October 13th, 2022

Dear Prof. Chambers,

We hereby resubmit the revised manuscript now titled "*Knowing why: Children's reflection on their own uncertainty about an adult's surprising claim increases their tendency to test that claim*" (MS #201) for consideration as a Stage 1 Registered Report to *PeerCommunitIn*. We are grateful for the opportunity to submit this revision, and for your suggested improvements. Considering the reviewer's substantial comments, we have thoroughly revised the manuscript to provide a clearer rationale and improved design. We address the points in more detail in the Reviewer Response section below.

In your editorial comments, you asked us to clarify our hypotheses and rationale and to reconsider the suitability of the scaffolding manipulation. In light of this, we have clarified what we mean by metacognitive abilities by distinguishing between the ability to feel uncertain and the ability to reason about the cause of that uncertainty. We elaborate on how we believe this change is implicated in children's increasingly targeted exploration (pp. 2-6). By re-writing the introduction, we have also responded to the Reviewers' request to elaborate on prior work and to provide a better justification for the chosen age range (pp.3-5). The title of the manuscript is modified to better reflect these changes.

In terms of the design, we have made substantial changes to the experimental procedures in line with the revised introduction and the reviewers comments on the previous draft (pp. 9-17). Specifically, responding to Reviewer 1's concerns about possible carry-over effects we have changed from a within- to a between-subject's design. In short, all children will now be assigned to one of two conditions of four trials: a prompted and an unprompted condition. In both conditions, all children are first presented with a surprising claim, before asked the extent to which they believe in the claim, and how certain they are in their belief. Next, children in the prompted condition are encouraged to reason about why they feel the way they do. Finally, children are asked whether they would like to figure out whether the claim they heard is true, and if so, how they would go about it. This new design will allow us to assess more directly the impact of children's ability to report on their uncertainty versus their ability to reason about the causes of that uncertainty on their decisions to test a claim, and their strategies for doing so. Furthermore, we decided to remove the third fully scaffolded condition from the main experiment, and rather introduce a separate task to index children's ability to select the appropriate test among a set of options at the end of the session. Children's scores will be introduced as a covariate in the relevant analyses.

Finally, in light of your comment about providing a more detailed reporting of sample size planning, as well as simulated data and code for the analyses we now include an R-script with a simulated dataset that illustrates what we expect the data might look, together with the specific parameters of the power analysis, as well as the main planned analyses. With the new design and coding plans, we have opted for a simpler analytical approach than previously described and provide details of the analyses associated with each of the main hypotheses in the attached R-script.

In the remainder of the letter, we respond point by point to the issues raised. The original comments are displayed in *italics*, and our responses are interspersed below in **bold**.

We look forward to receiving your feedback.

Kind regards,

Tone Hermansen

Postdoctoral research fellow,
Department of Psychology, University of Oslo
Forskningsveien 3A,
N-0373 Oslo, Norway
Phone: (+47) 99 25 28 22
E-mail: t.k.hermansen@psykologi.uio.no

Reviews

Reviewed by Elizabeth Lapidow, 09 Jul 2022 01:00

Review of To test or not to test: Uncertainty and information seeking following surprising claims.

The proposed study aims to investigate whether improvements in uncertainty awareness or experimentation abilities drive developmental change in children's information-seeking following surprising claims. It is an intriguing idea, and a better understanding of the mechanisms of development in children's experimental/exploratory inquiry would be of great value to the field. Unfortunately, I was disappointed by the content of the current proposal. The descriptions of the competing hypotheses were imprecise and occasionally confusing, and I am not convinced that the experimental manipulation is appropriate for detecting differences in the constructs of interest.

Below, I explain these concerns in more detail. They are roughly in the order they occur in the manuscript, and I have included the Stage 1 Review Criteria they speak to where appropriate.

Thank you for the opportunity to read and provide feedback on this proposal. I hope my comments are helpful in clarifying the design for this genuinely intriguing line of investigation. — Elizabeth Lapidow

Introduction.

Page 2. [Criteria 1A] The introduction is vague about what differentiates older and younger children's behavior. Paragraph 3 mentions "propensity to verify a claim," but the example in paragraph 2 is about using evidence, not generating it. Given that it is the focus of the current investigation, it is necessary to explicitly define what changes and when.

We agree that this could have been presented more clearly. We have now thoroughly revised our manuscript to better highlight what we believe to be the essential contribution of this study—comparing the ability of children to report on their feelings of uncertainty versus their ability to reason about the causes of that uncertainty—providing a more elaborate review of relevant previous literature (pp. 2-6). We have also revised the mentioned section by providing a more careful description of these prior studies now on pages 4-5. Specifically, we elaborate on the findings from Ronfard et al. (2018), focusing on the distinction between children's efficient and inefficient testing of claims. We also include descriptions of Cottrell et al. (2022) replicating the effect in a third-party task where children reason about how another child should act after hearing a surprising claim.

Line 36-47. I suggest the authors expand and elaborate their review of the previous literature in this paragraph. In particular, the statement that "only 5-6-year-olds make use of the information they acquire" during exploration following surprising claims seems very general for only being evidenced by one study? I would also draw the authors' attention to Köksal-Tuncer & Sodian (2018), in which the majority of 3-6-year-olds generated both appropriate empirical evidence and verbal arguments to disconfirm an informant's claims.

We appreciate the suggestion to look closer at the study by Köksal-Tuncer and Sodian (2018). Indeed, this illustrates developments and abilities in aspects of children's scientific reasoning that should be taken into consideration. However, an important difference between our prior work and that of Köksal-Tuncer and Sodian (2018) is that their study explicitly prompted children to "try and find out what makes the box light up". Thus, while Köksal-Tuncer and Sodian's study shows that children are able to execute an empirical test as young as 3 years of age when alerted to the need for information, we do not know from this study whether children would have recognized this need if left to their own devices. Because this finding is relevant for our argument that younger children may benefit from receiving prompts to explore, we have

now included a comment on this contribution in the section on children's increasingly targeted exploratory behaviors (now page 3).

Page 3. [Criteria 1A] I don't find the suggestion that lack of meta-cognitive awareness would result in 'single rather than comparative object exploration' intuitive. Wouldn't taking half the appropriate actions suggest that children are sensitive to their uncertainty (since they choose to act on it) but fail to complete all the necessary actions? This would seem to be the implication of sentence 51-55, but this reverses in 58-61, where the authors say that single object exploration results from a lack of awareness. It is also unclear what it would mean for children to recognize that a claim is surprising while lacking meta-cognitive awareness?

To clarify this point, and as mentioned above, we have now made substantial revisions to our introduction. In line with Kuhns descriptions of developments in epistemological thinking, we now emphasize the difference between feelings of uncertainty and reasoning about the cause of one's uncertainty and how those two abilities may be associated with children's exploratory behaviors (e.g., pp. 2-3 and pp.5-6). We recognize that both processes involve meta-cognitive components, but emphasize how we believe they may differently affect exploration. Moreover, with the new experimental design we now target this distinction more explicitly by assigning children to one of two conditions—an unprompted and a prompted condition—in which the degree to which children's reasoning about the cause of their uncertainty is manipulated.

Page 3. [Criteria 1A, 1B] Despite reading the description of the Scientific-Reasoning Hypothesis several times, I am still unsure what ability the authors intend to capture. The stated definition is "the ability to design/carry out an effective empirical test," separate from whether one is meta-cognitively aware that a test is required. What ability is this? Is it the conceptual understanding of what experimentation is? Is it the executive and motor control needed to follow through a series of planned actions? This issue is further confused by the fact that "scientific reasoning" is widely used in the literature to refer to a large set of interconnected abilities -- some of which certainly require meta-cognitive awareness of uncertainty. Deana Kuhn and others argue that explicit metacognitive understanding of the relationship between belief, uncertainty, and evidence is -required- for fully developed scientific reasoning. This lack of precision in one of the two competing hypotheses critically makes it very difficult to evaluate the empirical design.

Indeed, Kuhn (2001) argues for a differentiation between the competence to apply a specific method and the disposition to do so, while recognizing that both processes involve meta-cognitive components. We have now clarified what we mean to capture by focusing on children's feelings of versus reasoning about uncertainty on pages 2 and 3, and also elaborated on how these may be differently associated with children's exploratory decisions and behaviors on pages 5-6. Specifically, being able to identify the reasons for one's skepticism about a claim is likely to be a strong predictor of the types of action one envisions being informative and effective in assessing the veracity of a claim.

Page 4. The authors say, "Support for the Uncertainty-awareness-hypothesis comes from research showing that children's early exploratory behavior indicates sensitivity to uncertainty." In what way does this support the hypothesis? Wouldn't children showing sensitivity to uncertainty in their exploration at 4-5 years old be inconsistent with a proposal that sensitivity to uncertainty is what develops to support exploration between the ages of 4 and 7?

As we now describe in the revised introduction, there are several age-related changes in children's ability to report on their own ignorance across the ages 3-7 (e.g., Rohwer et al., 2012). While younger children may engage in selective exploration that indicates an implicit sensitivity to the information uncertainty at hand (e.g., Schulz & Bonawitz, 2007), the finding that they rarely engage in spontaneous, comparative exploration (e.g., Ronfard et al., 2018) suggests that their understanding of what constitutes an informative test in a given situation is limited. We suggest that this limitation can in part be explained by a limited understanding of the cause of

their uncertainty and draw parallels to the literature on diagnostic reasoning—the ability to reason about the cause on a phenomenon (Fermbach et al., 2012). That is, early selective exploration may be guided by the understanding that uncertainty requires further information, while later comparative exploration is guided by the understanding that uncertainty is not reduced unless alternative hypotheses are ruled out. We attempt to clarify this point across pages 5-6.

Page 4. It is misleading to say that Lapidow et al. (2022) "found no association between reports of uncertainty and exploration decisions" since no such test of association was conducted in that study.

Our understanding of this work was that while children's explicit reports of uncertainty—as measured through their confidence judgements—was not associated with the relative uncertainty of the presented stimuli in Experiment 1. However, as an indication of children's implicit uncertainty, Experiment 2 revealed a preference for the more ambiguous stimuli. When viewed in combination, these experiments appear to suggest that while children's exploration may be guided by an implicit sense of uncertainty, 4-5-year-olds do not readily report on this uncertainty. We have clarified our descriptions of these findings, now page 2-3 of the revised manuscript to better reflect the original paper.

Page 5. [Criteria 1C] I have reservations about the proposal to use scaffolding manipulation as an indication of the development of children's abilities. This presumes improvement in performance created by scaffolding is necessarily due to an underdeveloped awareness/understanding of what is scaffolded. This presumption is particularly concerning for the relationship between Hypothesis 2 and the Strategy Scaffolding condition. If children are successful at generating responses without options offered, this would certainly be evidence that they understand how to test uncertain claims. However, generating responses (of any kind) involves demands that selecting responses does not, including verbal ability, working memory, and attention. None of these are specific to experimentation ability and could readily lead to children performing better in the Scaffolded condition irrespective of their grasp of correct experimentation. Indeed, work by Azzurra Ruggeri and colleagues shows that 4- to 5-year-olds' evaluation- of information seeking via question-asking is sensitive to the same considerations as that of older children and adults, even though they are not yet able to -generate- questions of this complexity.
As a possible suggestion for revision -- a more promising approach to this investigation might be to employ separate assessments of the two constructs of interest and see if variation in performance on these measures can account for the difference in performance between older and younger children.

We agree that there may be different cognitive demands associated with the selection versus generation of a response. Indeed, prior work by for example Mayer and colleagues (2014) as well as Koerber and Osterhaus (2019) indicates that there is a relationship between scientific reasoning and, for example, intelligence, language, and reading. Importantly though, scientific reasoning remained a separate construct from these more basic information processing capacities.

With this comment in mind, we have now thoroughly revised our design as described on pages 6-8 and in further detail across pages 9-14. Specifically, we have now changed the design from a within- to a between-subjects and now assign children to one of two conditions: an unprompted and a prompted condition. The unprompted condition allows us to assess the previously observed age change in children's propensity to suggest an efficient test of a claim when left to their own devices, while the prompted condition allows us to assess the impact of uncertainty reasoning on children's inclination to construct an efficient test. To control for individual differences in children's scientific reasoning, and to emphasize Kuhn's distinction between the competency versus propensity to reason scientifically, we decided to leave what was previously the fully scaffolded condition as a separate task administered to all children and introduced as a covariate in the models (with detail on pages 8 and 13).

Methods.

Page 6. [Criteria 1C, 1E] I may be misunderstanding, but it seems like the three conditions were always presented to children in the same order — starting with no scaffolding and ending with scaffolding on both dimensions. If this is the case, it opens up the potential for confounding order effects. In particular, I worry that repeated exposure to/opportunities to think about the task question (as the trials are only very superficially different from each other) might improve children's performance even without the added scaffolding.

In light of this comment, and as mentioned above, we have now changed the design from a within- to a between-subjects design, assigning children to only one condition each and assessing scientific reasoning in a separate task. In addition, to assess possible implications of repeated trial exposure within a condition, we will include a set of control analysis to explore both trial and block effects. In the event that there are effects of trial, or block, order we will control for this in the main analyses.

Page 7-8. The authors provide a compelling reason for changing the task from first to third-person by having participants think about what a character in a story believes/should do, rather than themselves. However, it strikes me that there may be significant differences between awareness/sensitivity to uncertainty in oneself as opposed to others. The research cited in the introduction focuses on internal monitoring and meta-cognition. Similarly, "what do you think the protagonist should do?" is very different from past work examining children's spontaneous actions. Can the authors expand their discussion of this change to include considerations of what it would mean if they found differences between this and past work?

This is an important point, and indeed, studies on children's moral reasoning have found a somewhat protracted development in reasoning about first versus third persons intentions (e.g., Margoni & Nava, preprint retrieved from osf.io/vkf5q). However, looking at work with similar focus and design as the current study, there is preliminary support for similar developmental trajectories of first and third-person reasoning (e.g., Cottrell et al., 2022). That said, considering this comment and the fact that that the new design now more specifically targets the child's own uncertainty, we have re-considered the use of a third-person narrative and now focus on the child's own uncertainty and exploratory decisions.

Page 7. [Criteria 1E] Do the authors have a plan to rule out the possibility that the Uncertainty Scaffolding prompt simply clarifies the task goal rather than scaffolding participants' recognition/understanding of uncertainty in the story?

Upon reflection we realize that, for some children, the uncertainty prompt may clarify the task. Thus, with the revised experimental design we now explicitly ask all children about their belief in the adults claim, and their relative certainty in that belief before asking them whether they would want to find out the truth of the claim, and if so, how (see pp. 6-7, pp. 12-13, and Appendix 1). We made these changes primarily because it helps us clarify the distinction between feelings of uncertainty (as indexed by their responses to the Uncertainty question) and their reasoning about uncertainty (as indexed by children in the prompted condition's responses to the Reasoning questions). It should also reduce the chance that children perceive the task or situation as unclear.

Page 8. [Criteria 1D] What does "encouraging children's explicit representation of epistemic uncertainty in the trials where no scaffolding is provided" mean? Isn't the scaffolding meant to encourage/support the representation of epistemic uncertainty?

With the revised experimental design, this should no longer be an issue.

Minor Comments.

26 - The opening sentence of the introduction seems incomplete. Presumably it is meant to say control increases with age?

With the revised introduction, this has now been corrected on page 2.

44, 48, 92 - The general language used to refer to age ranges in these sentences should be revised. Given the highly developmental nature of the claims proposed, the authors need to be very specific about what ages are referred to.

The has now been revised and specified in the introduction (pp.2-5).

159 - "and to level children's prior knowledge…" this sentence is unclear, maybe a word is missing?

With the revised methods this should no longer be an issue.

Reviewed by Amy Masnick, 13 Jul 2022 21:09

This manuscript proposes a study examining young children's exploration and attempts at verification information they have been given. The research builds on work suggesting a developmental shift in skills, and wants to test if this development is due to children's increased skill at effective comparisons (designing comparison tests) or at their increased skepticism about information they're told. In particular, the task proposed is one in which participant children learn about another child who is exposed to a surprising piece of information, and then are asked the recommended next step.

I think the topic is interesting, but I have a number of issues I would like to see addressed to clarify the design and proposed analyses.

The researchers cite evidence that 5-6-year-olds who do seek new evidence after hearing a surprising claim can typically use the information. This is the study cited as evidence for a likely transition around ages 5-6, though the design is to compare 4-5-year-olds and 6-7-year-olds. I think that age choice could be more clearly explained.

The two potential explanations offered are that children get better at exploration because their metacognition improves, or because their scientific reasoning (operationalized as their ability to design an effective test) improves. In some ways, these are presented as mutually exclusive, as if both skills could not improve around the same ages. I understand the scientific reasoning hypothesis, in which good scientific reasoning ideally would lead to a clear, controlled test, and is clearly more informative than testing a single object. I am less clear about the argument behind the skepticism hypothesis. The argument as I understand it is that children who recommend a single object test might be doing so not because of a lack of understanding of scientific reasoning, but rather because they are unaware they are skeptical or should be skeptical because information provided violates their expectations/prior knowledge. I read these sections a few times, and still find the argument a little hard to follow. This argument needs to be explained and justified more clearly.

We realize that our description of the core argument may not have been as clear as we intended. Our point is that we believe the large range of prior work on children emerging reasoning capacities suggest that children can reason scientifically when provided with strong prompts to do so, but that the work on children's limited testing of surprising claims suggest that they do not always think of doing so when left to their own devices. Our proposal is that this limited propensity to test a claim is thus not due to a limited understanding of scientific reasoning but due to their limited insights as to why they are uncertain. We have tried to clarify this point already in the first paragraph of the revised manuscript, as well as across pages 5-6. That said, we agree that both metacognition and scientific reasoning skills may improve across the same period, and we therefore now acknowledge how these processes develop in parallel by elaborating on this across pages 3-4.

I am unclear about why the three conditions are no scaffolding, uncertainty scaffolding, and uncertainty and strategy scaffolding. I assumed when I started reading that there would be a scientific reasoning conditioning and an uncertainty condition, but that's not the design. Is there no strategy scaffolding as its own condition because the strategy is irrelevant if children don't recognize the uncertainty? Hypothesis 3 predicts a progressive increase from no scaffolding to uncertainty scaffolding to uncertainty and strategy scaffolding in younger children, presumably because they need help with each of the components, which suggests a more progressive development, rather than two equivalent competing hypotheses. That point should be spelled out and clarified.

We recognize these concerns as valid points, and as mentioned above, we have now thoroughly revised our experimental design in light of this and other related comments. Particularly relevant to this comment we have revised the design so that the fully scaffolded condition is now a separate Selection task administered to all. For details, see pages 6-7 and 12-13.

In the proposed analyses, it seems a little unusual to have three conditions, and then have one test that conditions 1 & 2 show a different pattern by age, and another that conditions 2 & 3 show a different pattern by age. I understand why the hypotheses are proposed separately but is there not one test that can assess these all in one analysis, with likely post hoc follow-ups?

Coding of the intent to test – in two of the conditions, children can give an open-ended response, which will then be classified as either an intent to test, an intent to explore but not test, or something else (or a non-response). There is no mention of any assessment of inter-rater reliability of the coding, which should be included.

This should have been included. We have now inserted a brief description of the responsibilities of the two research assistants on page 14 in the revised manuscript. We also describe how we plan to handle discrepancies between coders, and what we will do in the event of low inter-rater reliability.

Also, why are you coding for multiple possibilities, but only exploring analyses that collapse the outcome responses into testing or no testing? If you plan to code it, why wouldn't you look at that information to see what patterns may be shown by the additional types of responses? I don't think you would necessarily need to have a clear hypothesis of what the data might look like, but ignoring part of the data also seems unnecessarily limiting.

Indeed, to answer the main hypotheses of the paper, we use only parts of the coded data on children's suggestions for exploration. We do this to emphasize the key behavior of interest, namely children's ability to suggest an efficient test. However, we realize that including descriptive information about children's exploratory tendencies in more general could of course be relevant to include. We also recognize that further exploratory analyses may be of interest, but as we did not want to clutter the main argument of the paper, we only include a brief mention of this on page 17. Any later exploratory analyses will be appropriately marked as such in the final manuscript.

Sample size – I am not an expert in multilevel models, but from my basic understanding, it is important to consider sample size at each of the levels, not just the overall participant sample size. At the very least, the model and its justification should be better explained and supported.

We realize that this should have been better specified. However, in light of this comment, and upon revising our experimental design we now plan for a simpler analytical strategy. To clarify this point, we now provide a more a more detailed description our sample size planning in the attached R-script, together with a synthetic dataset intended to illustrate what we expect the data to look like as well as more detailed descriptions of the main statistical analyses.

In looking at the stimuli in the appendix, I wondered about olfaction as compared to the other categories of stimuli. I realize the study participants wouldn't have to actually manipulate objects, but asking someone to smell something unpleasant feels like something people might avoid for purely sensory reasons, independent of experimental design, in a way that would not be an issue with density and weight examples.

We acknowledge that this could be a concern, however this does not seem to be the case in the recent publication by Cottrell and colleagues (2022) using very similar stimuli tips—presenting children with trials referring to weigh, density, and olfaction, without registering any notable difference in children's responses. However, in light of this comment, we have decided to use fewer trials referring to olfaction and to reframe the olfaction claim to refer to good smells rather than bad (e.g., this banana smells much better than all the other fruit).

In sum, I think there are a lot of interesting ideas, but I am not fully convinced by the proposal as is, and would like to see better justifications and explanations for many parts of it, and perhaps some additional tweaks to the design.

We appreciate the reviewer's helpful comments, and we hope the revised manuscript reflects a better justification and explanation for our main hypotheses. As described above, we have also made adjustments to the experimental design to better align with the core of our argument and to clarify the task for the child participants.