—in Study 1 and 2—both used the same novel label ("toma") to refer to two different novel objects. In eye-tracking and interactive tasks, we then tested whether upon hearing this novel label used by a third person, children would selectively look at/ select the object that had been labeled by the reliable speaker relative to the one labeled by the unreliable speaker. The 5-year-olds selectively endorsed the label-object link taught by the reliable speaker, in both the eye-tracking and the interactive version of the task. The 24-month-olds, however, performed at chance level in both versions of the task, even in Study 2 when we tried to simplify the task's performance demands. In addition, children participated in eye-tracking and interactive versions of a Contrast task, in which they were presented with the same two novel objects (i.e., the objects labeled as "toma" by the reliable and unreliable speaker, respectively), but heard a third person use another completely novel label. We expected that if children selectively learned the information provided by the reliable source, then they should look at/ select the object labeled by the unreliable speaker. And indeed, the 5-year-olds showed this response in both the eye-tracking and the interactive version of the task. The 24-month-olds, however, again performed at chance level both in the eye-tracking and the interactive task.

Regarding the cognitive underpinnings of children's selective trust, we set out to test the premises of the attentional bias account, namely whether children attend more to information provided by a reliable speaker than an unreliable one and whether children's attentional bias predicts their selective trust in reliable over unreliable sources. To address these questions, we examined the duration of time children spent watching the reliable and unreliable speaker, respectively, as the two taught the participant novel information. Neither 2-year-olds, nor 5-year-olds spent more time observing the novel label events by the reliable speaker relative to the ones by the unreliable speaker. Thus, we found no evidence in favor of the suggestion that children attended more to the reliable than the unreliable speaker. Furthermore, the proportion of time a child spent looking at the teaching events of the reliable speaker, relative to the ones of the unreliable speaker, did not predict their selective endorsement of the label-object link by the reliable speaker (as indicated by successful performance in the endorsement tasks). Thus, these findings provide no evidence for the suggestion that children's selective trust is simply based on an attentional bias.

Instead, our findings are in line with research that suggests children's selective learning of novel labels is not based on inattention toward the information provided by questionable sources (ones that profess their ignorance or that have previously proven inaccurate in related matters), but on the selective encoding and consolidation of the semantic information provided by a more reliable source (Mangardich & Sabbagh, 2018; Sabbagh et al., 2003; Sabbagh & Shafman, 2009). As Sabbagh and colleagues have demonstrated, children can recount (at least immediately after training) how an unreliable speaker (in this case one who had professed ignorance) had referred to a novel object, but they do not endorse this label, when asked what that novel object is called. In other words, the distinction here is between the episodic recollection of the speaker's labeling event ("She called this an X") and the semantic encoding and consolidation of that label ("This is an X") (Mangardich & Sabbagh, 2018; Sabbagh & Shafman, 2009). Our findings, using a different methodological approach (the monitoring of looking behavior during training and its relation to children's selectivity), are in line with this proposal and they extend it by suggesting that this mechanism may not be limited to cases in which the speakers profess ignorance, but may also extend to cases in which the speakers' respective reliability needs to be inferred based on their previous behavior. This distinction is not trivial as in the former case the novel information (i.e., the novel label-object link) and the cues that provide information about the sources' quality are presented simultaneously and are explicitly linked (i.e., “I don't know what an X is, perhaps this is an X”), while in the latter case they are presented sequentially and are not explicitly linked.

In addition, our findings show that at the age that children showed selective learning from the more reliable source, they were also able to judge source reliability explicitly. This was shown by the 5-year-olds' successful performance in the explicit judgment task, in which they identified who was good at labeling objects and who was not. This is in line with previous research that suggests that children's selective learning from reliable sources correlates with—and may even depend on—the ability for trait ascription (Hermes et al., 2015; Koenig & Harris, 2005). For example, when preschoolers were presented with a choice between two individuals who had each displayed a positive trait (either expertise in naming or superior strength) children selected the individual whose trait matched the affordance of the task at hand, choosing the former for knowledge and the latter for strength-requiring tasks. Moreover, this selective pattern was only found among those children who could also answer trait questions correctly, suggesting that their trait ascriptions mediated their selective help-seeking (Hermes et al., 2015).

In sum, this body of research demonstrates that preschoolers from around 4–5 years of age selectively trust reliable over unreliable sources. This selectivity does not seem to be based on inattention to the information provided by the unreliable source. Instead, children this age are able to explicitly identify the more reliable source based on the previous accuracy of a specific speaker—an ability that goes hand in hand with their selective endorsement of the contested novel information provided by this source. Thus, these findings do not provide any evidence supporting the attentional bias accounts, but they are in line with the idea that preschoolers' selective trust may be based on trait ascription.

But what about the earlier ontogeny? We did not find any evidence that 2-year-olds selectively learned the novel information provided by the more reliable source. In contrast to previous research that had tested 2-year-olds with a two-informant design (Ganea et al., 2011; Hermes et al., 2019), we looked at toddlers' endorsement of information that was semantic in nature (e.g., which object is called a “toma”) rather than episodic (e.g., where an object was hidden). Even though the evaluation of semantic information is considered to be more informative than the evaluation of episodic information (Stephens & Koenig, 2015), toddlers performed at chance level when asked to choose between conflicting semantic information provided by a reliable and an unreliable speaker. This contradicts the suggestion that the use of episodic information had masked toddlers' early competence in selectively learning from reliable over unreliable sources.

Toddlers did not only show chance performance in the interactive tasks but also in the eye-tracking tasks that tested their selective endorsement of the more reliable source (see Study 1 and 2). Thus, regarding their ability to make person-specific ascriptions of trustworthiness it was *not* the case that an early sensitivity or competence could already be detected in toddlers' looking behavior. However, toddlers did show signs of learning novel label-object links when this did not require the person-specific ascription of trustworthiness. When in Study 3 the two speakers provided compatible information (i.e. each teaching a different novel label for a different object), toddlers looked significantly longer at the target objects upon hearing the novel labels (relative to the prenaming baseline). Note that we did not find significant differences in children's looking patterns depending on who had taught the novel label, the reliable or the unreliable speaker.

How do our findings with the 2-year-olds relate to previous research that explored infants' and toddlers' selective trust using the one-speaker design? There are a few studies showing that under certain circumstances children between 18 and 24 months learn novel labels at a higher rate, if the speaker teaching them had previously labeled familiar objects accurately rather than inaccurately (Crivello et al., 2018; Koenig & Woodward, 2010; Krogh-Jespersen & Echols, 2012). Even 8-month-olds track the reliability of potential informants, in the sense that they are quicker to look at a visual target if the gaze cues were provided by a face whose gaze direction had previously been predictive rather than nonpredictive for the appearance of rewarding visual targets (Tummeltshammer et al., 2014). Interestingly, recent work with 18-month-old infants showed that those infants who interacted with an unreliable speaker performed worse at test than those interacting with a reliable one, regardless of whether the task was to select an object whose label infants had known all along (as indicated by parental questionnaire, e.g., “Where is the ball?”) or had just been taught by the speaker (e.g., “Where is the dax?”; e.g., Crivello et al., 2018; Kuzyk et al., 2020). This is in line with the idea that the somewhat bizarre situation of encountering an adult consistently mislabeling familiar objects might slightly affect children's performance across the board. Thus, toddlers, and even young infants, may learn to use predictive cues and may learn slightly less under odd circumstances. But this does not necessarily mean that children this age use a specific person's past accuracy as an index of whether or not to endorse the contested novel information this person provided. Research to date provides no robust evidence that infants and toddlers show selective trust based on person-specific differences in reliability.

One curious point regarding the research designs used across the different age groups, is that the one-informant design typically used to explore selective trust in toddlers is the very paradigm in which preschoolers fail to demonstrate selectivity (Jaswal et al., 2010, 2014; Kim et al., 2017; Vanderbilt et al., 2014). In fact, Vanderbilt et al. (2014) found that 3- to 4-year-olds trusted a single inaccurate informant at the same level as they trusted a single accurate informant, with no significant difference between these two conditions. Especially when faced with intentionally communicated information, young preschoolers are remarkably trustful (Mascaro & Morin, 2014).

Taken together, the present findings raise the interesting, yet speculative, idea that young children may apply different learning strategies at different developmental stages. While infants and toddlers may engage in default trust without ascribing reliability to specific individuals, preschoolers do engage in selective trust, ascribing reliability to specific individuals based on the informants' relevant track record. This change in learning strategies may present a conceptual development, reflecting the more sophisticated processes required, for example, trait concepts, and thus may emerge gradually and in protracted ways (e.g., Liu et al., 2007). This might even present a sensible ontogenetic adaptation: Interactions in children's first years of life usually involve benevolent and more knowledgeable communicators, thus default trust might pay off initially (esp. when these interaction partners act confidently, see, e.g., Birch et al., 2010). Later, however, as children become more mobile and independent and start to acquire much more nuanced and specialized knowledge and skills, it may be more adaptive to be selective in whom they learn from (see, e.g., Gopnik et al., 2015; Mascaro & Morin, 2014; Tomasello, 2019). The cognitive underpinnings of such selectivity seem to be neither general confusion nor attentional biases. Instead, research findings suggest that young children's selective learning is mediated by trait-like person-specific ascriptions of trustworthiness.

ACKNOWLEDGMENTS

We thank all the children who participated in our studies as well as their parents. Thank you to Marlen Kaufmann and Konstanze Schirmer for help with recruiting children and organizing the study. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 254142454/GRK 2070. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data and code necessary to reproduce the analyses presented here are publicly accessible on OSF (https://osf.io/qsyu9/). The materials necessary to attempt to replicate the findings are available from the corresponding author. The analyses presented here were not preregistered.

ORCID

Natalie Bleijlevens https://orcid.org/0000-0002-4421-4153
Nivedita Mani https://orcid.org/0000-0002-9843-4629
Tanya Behne https://orcid.org/0000-0001-7649-4949

REFERENCES

Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLoS One*, *7*, e49881. https://doi.org/10.1371/journal.pone.0049881

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*, 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three-and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034. https://doi.org/10.1016/j.cognition.2007.12.008

Birch, S. A. J., Akmal, N., & Frampton, K. L. (2010). Two-year-olds are vigilant of others' non-verbal cues to credibility. *Developmental Science*, *13*, 363–369. https://doi.org/10.1111/j.1467-7687.2009.00906.x

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, *18*, E46–E68. https://doi.org/10.1111/infa.12027

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400.

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

Crivello, C., Phillips, S., & Poulin-Dubois, D. (2018). Selective social learning in infancy: Looking for mechanisms. *Developmental Science*, *21*, e12592. https://doi.org/10.1111/desc.12592

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*, 521–536. https://doi.org/10.1037/dev0000083

Ganea, P. A., Koenig, M. A., & Millett, K. G. (2011). Changing your mind about things unseen: Toddlers' sensitivity to prior reliability. *Journal of Experimental Child Psychology*, *109*, 445–453. https://doi.org/10.1016/j.jecp.2011.02.011

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, *24*, 87–92. https://doi.org/10.1177/0963721414556653

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*, 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard University Press.

Hermes, J., Behne, T., Bich, A. E., Thielert, C., & Rakoczy, H. (2018). Children's selective trust decisions: Rational competence and limiting performance factors. *Developmental Science*, *21*, e12527.

Hermes, J., Behne, T., & Rakoczy, H. (2015). The role of trait reasoning in young children's selective trust. *Developmental Psychology*, *51*, 1574–1587. https://doi.org/10.1037/dev0000042

Hermes, J., Behne, T., & Rakoczy, H. (2018). The development of selective trust: Prospects for a dual-process account. *Child Development Perspectives*, *12*, 134–138. https://doi.org/10.1111/cdep.12274

Hermes, J., Rakoczy, H., & Behne, T. (2019). Making sense of conflicting information: A touchscreen paradigm to measure young children's selective trust. *Infant and Child Development*, *28*, e2119. https://doi.org/10.1002/icd.2119

Heyes, C. (2017). When does social learning become cultural learning? *Developmental Science*, *20*, e12350. https://doi.org/10.1111/desc.12350

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology*, *39*, 61–70. https://doi.org/10.1037/0012-1649.39.1.61

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*, 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., Pérez-Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science*, *17*, 965–976. https://doi.org/10.1111/desc.12187

Kim, S., Paulus, M., & Kalish, C. (2017). Young children's reliance on information from inaccurate informants. *Cognitive Science*, *41*, 601–621. https://doi.org/10.1111/cogs.12471

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*, 815–826. https://doi.org/10.1037/a0019664

Krogh-Jespersen, S., & Echols, C. H. (2012). The influence of speaker reliability on first versus second label learning. *Child Development*, *83*, 581–590. https://doi.org/10.1111/j.1467-8624.2011.01713.x

Kuzyk, O., Grossman, S., & Poulin-Dubois, D. (2020). Knowing who knows: Metacognitive and causal learning abilities guide infants' selective social learning. *Developmental Science*, *23*, e12904.

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, *78*, 1543–1558. https://doi.org/10.1111/j.1467-8624.2007.01082.x

Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science*, *21*, e12663. https://doi.org/10.1111/desc.12663

Mangardich, H., & Sabbagh, M. A. (2018). Children remember words from ignorant speakers but do not attach meaning: Evidence from event-related potentials. *Developmental Science*, *21*, e12544. https://doi.org/10.1111/desc.12544

Mascaro, O., & Morin, O. (2014). Gullible's travel: How honest and trustful children become vigilant communicators. In E. J. Robinson & S. Einav (Eds.), *Trust and skepticism* (pp. 69–82). Psychology Press.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*, 404–418. https://doi.org/10.1037/a0029500

Robinson, E. J., & Einav, S. (2014). *Trust and skepticism: Children's selective learning from testimony*. Psychology Press.

Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers. *Cognition*, *112*, 415–422. https://doi.org/10.1016/j.cognition.2009.06.005

Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers? *Journal of Child Language*, *30*(4), 905–924. https://doi.org/10.1017/S0305000903005828

Schütte, F., Mani, N., & Behne, T. (2020). Retrospective inferences in selective trust. *Royal Society Open Science*, *7*, 191451. https://doi.org/10.1098/rsos.191451

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*, 779–797. https://doi.org/10.1037/a0034191

Stephens, E. C., & Koenig, M. A. (2015). Varieties of testimony: Children's selective learning in semantic versus episodic domains. *Cognition*, *137*, 182–188. https://doi.org/10.1016/j.cognition.2015.01.004

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development*, *92*, e20–e38. https://doi.org/10.1111/cdev.13429

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science*, *25*, 1730–1738. https://doi.org/10.1177/0956797614540178

Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science*, *17*, 443–451. https://doi.org/10.1111/desc.12134

How to cite this article: Schmid, B., Bleijlevens, N., Mani, N., & Behne, T. (2024). The cognitive underpinnings and early development of children's selective trust. *Child Development*, *00*, 1–18. https://doi.org/10.1111/cdev.14073

APPENDIX

DETAILS REGARDING THE ADDITIONAL EYE-TRACKING DATA AND ANALYSES IN STUDY 1

Here, we report (1) whether children's gaze behavior differed depending on the reliability of the speaker during the accuracy exposure phase and during the novel label training and (2) whether children's proportion of looking at the reliable speaker during novel label training predicts their success in the interactive endorsement task.

Gaze behavior during the accuracy exposure phase and the novel label training

Here, we tested whether children's gaze behavior during the Accuracy exposure and the Novel label training differed between the labeling events by the reliable and unreliable speaker. Therefore, we ran two linear mixed effects models (one for the Accuracy exposure and one for the Novel label training). We predicted the total looking duration by speaker reliability, fixation (object vs. speaker), age group, and all of their possible interactions, and added random intercepts for participants.

During the accuracy exposure, the gaze behavior of both age groups did not differ depending on whether they watched the labeling events by the reliable speaker or by the unreliable speaker, neither for their looking at the speaker (2-year-olds: $b=-118.10$, $SE=156.20$, $p=.450$; 5-year-olds: $b=-266.30$, $SE=147.10$, $p=.070$) nor the object (2-year-olds: $b=202.50$, $SE=156.21$, $p=.195$; 5-year-olds: $b=208.84$, $SE=147.08$, $p=.156$). During the novel label training, no differences in gaze behavior were observed, either: Children's looking durations did not differ depending on the speaker reliability, neither while looking at the speaker (2-year-olds: $b=402.70$, $SE=220.60$, $p=.068$; 5-year-olds: $b=125.20$, $SE=208.00$, $p=.548$) nor the object (2-year-olds: $b=-30.50$, $SE=220.60$, $p=.890$; 5-year-olds: $b=-135.30$, $SE=208.00$, $p=.516$) (see Table A1).

Is success in the endorsement task predicted by looking times during the novel label training?

Here, we investigated whether children's looking times during the novel label training would predict their success at test (the eye-tracking or interactive endorsement task). Thus, we first calculated the proportion of time children looked at the training provided by the reliable speaker versus the reliable and unreliable speaker combined (aggregated across trials). For the interactive endorsement task, we then ran a generalized linear model with binomial error distribution: We predicted children's object selection by the proportion of looking to the reliable speaker during the label training, age group, and their interaction. For the eye-tracking endorsement task, we ran a generalized linear mixed model with beta error distribution: We predicted children's proportion of target looking by the proportion of looking to the reliable speaker during the label training, age group and their interaction, including random intercepts for participants.

Children's proportion of looking at the novel label training by the reliable speaker (vs. reliable and unreliable combined) was balanced for both 2-year-olds ($M=0.48$, $SD=0.06$) and 5-year-olds ($M=0.50$, $SD=0.02$). Importantly, this proportion did not predict their success in the subsequent endorsement tasks, neither in the eye-tracking (2-year-olds: $b=-5.73$, $SE=3.44$, $p=.095$; 5-year-olds: $b=-10.84$, $SE=11.52$, $p=.347$), nor the interactive version (2-year-olds: $b=4.30$, $SE=6.64$, $p=.517$; 5-year-olds: $b=61.20$, $SE=95.38$, $p=.521$).

TABLE A1 Children's mean looking times (in ms) during the accuracy exposure phase and label training in Study 1, separated by speaker role and age group.*

	Reliable speaker	Unreliable speaker
5-year-olds		
Accuracy exposure		
Looking at speaker	5254 (1307)	4987 (1350)
Looking at object	2555 (1022)	2763 (1161)
Total looking time	7809 (1210)	7750 (1102)
Label training		
Looking at speaker	4365 (1588)	4491 (1511)
Looking at object	3299 (1338)	3164 (1304)
Total looking time	7665 (1155)	7655 (1028)
2-year-olds		
Accuracy exposure		
Looking at speaker	4506 (1661)	4388 (1840)
Looking at object	3092 (1652)	3295 (1731)
Total looking time	7598 (1574)	7682 (1641)
Label training		
Looking at speaker	3985 (1708)	4388 (1765)
Looking at object	3137 (1672)	3106 (1667)
Total looking time	7122 (1883)	7494 (1490)

*The numbers in brackets are standard deviations.