Abstract

Hearing about surprising phenomena triggers exploration, even in young children. This exploration increases and changes with age. It becomes more targeted and efficient with children around 6-years-old clearly exploring with the intent to verify what they have been told. What underlies this development? In this study [anticipated total N=175, 48-84 months, 50% female], we tested the hypothesis that children's ability to reflect on the *causes* of their uncertainty about a surprising claim allows them to better target their empirical investigation of that claim—and that this ability increases with age. To test this developmental account, we assigned children to two conditions: a prompted and an unprompted condition. In each condition, children witnessed a series of vignettes where an adult presented them with a surprising claim about an object. Children were then asked whether they thought the claim was true or not, and how certain or uncertain they were in that belief. Then, in the prompted condition, children were asked *why* they felt that way. Finally, in both conditions, children were asked to recommend a course of action to determine whether the adult's claim was true or not. The findings from this study, revealed that [anticipated results: children express more uncertainty with age, older children also provide more reasons for their uncertainty, and controlling for children's ability to design an effective test, younger children recommend targeted empirical tests for a surprising claim at similar rates to older children, but only when prompted to reflect on the causes of their uncertainty]. This provides [support/some support/no support for the notion that developments in children's reasoning about their own uncertainty drive changes in their empirical evaluation of surprising claims].

Keywords: information seeking, uncertainty reasoning, testing claims, exploration

Deleted: TO TEST OR NOT TO TEST

Deleted: or witnessing

Deleted: 220

Deleted: two possible developmental mechanisms: (1) increases in children's ability to design an efficient test of what they have been told; (2) increases in children's metacognitive awareness of their skepticism about what they have been told—a development which may allow them to better empirically evaluate surprising claims

Deleted: these

Deleted: s

Deleted:

Deleted: presented children with a series of vignettes where they witnessed an adult present a surprising claim about an object to a child protagonist. Children were then asked to recommend a course of action for the child.

Deleted: To manipulate the degree to which children were prompted to test the claim, we varied within subject whether children had to generate a response by themselves, whether the protagonists' epistemic uncertainty was made explicit or not (to scaffold metacognitive awareness), and whether children could select between a set of pre-defined responses (to scaffold designing an empirical test)

Deleted: 4

Deleted: -5-year-olds rarely seek information to test a surprising claim, while 6-7-year-olds frequently do

Deleted: uncertainty monitoring and/or scientific reasoning drive changes in how children

Deleted: ly

Deleted: e

Deleted: I

Deleted: Scientific

Deleted: R

Deleted: T

Deleted: E

Knowing why: Children's reflection on their own uncertainty about an adult's surprising claim increases their tendency to test that claim

Across domains, older children appear to be more intentional in their actions (Siegler, 1996). Of particular interest to the current paper is how this increased intentional control develops when children are seeking information (Meder et al., 2021; Ruggeri et al., 2019), especially when such information seeking follows exposure to a surprising claim (Ronfard et al., 2018). The current study tests the hypothesis that older children's more targeted and efficient exploration following surprising claims may reflect a greater awareness of *why* they are skeptical about a surprising claim and that knowing *why* one is uncertain about a claim allows children to devise and implement more targeted information seeking strategies. In doing so, the current study advances our understanding of the transition from intuitive science (Gopnik, 2012) to explicit scientific thinking and reasoning (Kuhn, 2001).

Infants and young children seek information when they are uncertain. These information seeking behaviors are informative and facilitate belief revision (Gopnik, 2012). For example, infants will selectively explore a train they have just seen float in midair rather than play with a novel toy (Stahl & Feigenson, 2015), 20-month-old infants will ask for help when they are unsure where a toy is hidden (Goupil et al., 2016), and 4-year-old children engage in more exploratory play when evidence is confounded rather than unconfounded (Schulz & Bonawitz, 2007). These adaptive behaviors suggest an early emerging sensitivity to uncertainty that may be distinct from the ability to report on one's uncertainty (Baer & Kidd, 2022). In support of this hypothesis, Lapidow and colleagues (2022) found that 4- and 5-year-old children's explicit reports of their confidence about the presence of a target shape did not differ significantly when the shape was visible, partially hidden, or fully hidden. In contrast, when asked to choose which window to explore, children's exploration differed significantly with children most often choosing to explore the fully occluded shape. By

Deleted: TO TEST OR NOT TO TEST

Deleted: **To test or not to test: Uncertainty and information seeking following surprising claims** ¶

Deleted: we see that

Deleted: gain increasing

Deleted: control over

Formatted: Indent: First line: 0 cm

Deleted: Giron et al. 2022;

Deleted: On the one hand, children's awareness of their own uncertainty is likely to affect their motivation to seek clarifying information and their ability to select the most appropriate test for a surprising claim (*Uncertainty-awareness-hypothesis*). On the other hand, children's ability to design an effective test will determine whether the necessary information can be sought (*Scientific-reasoning-hypothesis*). In this study, we test these two developmental accounts. …

Deleted:

Deleted: most

Deleted: se

implication, young children's selective and informative information seeking in response to uncertainty may not need to (and may often not) reflect an explicit awareness of their uncertainty. However, children's ability to reflect on their uncertainty may influence *how* they seek information and therefore how efficiently they seek information when they feel uncertain.

Improvement in children's metacognitive abilities—the "ability to be aware of and contemplate thinking" (Kuhn, 2022, p. 73)—may allow children to improve the efficiency and effectiveness of their exploration. Indeed, with increasing age, children improve in their judgements and reasoning about their own uncertainty (Rohwer et al., 2012) alongside the efficiency of their exploration (Pelz & Kidd, 2020) and their reasoning about how to gather information (Piekny & Maehler, 2013). Between 3- and 7-years-old, children's uncertainty judgements become better calibrated (Baer & Odic, 2019; Rohwer et al., 2012), their ability to reason about uncertain causes improves (Fernback et al., 2012; Erb & Sobel, 2014), and they are increasingly able to reflect on the relation between beliefs and evidence (Astington et al., 2002). During the same time, children's exploration becomes increasingly attuned to the opportunities and constraints of a situation. For example, from 3-9-year-olds, children's exploratory search strategies become increasingly targeted to the task (Pelz & Kidd, 2020; Ruggeri et al., 2019), they also improve in their ability to identify the most informative questions (Mills et al., 2011; Ruggeri et al., 2017), and become more likely to exploit the environment to gather the information they need (Meder et al., 2021). Finally, children's ability to reason about how to test a claim improves. Between 4- and 8-years old children increasingly come up with relevant exploration strategies when explicitly asked to do so (Koksal-Tuncer & Sodian, 2018) and improve in their ability to generate and test hypotheses using unconfounded experimental designs (Koerber et al., 2015; Mayer et al., 2014; Piekny et al., 2014; Van der Graaf et al., 2015; Zimmerman & Klahr, 2018).

Deleted: TO TEST OR NOT TO TEST

Deleted: that

Deleted: , yet remain informative

Deleted:

Deleted: However, while being able to reflect on one's uncertainty may not impact *whether* children seek information when faced with uncertainty, it

Deleted: .

These data suggest an important role for children's metacognitive abilities for *how* they search for information. The current study seeks to test this connection by examining links between children's reasoning about their own uncertainty and their suggestions for whether and how to test a surprising claim. In a set of recent studies, researchers have investigated young children's exploratory behaviors following a surprising claim, finding that their exploration is generally informative and can lead to knowledge updating (Cottrell et al., 2022; Hermansen et al., 2020; Hermansen et al., 2021; Ronfard et al., 2018, 2021). Moreover, and consistent with the above referenced research on children's exploration following their *observation* of surprising phenomena, children's exploration following surprising *claims* becomes increasingly targeted and efficient. For example, in past research, 4- to 7-year-old children were presented with a set of 5 Russian dolls. Children were asked which of the dolls was the heaviest. Consistent with prior research demonstrating that even infants associate size with weight (Mounoud & Bower, 1974), all children replied that the biggest doll was the heaviest. Children were then assigned to have that intuition confirmed or contradicted. When their intuitions were contradicted, they were told that the smallest doll rather than the biggest doll was the heaviest. Children were then left alone with the dolls. Compared to children whose intuition were confirmed, children whose intuitions were contradicted selectively explored the dolls: on average, they were more likely to pick up the smallest doll and the biggest doll. However, only older children (6- and 7-years-old) engaged in targeted testing of the surprising claim by picking up the biggest and the smallest doll *at the same time*, a direct test of the surprising claim (Ronfard et al., 2021). This age change in the efficiency of children's investigation of a surprising claim was recently replicated on a third-party task where children were asked to reason about how another child should act after hearing a set of 8 surprising claims. These claims targeted different object properties and varied in whether they were simple claims, e.g., this [the smallest] object is very heavy, or comparative claims,

Deleted: TO TEST OR NOT TO TEST

Formatted: Left

Deleted: explored

Deleted: *under review*

Deleted: TO TEST OR NOT TO TEST

e.g., this [the smallest] object is the heaviest [compared to much larger objects]. Results revealed that with increasing age children made more targeted exploration recommendations—for example, suggesting that the child pick up only the target object for simple claims and suggesting that the child pick up the smallest and the biggest object for the comparative claims. With increasing age, children were also more likely to justify such exploration decision by expressing uncertainty about the truth of the surprising claim. Such reports of uncertainty were associated with children's suggestion of more targeted testing, controlling for age (Cottrell et al., 2022).

Deleted: However, although 3-6-year-olds disregard an informant's claim if they are presented with contradicting evidence (Hermansen et al., 2020), they do not consistently seek out evidence themselves to dispel their uncertainty if left alone with the target object(s) (Hermansen et al., 2021). Among those who do seek further information, only 5-6-year-olds make use of the information they acquire to solve a given task (Hermansen et al., 2021). This age change in children's information seeking following claims that contradict recent empirical evidence is also evident when children are faced with a claim that contradicts their prior intuitions (Cottrell et al., *under review*; Ronfard et al., 2018, 2021).

Why might children's report of their uncertainty about a surprising claim be associated with more efficient and targeted recommendations for testing that claim? Children's report of their uncertainty about the surprising claim in this prior study could reflect two distinct aspects of uncertainty: (1) the ability to accurately report feelings of uncertainty and (2) the ability to introspect on the causes of that uncertainty, specifically being able to provide an explanation for why the adult's claim is unlikely to be true. These two aspects are distinct because being uncertain about a claim does not tell you about why you feel this way (Baer & Kidd, 2022). Between the ages of 3- and 7-years-old, children become increasingly able to report on their own ignorance and uncertainty (Rohwer et al., 2012), with children from the age of 5 evincing increasingly precise reasoning about their own knowledge (Baer & Odic, 2019). During the same time period, children also become better able to engage in diagnostic reasoning about uncertain causes, with older children providing more correct assessments (Erb & Sobel, 2014) and being less sensitive to task constraints (Weisberg et al., 2020). Diagnostic reasoning is the ability to identify the cause of a phenomenon among a set of possibilities (Fermbach et al., 2012). Such reasoning is similar to being able to identify the causes of one's uncertainty about a claim as identifying the source of one's uncertainty requires identifying the most likely explanation for why the claim

Deleted: Older children's propensity to verify a claim may be linked to developments in two important aspects of cognitive development

Deleted: 1) metacognition, being aware of the uncertainty elicited by a surprising claim, and 2) scientific reasoning, being able to come up with an effective test for a claim

Deleted: TO TEST OR NOT TO TEST

is wrong. Being able to identify the reasons for one's skepticism about a claim is likely to be a strong predictor of one's ability to design an informative test of that claim given that explanations drive children's exploratory behaviors (see Legare, 2012 for a review). Thus, we make four main predictions: (1) with increasing age, children will express more uncertainty about the possibility of a surprising claim; (2) with increasing age, children will be more likely to provide a plausible reason for their uncertainty about the possibility of a surprising claim; (3) with increasing age, children will be more likely to suggest targeted empirical tests for a claim—tests that provide the needed evidence to confirm or disconfirm the truth of a claim; (4) prompting children to reflect on the causes of their uncertainty about a claim will increase the likelihood that they generate an efficient test of that claim. This effect of prompting is expected to be stronger among younger children.

Deleted: For example, when presented with the claim that a golf ball bounces better than a tennis ball, a child who drops only the golf ball on the floor (single object exploration) may be guided by skepticism towards said qualities of the ball but explores in a manner that would *not* provide evidence for or against the claim that the golf ball bounces higher than the tennis ball. On the other hand, a child who drops both balls on the floor (comparative exploration) will possess the evidence needed to verify what they were told. Children's engagement in single rather than comparative object exploration could result from a lack of meta-cognitive awareness that one is skeptical of a claim. In this case, children may explore the object mentioned in the claim (because it is surprising) but lacking an awareness of their skepticism will not implement an exploration strategy designed to assuage their doubts (*Uncertainty-awareness-hypothesis*). Alternatively, children may have the metacognitive awareness that they do not believe a claim but lack the ability to design an empirical test that could empirically test it (*Scientific-reasoning-hypothesis*)

To test these hypotheses, we assigned children to two conditions: a prompted and an unprompted condition (see Figure 1 for an illustration of the overall procedure, and Figure 2 for an illustration of a single trial). In each condition, children were presented with 4 surprising claims. For example, they were shown pictures of three rocks of increasing size. They then heard a speaker say that the smallest rock is much heavier than the other rocks before being asked whether they think that this is true or not true (Belief question) and whether they are sure or not sure about this (Uncertainty question). Then, in the prompted condition only, children were asked why they feel this way (Reasoning question). These responses were coded for the presence or absence of a plausible explanation, for example, the explanation that "bigger objects are typically heavier than smaller objects". Finally, for both conditions, children were asked whether they thought it might be worth to try and determine whether the adult's claim was true or not (Explore question), and if so, how they would proceed (Design question). Children's responses were coded based on whether children's exploration suggestions would provide evidence that the adult's claim is true, and whether

Deleted: Both hypotheses are motivated by previous research. Support for the *Scientific-reasoning-hypothesis* comes from the fact that children become more skilled in their scientific reasoning during the preschool years, showing gradual improvements in their ability to evaluate evidence, experiment, and generate hypotheses (Koerber et al., 2015; Mayer et al., 2014; Piekny & Mahler, 2013). By the age of 4, children are able to evaluate available evidence when the task does not require prior knowledge (Piekny et al., 2014; Piekny & Mahler, 2013), and when required to evaluate patterns of covariation (Koerber et al., 2005). However, when presented with non-covariation evidence, even 6-year-olds require prompts to perform adequately (Koerber et al., 2005). Furthermore, studies have also shown that while 4-year-olds sometimes engage in experimentation (Van der Graaf et al., 2015), their understanding of experimentation continues to increase rapidly from 5 to 6 years of age (Piekny et al., 2014) with concomitant increases in their abilities to generate and express hypotheses (Koerber et al., 2015; Mayer et al., 2014). ¶
However, although the ability to devise an effective empirical test for a claim is a necessary skill for verifying surprising claims, it may not be sufficient. For children to strategically test a surprising claim—for example, the claim that a golf ball is actually bouncier than a tennis ball—it might be necessary for children to be explicitly aware of their epistemic uncertainty about a claim. This is because such awareness may allow children to better target their exploration. Support for the *Uncertainty-awareness-hypothesis* comes from research showing that children's early exploratory behavior indicates a sensitivity to uncertainty, albeit implicitly, with young children preferring to explore surprising events (Shulz 2012: Schulz & Bonawitz, 2007) or perceptually occluded stimuli (Lapidow et al., 2022). Evidence for the notion that these exploration decisions do not reflect an explicit awareness of uncertainty comes from the fact that Lapidow and colleagues (2022) found no association between 4- to 5-year-old children's reports of their uncertainty and their exploration decisions. Thus, while children`s early exploration may be guided by an implicit sense of uncertainty, their lack of a conscious awareness of this uncertainty may limit their chances of dealing with a surprising claim in a strategic manner. This interpretation is supported by the recent correlational data suggesting that, with increasing age, children increasingly justify their exploration as a way to test if a claim is true (Cottrell et al., under review). ¶

they did so inefficiently (e.g., they suggested gathering either too little or too much data) or efficiently (e.g., they suggested gathering enough data and the right type of data). In the preceding example, this would mean comparing the weight of all of the rocks which is the only way to know if the smallest rock is the *heaviest* of them all.

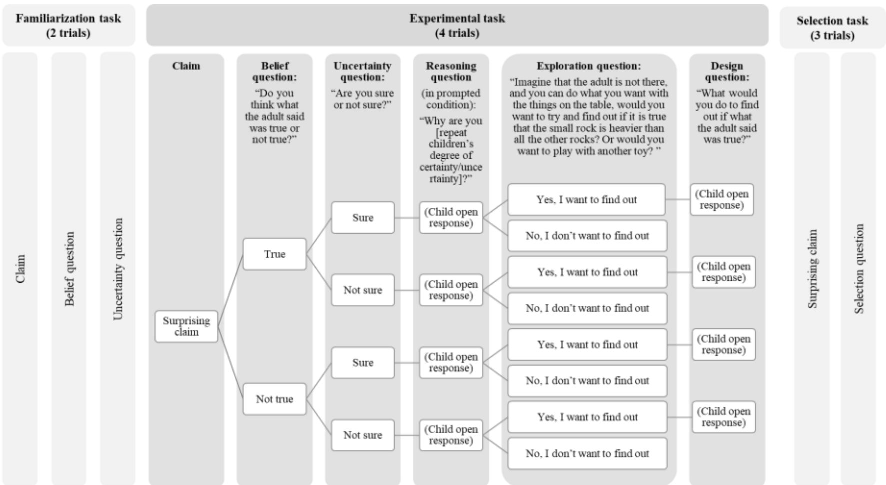


Figure 1. Illustration of overall testing procedure. Note that the Reasoning question is only presented to children in the prompted condition.

To test our four hypotheses, we will conduct the following confirmatory analyses. We will assess prediction (1) by combining the certainty response of children across both conditions and testing whether children express more uncertainty with age. Prediction (2) will be tested by looking at whether older children in the prompted condition provide more plausible reasons for their skepticism than younger children. Prediction (3) will be tested by looking at whether older children in the unprompted condition provide more targeted empirical tests than younger children. Prediction (4) will be assessed by testing whether children in the prompted condition are more likely to provide a more targeted empirical test than children in the unprompted condition.

Deleted: TO TEST OR NOT TO TEST

Deleted: ing to

Deleted: suggest

Deleted: ing

Deleted: to

Formatted: Font: Italic

Formatted: Font: Bold

Of course, children's awareness of the uncertainty of a claim may not be sufficient for children to engage in targeted testing of that claim. Children's ability to design an effective test will determine whether the necessary information can be sought. Thus, at the end of the experiment, we presented children with 3 additional surprising claims, this time with three options for assessing the truth of those claims. Children were asked to select the alternative that best enables them to test whether the adult's claim is true or not (Select question). Given prior work showing that 4-year-old children are able to select the most informative test to understand a causal system (Lapidow & Walker, 2020) and to select the most informative question when presented with multiple options (Ruggeri et al., 2017), our final prediction (5) is that we expect all children in our study to do so. However, this additional test is important to rule out the possibility that age related differences in the type of empirical test suggested by older children does not simply reflect age-related differences in knowledge of testing strategies.

Methods

[Note that we wrote this section in past tense to indicate what these sections would look like after data collection. No data collection took place yet.]

Participants

The final sample consisted of [anticipated N=175 (M_{age}=XX years, Range = 4-7 years, 50% female)]. An additional [N=X] children were tested but were excluded from the final analyses due to the following: 1) Withdrawal from study [N=X], 2) Technical or experimenter error [N=X], and 3) Less than one response to a trial where response options are provided (i.e., the Belief question, the Uncertainty question, the Explore question, and/or the Select question) [N=X]. To enable a diverse sample from rural and urban areas, participants were recruited through childcare centers across the country, as well as through social media channels, with

Deleted: TO TEST OR NOT TO TEST

Formatted: Left

Deleted: In this study, we aim to provide causal evidence for the link between children's uncertainty monitoring and scientific reasoning, and their strategic exploration. We investigate the impact of these two factors by experimentally manipulating the degree to which we scaffold children's uncertainty awareness and scientific reasoning. Specifically, we present children with a series of vignettes where an adult informant presents a claim that runs counter to what a child protagonist believes. The *Uncertainty-awareness-hypothesis* (Hypothesis 1) predicts that younger children are more likely to suggest comparative rather than single object exploration when the protagonist's uncertainty is stated (Uncertainty Scaffolding) rather than not stated (No Scaffolding). Alternatively, if younger children recognize the uncertainty generated by those claims but do not know how to test them, the *Scientific-reasoning-hypothesis* (Hypothesis 2) predicts that younger children are more likely to identify the sufficient test when this is explicitly presented to them among a set of options (Uncertainty and Strategy Scaffolding) as compared to when they have to generate a response themselves (Uncertainty Scaffolding). If developments in uncertainty monitoring *and* scientific reasoning underlie age-related changes in children's testing of surprising claims, the *Uncertainty-reasoning-hypothesis* (Hypothesis 3) predicts a progressive increase in younger children's inclination to test claims with increasing scaffolding (No Scaffolding < Uncertainty Scaffolding < Uncertainty and Strategy Scaffolding)…

Deleted: preregistration or

Deleted: 220

Deleted: the Uncertainty and Strategy Scaffolding condition

Deleted: TO TEST OR NOT TO TEST

[insert description of sample diversity, parental education level and parental income]. Informed consent was obtained from the child's parents in advance of testing.

We base our sampling plan on a power analysis, using effect size information from previous studies with a similar design and age range (Hermansen et al., 2021; Ronfard et al., 2021). Using GPower 3.1 (Faul et al., 2007), this analysis revealed that to detect a low-medium to medium effect (f^2=.08-.15), with 80% power at alpha = .05, the analyses associated with the most resource demanding analysis will require a sample of 175 participants. See attached R-script for details on these estimates.

Formatted: Superscript

Deleted: analysis will

Ethics and data handling

The overall project was approved by the internal ethical research committee (Department of Psychology, University of Oslo, no.: 16842024) and the local authorities on data protection (NSD, no.: 843823), and supported by a departmental research grant to the first author from the Department of Psychology, University of Oslo. Person-identifying video recordings from the testing session are stored in the University's internal, secure storage system TSD, accessible only to the two first authors and research assistants associated with the two first authors. Anonymized data supporting the findings of the study are openly available in the Open Science Framework ([URL, reference number]), together with study stimuli and analysis scripts.

Procedure

Parents agreeing to take part in the study with their child were invited to perform the tasks through an online web-portal at a time of their convenience. One week prior to their scheduled participation, parents received a short online questionnaire on demographic background information, and instructions as to the procedures of the main testing session. On the day of testing, parents were asked to log onto the online platform together with their child, from which the experimenter introduced the Familiarization task, the Experimental task, and the Selection task. The session was video recorded for later coding.

A **Familiarization task** was presented to all children prior to starting the main experiment. We used this task to inform children about what was about to happen (i.e., hear a set of claims, and be asked a set of questions about what they believe and how certain they are), and that they now had the opportunity to practice two times first.

For the **Experimental task**, all children were assigned to one of two conditions of four trials: a prompted and an unprompted condition. In each condition, an adult informant presented children with a surprising claim about an object in front of them on the screen, following which the experimenter asked the child: 1) whether they believed the claim or not (Belief question), and 2) how certain or uncertain they were in their belief (Uncertainty question). In addition, children in the prompted condition only were asked to reflect on *why* they felt the way they did about their belief (Reasoning question). Finally, children in both conditions were asked whether they wanted to figure out whether the adult's claim was true or not (Explore question), and if so, how they would go about doing so (Design question). The type of trial and placement of the referent object was counterbalanced across trials and participants.

To conclude the session, children were presented with the **Selection task** to assess their ability to identify an effective test. In this task, children were again presented with a series of three surprising claims, but rather than asking children whether they want to find out the truth about the claim and if so how, they were simply asked to select between a set of three pre-specified options as to how they might go about testing the claim. The options varied in the degree to which the claim would be sufficiently tested, allowing children to select between manipulating the referent object, the referent object and one of the alternative objects, or the referent object and both the alternative objects. The order of the action options was

Deleted: TO TEST OR NOT TO TEST

Deleted: experimental

Deleted: All children were assigned to one of three blocks of nine trials (three trials per condition), presenting children with a surprising claim and followed by a request to suggest a course of action (see Appendix 1 for an overview of the three fixed trial sequences). All children were exposed to all three conditions in a pseudorandomized order (No Scaffolding before Uncertainty Scaffolding before Uncertainty and Strategy Scaffolding), but with the content of the claims varied across trials. To avoid potential issues of fatigue, two brief distractor tasks were administered between the three conditions. ¶
Experimental procedure¶

Deleted: accompanied by a Belief and an Uncertainty question, counterbalanced across trials and participants. To ensure that the trials used in the selection task was not perceived differently to the trials used in the experimental task, all trials for both tasks were randomly selected from the same pool of trials.

Familiarization task

During the familiarization task, children were presented with two trials containing a familiar object, a blue ball and a yellow box, and faced with a claim that was either easy to be sure of (e.g., that a blue ball is blue), or something to be less sure of (e.g., that a yellow box is pink) (see Appendix 1 for details of the stimuli and procedure). In the first trial, children were presented with a claim about an object's which matched its visible properties: "Lets imagine that you see a picture of a ball—like this one—and I tell you that it is blue (a blue ball is shown on the screen). Do you think it is true or not true that this ball is blue?" To guide the child in how to respond, the informant followed up by saying: "If you think it is true, you can click on the green button, and if you think it is not true you can click on the red button." After the child responded, the experimenter went on to ask the child: "Are you sure or not sure?" Again, to guide the child in how to respond, the informant followed up by saying: "If you are sure, you can click on the green button, and if you are not sure you can click on the red button." If the child did not respond or answered something not in line with the task on either of these two questions, the experimenter repeated the question once before moving to the next practice trial. In the second trial, children were presented with a claim about an object's which did not match its visible properties: "Lets imagine that you see a picture of a box—like this one—and I tell you that it is pink (a yellow box is shown on the screen). Do you think it is true or not true that this box is pink?" followed by the same instructions on how to respond as in trial one. After the child responded, the experimenter went on to ask the child: "Are you sure or not sure?" followed by instructions on how to respond. If the child did not respond or answered something not in line with the task on either of these two questions, the experimenter repeated the question

Deleted: TO TEST OR NOT TO TEST

Deleted: In the main experiment, each trial consisted of three phases, played out in a fixed order, following a predefined script (see Appendix 1). In each trial, children were presented with a scenario in which a child protagonist hears a surprising claim from an adult informant about one of three objects visible in front of this child character. For example, presenting a claim that runs counter to what the child protagonist thinks about the relative weight of small objects as compared to big objects, the adult informant referred to a small rock next to a large rock saying: "This small rock is much heavier than the big rock". There was also a non-target object present in the scene, such as a toothbrush. To maximize the perceived level of surprise, and to level children's prior knowledge about the subject matter (Piekny et al., 2014), the protagonist's initial belief was made explicit at the beginning of the trial. Considering previous work showing that young children are sensitive to an informant's knowledge (Butler & Markman, 2012), the claim was presented without any reference to knowledgeability. After providing the claim, the adult excused themselves from the scene, leaving the child protagonist alone and free to play with the objects. Following each of the scenarios in the first three trials (No Scaffolding), the experimenter asked what the child believed that the protagonist should do, saying: "Now the child is alone, what do you think they should do?". In the next three trials (Uncertainty Scaffolding), children were alerted to the protagonist's uncertainty, with the experimenter saying for example: "Now the child is alone, and the child is not sure if the small rock is actually heavier than the big rock. What do you think they should do?". Following each of the scenarios in the final three trials (Uncertainty and Strategy Scaffolding), children were again prompted to the protagonist's uncertainty, but rather than generating their own action suggestions, they were asked to choose between three options that reflected either: 1) exploration of both relevant target objects; or 2) exploration of only one of the relevant target objects; or 3) exploration of aspects of the scenario irrelevant to the claim. Across all nine trials, if children did not select any of the action options, the experimenter moved on to the next scenario after having repeated the question once. Given that children may vary in the extent to which they find it acceptable to question an adult's assertions (Chen & French, 2008; Corriveau et al., 2013), the child participant was asked to suggest a course of action on behalf of the protagonist, rather than to act themselves, and only when the adult informant had left the scene. Furthermore, presenting the task in a third-person format allowed us to encourage children's explicit representation of epistemic uncertainty also in the trials where no scaffolding is provided. Illustrations of a trial from each of the three conditions is presented below in Figure 1. Note that each child was presented with a selection of different claims across the trials and Figure 1 is only meant as an example of the general experimental procedure and to highlight key differences between the three conditions. ¶
¶ ...

Deleted: *Warm-up*

Deleted: To

Deleted: e

once before moving to the next practice trial. If the child never responded to any of these preliminary questions, the experimental session ended.

Experimental task

In each trial of the main experiment, children were presented with a scenario in which an adult informant presents a surprising claim about one of three objects visible on screen followed by a set of questions (for the details of the full script, see Appendix 1). For example, presenting a claim that runs counter to what children typically think about the relative weight of small objects as compared to big objects, the adult informant referred to a small rock next to two larger rocks saying: "This small rock is much heavier than all the other rocks". Considering previous work showing that young children are sensitive to an informant's knowledge (Butler & Markman, 2012), the claim was presented without any reference to knowledgeability. After hearing the informant's claim, the experimenter first asked the child the Belief question: "Do you think what the adult said was true or not true?" before asking the child the Uncertainty question: "Are you sure or not sure?" Next, children in the prompted condition were asked the Reasoning question: "Why are you [repeat children's degree of certainty/uncertainty]?" Finally, children in both conditions were asked the Exploration question: "Imagine that the adult is not there, and you can do what you want with the things on the table, would you want to try and find out if it is true or not true that the small rock is heavier than all the other rocks?" Following a confirming response to this question, they were asked the Design question: "Can you tell me what you would do to find out if what the adult said was true or not true?" Across all trials, if children did not respond to a given question, the experimenter moved on to the next scenario after having repeated the question once. Illustrations of a trial from each of the two conditions is presented below in Figure 2. Note that each child was presented with a selection of different claims across the trials and Figure 2 is

Deleted: TO TEST OR NOT TO TEST

Deleted: with the situation, the experimenter presented children with an image of a farmyard with a range of different animals and asked them to identify the type of animal gathered around the farmhouse (pigs) and to count the number of cows (eight)

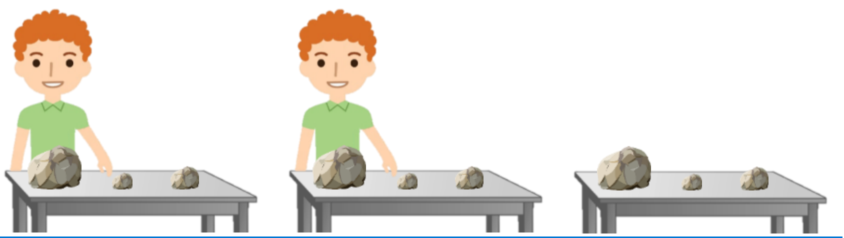
Deleted: 1

Deleted: 1

only meant as an example of the general experimental procedure and to highlight the key difference between the two conditions (or a full overview of the trial types, see Appendix 1).

Selection task

To control for individual differences in children's ability to identify an efficient test of a surprising claim, children were presented with three additional surprising claims along with three options for assessing the truth of those claim and asked to select which option they would use to find out if what the adult said is true. Thus, rather than generating their own action suggestions, they were asked to choose between three options that reflected inefficient non-comparative exploration (1), inefficient comparative exploration (2), or efficient exploration (3) (for details, see Appendix 1). Across all trials, if children did not select any of the action options, the experimenter moved on to the next scenario after having repeated the question once.



1. Experimenter says: "This adult says that: "This small rock is much heavier than the big and the medium rock.""

2A. Experimenter asks: "Do you think what the adult said is true or not true?"

2B. Experimenter follows by asking: "Are you sure or not sure?"

3A. Experimenter asks: "Imagine that the adult is not there, and you can do what you want with the things on the table, would you want to try and find out if it is true that [repeat the adult's claim]? Or would you want to play with another toy?"

3B. If yes to the first question, experimenter follows by asking: "Can you tell me what you would do to find out if what the adult said was true?"

Deleted: TO TEST OR NOT TO TEST

Deleted: *Distractor*

Deleted: *s*¶

Formatted: Indent: First line: 0 cm

Formatted: Indent: First line: 1,27 cm

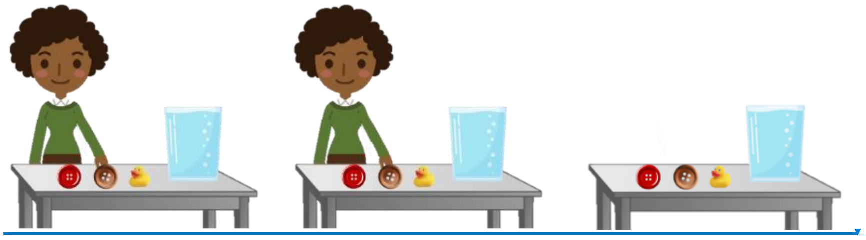
Deleted: In the first distractor task, children were presented with an image of multiple sheep in a meadow of various colors and asked to count the number of white sheep. In the second distractor task, children were presented with an image of a street map with cars and asked to count the number of yellow cars…

Deleted: all

Deleted: the other

Deleted: s

Deleted: the small rock is heavier than all the other rocks



1. Experimenter says: “This adult says that: “This wood button sinks much faster than the rubber duck.””	2A. Experimenter asks: “Do you think what the adult said is true or not true?” 2B. Experimenter follows by asking: “Are you sure or not sure?” 2C. And finally: “Why are you [repeat children’s degree of certainty/uncertainty]?”	3A. Experimenter asks: “Imagine that the adult is not there, and you can do what you want with the things on the table, would you want to try and find out if it is true that [repeat the adult’s claim]? Or would you want to play with another toy?” 3B. If yes to the first question, experimenter follows by asking: “Can you tell me what you would do to find out if what the adult said was true?”

Figure 2. Examples of experimental procedure in unprompted (top) and prompted (bottom) condition, showing two different comparative claims: one comparing the target object to all the other objects (top) and the other comparing the target object to one of the other objects (bottom).

Data processing

Coding

Video-recordings from the three tasks will be coded post-testing according to the below categories for the different questions by the second author and a research assistant blind to the study’s hypotheses, with one coding all videos and the other coding 20% to allow for reliability estimation. Reliability was assessed as percentage agreement, with agreement above 75% on each item considered acceptable. Reliability below 75% lead to a reassessment of the coding manual, and a new set of videos will be coded by two research assistants before a new estimate of reliability was calculated. In the event of coder discrepancies, differences due to coding

Deleted: TO TEST OR NOT TO TEST

Deleted:

Deleted: metal button

Deleted: the wood button sinks faster than the metal button

Deleted: 1

Deleted: ¶
Hypothesis overview¶
Hypothesis 1 (H1): Testing the notion that younger children struggle to explicitly represent the uncertainty created by the surprising claim, and the need for testing it, the *Uncertainty-awareness-hypothesis* predicted an interaction between Condition (No Scaffolding versus Uncertainty Scaffolding) and Age (illustrated in Figure 2). We expected that younger children would be significantly more likely to spontaneously recommend that the child protagonist *Test* the surprising claim in the Uncertainty Scaffolding trials (trials 4-6) compared to the No Scaffolding trials (trials 1-3). For older children, we expected to see little difference in testing between the two conditions. ¶
Hypothesis 2 (H2): Testing the alternative notion, that younger children struggle to identify the appropriate test to reduce their uncertainty, the *Scientific-reasoning-hypothesis* predicted an interaction between Condition (Uncertainty Scaffolding versus Uncertainty & Strategy Scaffolding) and Age (illustrated in Figure 2). We expected that younger children would be more likely to suggest that the child protagonist *Test* the surprising claim in the Uncertainty and Strategy Scaffolding trials (trials 7-9) compared to the Uncertainty Scaffolding trials (trials 4-6). For older children, we expected to see little difference in testing between the two conditions.¶ ...

Deleted: experiment

Deleted: twoa

Deleted: s

errors were corrected, while discrepancies due to coder disagreement were resolved through discussion with the first author.

Belief question. Children's responses to the Belief question was coded based on whether they said they thought the claim was true (0) or not true (1). Responses that were not relevant for the question were coded as 2, but discarded from the main analyses. A lack of response was coded as missing. Following the individual coding of each of the four trials in the Experimental task, and excluding responses coded as 2 or missing, we created an average score to reflect children's propensity to believe the surprising claim.

Uncertainty question. Children's responses to the Uncertainty question was coded based on whether they were sure (0) or not sure (1) in their beliefs. Responses that were not relevant for the question were coded as 2, but discarded from the main analyses. A lack of response was coded as missing. Again, excluding responses coded as 2 or missing, an average score was created based on the four trials in the Experimental task.

Reasoning question. We coded children's spontaneous responses to the Reasoning question into two broad categories depending on whether they provided a response that reflected the absence (0) or presence (1) of a plausible explanation. That is, a response was considered a plausible reason if including a description of a mechanism (e.g., saying "some things can be filled with stuff") or statistical regularity (e.g., saying "bigger things are often heavier") relating to the objects at hand, while responses that did not include a reference to such characteristics (e.g., saying "maybe it just is" or "I don't know") where considered lacking a plausible explanation. A lack of response was coded as missing. For children in the prompted condition, an average score was created based on the four trials in the Experimental task. For children in the unprompted condition, this average core was coded as missing.

Exploration question. Children's responses to the Exploration question was coded based on whether they said they did not want to find out the truth of the claim or simply wanted

Deleted: TO TEST OR NOT TO TEST

Deleted:

Deleted: The same procedure was followed for the three trials in the Selection task.

Deleted: one

Deleted: , and another was created based on the three trials in the Selection task

Deleted: TO TEST OR NOT TO TEST

to play (0) or whether they said they wanted to find out if the claim was true (1). Responses that were not relevant for the question were coded as 2, but discarded from the main analyses. A lack of response was coded as missing. An average score was created based on the four trials in the Experimental task, excluding responses coded as 2 or missing.

Design question. Children's spontaneous responses to the Design question was first coded into three broad categories depending on whether they provided a response that reflected inefficient non-comparative (1), inefficient comparative (2), or efficient (3) exploratory behaviors. Efficient exploration including descriptions of a comparative manipulation of all relevant objects only (e.g., lifting all rocks to assess which is the heaviest of them all, or dropping the wood button and the metal button into the water container to assess which of the two sinks faster). Inefficient relevant exploration encompassing responses that are comparative but reflect too much exploration (e.g., exploring all objects, when the claim relates to only two of the objects) or too little exploration (e.g., exploring two of the objects when all three objects require assessment), and inefficient irrelevant exploration involving only suggestions to explore a single key referent object when a comparative assessment is required. Children who suggested irrelevant exploratory behaviors unrelated to the claim (e.g., suggesting to smell a stone) or who expressed that they did not know were coded as 0. A lack of response was coded as missing. For the purpose of the main planned analysis, we the recoded the Design-variable by focusing on the contrast between children's suggestions of inefficient (0) versus efficient exploration (1), collapsing across the inefficiency categories and creating an average score based on the four trials in the Experimental task, excluding responses coded as missing. We make this simplification to highlight the main argument of our paper, while maintain the option to run future exploratory analyses of children's patterns of exploration.

Deleted: two

Selection question. Children's responses to the Selection question was first coded in a similar vein as their responses to the Design question, and based on whether they selected the option reflecting inefficient non-comparative (1), inefficient comparative (2), or efficient (3) exploratory behaviors. A lack of response was coded as missing. For the purpose of the main analysis, an average score was created based on the four trials in the Experimental task, excluding responses coded as missing. An average score was created based on the three trials in the Selection task, excluding responses coded as missing.

Exclusion

To be considered eligible for inclusion in the main analysis, children are required to provide at least one response to a question where response options are readily available (i.e., Belief question, Uncertainty question, Explore question, and Selection question). A complete lack of response on these questions will be considered a failure to comply with the task. A lack of responses to a trial without such response options (i.e., Reasoning question or Design question) will not be considered a failure to comply with the task, as some children are expected to struggle to spontaneously generate a response for these questions. We will exclude all cases where there is a disruption of the trial due to technical or experimenter error, or if the parent or child decided to withdraw from the experiment.

Overview of main hypotheses and planned statistical tests

Details about the planned statistical models are presented in the attached R-script, together with a simulated data set and the power analyses used to determine sufficient a sample size.

Discussion

Deleted: TO TEST OR NOT TO TEST

Deleted: The key variable of interest is whether children recommend an action that reflects an intent to test the claim or not. Because the child protagonist in the scenario was presented with a comparative claim, the child is required to suggest multiple object exploration, as opposed to single object exploration to be coded as testing the claim. In the Uncertainty and Strategy Scaffolding condition, these alternatives were made readily available to the child through the three response options. In the No Scaffolding and Uncertainty Scaffolding conditions however, we will code the child's spontaneous responses according to the following four categories: 1) *Test*: Child suggested that the protagonist should manipulate the key object as well as an alternative relevant object (e.g., the target rock and the alternative rock); 2) *Explore*: Child suggested that the protagonist should manipulate either the key object or the alternative, but did not both (as would be necessary for comparison); 3) *Other*: Child suggested that the protagonist should manipulate the alternative relevant object (e.g., an alternative rock), or alternative non-relative object (e.g., the sponge), but did not mention anything about manipulating the target object (e.g., the target rock). Alternatively, the child talked about the objects on the screen, but without suggesting any form of exploratory behaviors for the protagonist to engage in, or the child said explicitly that the protagonist did not need to do anything; 4) *No response*: Child made no response to the task

Deleted: a response to at least one of the Uncertainty and Strategy Scaffolding trials

Deleted: A lack of response on all three trials of the Uncertainty and Strategy Scaffolding condition will be considered a failure to comply with the task.

Deleted: the No Scaffolding or Uncertainty Scaffolding

Deleted: trials

Deleted: ***P***

Deleted: ***and power estimates***

Formatted: Indent: First line: 0 cm

Deleted: *Sample size*¶
We base our sampling plan on a power analysis, using effect size information from previous studies with a similar design and age range (Hermansen et al., 2021; Ronfard et al., 2021). Using GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), this analysis revealed that to detect a medium effect (w =.30), with 95% power at alpha = .05, the analyses associated with Hypothesis 1 and 2 will require a sample of 191 participants. However, given that the analyses associated with Hypothesis 3 requires a sample of 220 participants, we will continue recruitment and testing until we have data from 220 eligible participants (see above plans for exclusion). ¶
Analysis¶
To test Hypothesis 1, children's responses on No Scaffolding and Uncertainty Scaffolding trials are coded as '1' if the suggested action reflects an intent to *Test* the claim (i.e., the manipulation of both target objects), and '0' for any other suggestions (i.e., including the above categories *Explore*, *Other*, and *No response*), this is the dependent variable. We will use a multi-level logistic regression to test the probability that children decide that gathering further information is helpful to dispel the uncertainty (within subject factor is Uncertainty scaffolding (2: No Scaffolding vs. Uncertainty ...

Formatted: Font: Not Bold

Table 1. Proforma study design template.

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
Q1: Do children express more uncertainty about the possibility of a surprising claim with increasing age?	H1: With increasing age, children will express more uncertainty about the possibility of a surprising claim.	We base our sampling plan on a power calculation, using GPower 3.1 (Faul et al., 2007), indicating a need for ca. 175 participants to meet the criteria of the most resource demanding analyses.	Linear regression • Dataset: Full • Dependent variable: Uncertainty_Average • Independent variable: Age	Effect size estimates used in the power analyses are derived from previous studies with a similar design and age range indicating that we should expect low-medium to medium effects (f^2 =.08-.15)(Hermansen et al., 2021; Ronfard et al., 2018, 2021). We use standard norms to set alpha= .05 and power = 80%.	H1 is supported if the effect of Age is significant below the threshold of alpha = .05. If Age is not a significant predictor, this questions the age change observed in prior studies showing an increased sensitivity to the uncertainty of surprising claims.	Older children do not express increased rates of uncertainty about the possibility of a surprising claim compared to younger children.
Q2: Do children express more plausible reasons for their uncertainty about the possibility of a claim with increasing age?	H2: With increasing age, children are more likely to provide a plausible reason for their uncertainty about the possibility of a surprising claim.		Linear regression • Dataset: Prompted condition only • Dependent variable: Reasoning_Average • Independent variable: Age		H2 is supported if the effect of Age is significant below the threshold of alpha = .05. If Age is not a significant predictor, this questions the age change observed in prior studies showing improvements in their	

Deleted: TO TEST OR NOT TO TEST
Deleted: Is younger children's testing of claims more
Deleted: The *Uncertainty-awareness-hypothesis* …1 is
Deleted: Prompts to the uncertainty of surprising claims
Deleted: a…power analysis…alculation, using GPower 3
Formatted
Formatted
Formatted
Deleted: only …ge is not a significant predictor, this
Deleted: To assess the impact of uncertainty prompts on
Deleted: *Uncertainty-awareness-hypothesis* (Hypothesis 1
Deleted: ¶
Deleted: Is younger children's testing of claims more
Formatted
Deleted: The *Scientific-reasoning-hypothesis*
Deleted: Prompts to testing strategies for surprising claim
Formatted
Formatted
Deleted: the Chi square test is significant below the
Formatted
Deleted: Using GPower 3.1 (Faul, Erdfelder, Lang, &
Deleted: Using the same coding as above, we will run a
Formatted
Deleted: To detect a medium effect (w =.30), with 95%
Formatted
Deleted: To detect a medium effect size (w =.30), with 95
Deleted: *Scientific-reasoning-hypothesis* (Hypothesis 2):

KNOWING WHY

					ability to express reasons for their uncertainty.	
Q3: Are children more likely to suggest targeted empirical tests for a claim with increasing age? And does this effect hold also when controlling for children's ability to select an efficient test?	H3: Children are more likely to suggest targeted empirical tests for a claim with increasing age. And this is not driven by their ability to select an efficient test		Linear regression • Dataset: Full • Dependent variable: Reasoning_Average • Control variable: Condition • Independent variable: Age • Covariate: Selection_Average		H3 is supported if the effect of Age is significant below the threshold of alpha = .05 If Age is not a significant predictor, this questions the age change observed in prior studies showing that older children are more likely to test or suggest a test for a surprising claim.	Prompts to uncertainty alone does not increase younger children's testing of a surprising claim
Q4: Does prompting children to reflect on their uncertainty increase the likelihood that they generate an efficient test for that claim?	H4: Prompting children to reflect on the causes of their uncertainty about a claim will increase the likelihood that they generate an efficient test of that claim. This effect of prompting is		Linear regression • Dataset: Full • Dependent variable: Design_Average • Independent variables: Condition, Age, Condition BY Age • Covariate: Selection_Average		H4 is supported if the main effect of Condition is significant below the threshold of alpha = .05, and/or there is a significant Condition BY Age interaction. If there is only an effect of Age, this suggest that reasoning about the causes of	Prompts to reason about their uncertainty about surprising claims does not increase younger children's inclination to suggest an efficient test any more than older children.

Deleted: TO TEST OR NOT TO TEST

Deleted: If only Age is a significant predictor, this confirms the age change observed in prior studies, indicating that older children are more likely to test a claim. However, the analysis will not reveal the underlying mechanism of this age change.¶

Deleted: If only Condition is a significant predictor, this indicates that presenting children with different testing strategies has an impact at all ages, and that no age change in children's testing is explained by this form of prompting.¶

Deleted: Is younger children's testing of claims more dependent on the combination of cues to uncertainty and testing strategies than older children?

Deleted: The *Uncertainty-reasoning-hypothesis*

Formatted: Font: Not Italic

Formatted: Font: Times New Roman, 10 pt

Formatted ...

Deleted: testing strategies for surprising claims

Deleted: the Chi square test is significant below the ...

Deleted: the

Deleted: s any more than older children's testing

Formatted: Font: Times New Roman, 10 pt

Deleted: .

Deleted: Uncertainty-reasoning-hypothesis (Hypothesis 3) ...

Deleted: Using the same coding as above, we will run a ...

Deleted: If only Age is a significant predictor, this confirm ...

Formatted: Font: Times New Roman, 10 pt

Formatted ...

Formatted: Font: Times New Roman, 10 pt

Formatted: Font: Times New Roman, 10 pt

Formatted: Font: Times New Roman, 10 pt

Deleted: TO TEST OR NOT TO TEST

And is this effect stronger for younger than older children? And do these effects hold also when controlling for children's ability to select an efficient test?	expected to be stronger among younger children. And these effects are not driven by their ability to select an efficient test.				one's uncertainty may not be the key driver of the propensity to test the veracity of a claim. If the there is only a significant effect of the covariate reflecting children's ability to select the most efficient test, this suggest that scientific reasoning may be a key driver of children's inclination to test a claim.	Scientific reasoning is not key driver of children's inclination to test a claim.

Deleted: TO TEST OR NOT TO TEST

References

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology, 20*, 131–144. doi:10.1016/S0732-118X(02)00005-3

Baer, C., & Kidd, C. (2022). Learning with certainty in childhood. *Trends in Cognitive Sciences, 26*, 887-8896. doi:10.1016/j.tics.2022.07.010

Baer, C., & Odic, D. (2019). Certainty in numerical judgments develops independently of the approximate number system. *Cognitive Development, 52*, e100817. doi:10.1016/j.cogdev.2019.100817

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development, 83*, 1416–1428. doi: 10.1111/j.1467-8624.2012.01775.x

Cottrell, S., Torres, E., Harris, P.L., & Ronfard, S. (2022). Older children know they are exploring surprising claims to verify them. *Child Development, early online publication*. doi: doi.org/10.1111/cdev.13847

Erb, C. D., & Sobel, D. M. (2014). The development of diagnostic reasoning about uncertain events between ages 4–7. *PloS one, 9*, e92285. doi:10.1371/journal.pone.0092285

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. doi:10.3758/BF03193146

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*, 1623-1627. doi:10.1126/science.1223416

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences, 113*, 3492-3496. doi:10.1073/pnas.1515129113

Hermansen, T. K., Ronfard, S., Harris, P. L., Pons, F., & Zambrana, I. M. (2020). Young

Formatted: Norwegian (Bokmål)

Deleted: Chen, X., & French, D. C. (2008). Children's social competence in cultural context. *Annual Review of Psychology, 59*, 591–616. doi:10.1146/annurev.psych.59.103006.093606
Corriveau, K. H., Kim, E., Song, G., & Harris, P. L. (2013). Young children's deference to a majority varies by culture. *Journal of Cognition and Culture, 13*, 367–381. doi:10.1163/15685373-12342099

Deleted: Cottrell, S., Torres, E., Harris, P.L., & Ronfard, S. (*under review*). Older children know they are exploring surprising claims to verify them.
Giron, A. P., Ciranka, S. K., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2022, April 4). Developmental changes in learning resemble stochastic optimization, *PsyArXiv*. doi: 10.31234/osf.io/9f4k3

Deleted:

children update their trust in an informant's claim when experience tells them otherwise. *Journal of Experimental Child Psychology*, *205*, 105063. doi:10.1016/j.jecp.2020.105063

Hermansen, T. K., Ronfard, S., Harris, P. L., & Zambrana, I. M. (2021). Preschool children rarely seek empirical data to resolve conflicts between observation and testimony. *Child Development*, 92, 2546-2562. doi: 10.1016/j.jecp.2020.105063

Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, *86*, 327-336. doi:10.1111/cdev.12298

Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, *48*, 135–145. doi:10.1016/j.cogdev.2018.06.011.

Kuhn, D. (2001). How do people know? *Psychological Science*, *12*, 1-8. doi:10.1111/1467-9280.00302

Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist*, *57*, 73-86. doi:10.1080/00461520.2021.198860

Lapidow, E., Killeen, I., & Walker, C. (2022). Learning to recognize uncertainty vs. recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Developmental Science*, *25*, e13178. doi:10.1111/desc.13178

Lapidow, E., & Walker, C. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315. doi:10.1016/j.cognition.2020.104315

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child development*, *83*, 173-185. doi:10.1111/j.1467-8624.2011.01691.x

Deleted: TO TEST OR NOT TO TEST

Deleted: *, early online publication*

Formatted: Font: Not Italic

Deleted:

Deleted: *early online publication*

Deleted:

Deleted: Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young ¶
children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*, 141-152. doi: 10.1024/1421-0185.64.3.141¶

Deleted:

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction, 29*, 43–55. doi:10.1016/j.learnin-struc.2013.07.005

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science, 2021*, e13095. doi:10.1111/Desc.13095

Mounoud, P., & Bower, T. G. R. (1974). Conservation of weights in infants. *International Journal of Cognitive Psychology, 3*, 29–40.

Pelz, M., & Kidd, C. (2020). The elaboration of exploratory play. *Philosophical Transactions of the Royal Society B, 375*, 20190503. doi:10.1098/rstb.2019.0503

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31*, 153-179. doi:10.1111/j.2044-835X.2012.02082.x

Piekny, J., Grube, D., & Maehler, C. (2014). The development of experimentation and evidence evaluation skills at preschool age. *International Journal of science Education, 36*, 334-354. doi:10.1080/09500693.2013.776192

Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from metaignorance: How children develop an understanding of their own lack of knowledge. *Child development, 83*, 1869-1883. doi:0.1111/j.1467-8624.2012.01830.x

Ronfard, S., Chen, E. E., & Harris, P. L. (2018). The emergence of the empirical stance: Children's testing of counter-intuitive claims. *Developmental Psychology*, 54, 482-493. doi:10.1037/dev0000455

Ronfard, S., Chen, E. E., & Harris, P. L. (2021). Testing what you're told: The role of culture and parents' authoritarian beliefs. *Journal of Cognition and Development*, 22, 426-447. doi:10.1080/15248372.2021.1891902

Deleted: TO TEST OR NOT TO TEST

Deleted:

Deleted:

Deleted:

Deleted:

Deleted: *accepted for publication*. doi: XX

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental psychology, 53*, 1620–1632. doi:10.1037/dev0000340

Ruggeri, A., Swaboda, N., Sim, Z. L., & Gopnik, A. (2019). Shake it baby, but only when needed: Preschoolers adapt their exploratory strategies to the information structure of the task. *Cognition, 193*, 104013. doi:10.1016/j.cognition.2019.104013

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*, 1045-50. doi:10.1037/0012-1649.43.4.1045

Siegler, R. (1996). *Emerging minds: The process of change in children's thinking.* New York: Oxford University Press.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infant's learning and exploration. *Science, 348*, 91-94. doi:10.1126/science.aaa3799

van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science, 43*, 381-400. doi:10.1007/s11251-015-9344-y

Weisberg, D. S., Choi, E., & Sobel, D. M. (2020). Of blickets, butterflies, and baby dinosaurs: Children's diagnostic reasoning across domains. *Frontiers in psychology, 11*, e2210. doi:10.3389/fpsyg.2020.02210

Zimmerman, C., & Klahr, D. (2018). Development of scientific thinking. *Developmental and Social Psychology, 4*, 1-25. doi:10.1002/9781119170174.epcn407

Deleted: TO TEST OR NOT TO TEST

Deleted:

Deleted: Schulz, L.E. (2012). The origins of inquiry: Inductive inference and exploration in early ¶
childhood. *Trends in Cognitive Science, 16*, 382-389. doi: 10.1016/j.tics.2012.06.004¶

Deleted:

Formatted: Norwegian (Bokmål)

Deleted:

Formatted: Indent: First line: 1,25 cm