Received: 22 January 2021 | Revised: 12 April 2021 | Accepted: 12 April 2021

DOI: 10.1111/sode.12533

ORIGINAL ARTICLE

WILEY

Young children evaluate and follow others' arguments when forming and revising beliefs

Hannes Rakoczy | **Nadja Miosga** | **Thomas Schultze**

Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

Correspondence
Dr. Hannes Rakoczy, Department of Developmental Psychology, University of Göttingen, Waldweg 26, D-37073 Göttingen, Germany.
Email: hannes.rakoczy@psych.uni-goettingen.de

Funding information
German Science Foundation

Abstract

What do young children understand about arguments? Do they evaluate arguments critically when deciding whom to learn from? To address this question, we investigated children at age 4–5, when robust selective social learning is in place. In Studies 1a/b, children made an initial perceptual judgment regarding the location of an object under varying perceptual circumstances; then received advice by another informant who had better/worse perceptual access than them; and then made their final judgment. The advice by the other informant was sometimes accompanied by utterances of the form "I am certain … because I've seen it". These utterances thus constituted good arguments in some conditions (informant could see clearly), but not in others (informant had poor perceptual access). Results showed that children evaluated argument quality in context-sensitive ways and used them differentially for belief-revision. They engaged in more belief-revision when the informant gave this argument only when her perceptual condition, and thus her argument, was good. In Study 2, children were asked to find out about different properties (color/texture) of an object, and received conflicting testimony from two informants who supported their claims by utterances of the form "because I've seen it" (good argument regarding color/poor regarding texture) or "because I've felt it" (vice versa). Again, children

engaged in context-relative evaluation of argument quality, selectively learning from the agent with the appropriate argument. Taken together, these finding reveal that children from age 4 understand argument quality in sophisticated, context-relative ways, and use this understanding for selective learning and belief-revision.

KEYWORDS
argumentation, belief-revision, selective trust, social cognition

1 | INTRODUCTION

Most of what we know about the world, we learn from what others tell us. This learning by testimony raises the challenge to decide whom to learn from. One important source for such decisions is to keep track of not just what informants say, but how they justify what they say with arguments. Imagine you want to find out about the surface of the moon and find yourself faced with two sources of conflicting testimony. Neil Armstrong claims the surface of the moon is made of stuff A, whereas John Doe claims it is made of stuff B. Armstrong justifies his claim by the argument "…been there, seen it" whereas John Doe's argument is "…I have seen the Wallace & Gromit episode in which they fly to the moon". You yourself have no capacity to distinguish between the prior veracity or plausibility of the A versus B options as such. But you do recognize that A is more plausible in light of the argument adduced for it than B is.

From the point of view of cognitive development, the central question is when and how children develop such capacities. When do they acquire the (meta-)representational ability to critically evaluate the quality of others' arguments in order to form new or revise existing beliefs in response to testimony?

Much research has documented that preschool-aged children selectively learn from others when forming new beliefs in multifarious ways (for reviews, see Harris, 2012; Hermes et al., 2018; Sobel & Kushnir, 2013). They take into account perceptual access (e.g., Robinson et al., 2011, 1999), epistemic reasons (e.g., Koenig, 2012), previous epistemic history (e.g., Koenig et al., 2004), linguistic expression of evidentiality (Fitneva, 2009), confidence (e.g., Sabbagh & Baldwin, 2001) and other relevant information when deciding who to learn new facts from. Recent research has also shown that children from around age 4 engage in selective social belief revision, updating and modifying their judgments more in response to an agent's advice when this agent is better informed or more competent (Miosga et al., 2020; Rakoczy et al., 2015; Robinson & Whitcombe, 2003).

Yet, relatively little is known about how children take into account other agents' arguments in their selective learning and belief-revision. In general, understanding, evaluating and using arguments is a complex form of higher cognition with protracted (potentially lifelong) development (Kuhn, 2000). For example, scientific argumentation skills keep on developing until adolescence and beyond (Kuhn & Udell, 2003)[1]. The most basic forms of using, understanding and evaluating arguments, however, develop much earlier in ontogeny, during the preschool years and potentially even earlier (Bernard et al., 2012; Castelain et al., 2016, 2018; Mercier et al., 2014, 2018). Children as young as 4–5 years of age (sometimes even as young as 2 years) produce reasonable arguments themselves (Dunn & Munn, 1987; Köymen & Tomasello, 2018; Köymen et al., 2014, 2016, 2020); and they differentiate between better and worse arguments given by others: they prefer strong (e.g. "The dog went in this direction…because I have seen him go there") over empty (circular) arguments (e.g., "The dog went in this direction…because he went in this direction") (Castelain et al., 2018; Mercier et al., 2014). Relatedly, children this age are sensitive to the appropriate use of relevant connectives such as "because" in good arguments (Bernard et al., 2012). Recent research also suggests that children this age put this sensitivity towards argument quality to use in their selective belief formation (Corriveau & Kurkul, 2014) and belief

revision (Castelain et al., 2018): they acquire new information and modify existing judgments more readily in response to an informant with good rather than bad arguments.

One limitation of existing studies, though, is that argument quality was always confounded with certain formal properties of the corresponding utterances including syntax, semantics, length etc.: the utterances either did or did not include relevant connectives such as "because"; or better arguments had more informational content that poorer ones (with the limiting case of contrasting informative ("p because q") with circular arguments that lack substantial content ("p because p"). Formally speaking, however, what determines the quality of an utterance as an argument is not its informational content per se, but the relation between its content and the conclusion it is supposed to license: how strongly does the premise support and warrant the conclusion (e.g., Toulmin, 1958)? The quality of an utterance as an argument is thus a context-relative matter. One and the same utterance with the same informational content can be a good argument in one context and a poor one in another context, depending on its justificatory relation to the to-be-supported conclusion. To illustrate this point, consider the following dialogue (from Monty Python's famous Dead Parrot Sketch[2]):

Customer: "Hello, Miss!"
(Male) Shopkeeper: "What do you mean, 'Miss'?"
Customer: "Oh, I am sorry, I have a cold."

Now, the utterance "I am sorry, I have a cold" which here is a complete non-sequitur (and thus funny in the first place) may well be a good explanatory argument in another context. Think of the following dialogue:

Customer: "Hello, Miss!"
(Female) Shopkeeper: "Pardon? I can't hear you. You have to speak up"
Customer: "Oh, I am sorry, I have a cold."

To illustrate the role such context-sensitive argument evaluation can play for selective learning, consider the following less humoristic example: Imagine you want to know about the chemical composition of some substance and you ask informants 1 and 2.

Informant 1 (after having looked at the substance through sunglasses): "It's made of C; I know this because I have looked at it".
Informant 2 (after having looked at the substance through a microscope): "It's made of D; I know this because I have looked at it".

Both informants make the same utterance ("I know this because I have looked at it"). This utterance amounts to a good argument in the case of informant 2, but fails to do so in the case of informant 1. Understanding this context-sensitivity licenses a number of inferences: First, it licenses judgments about the matter at hand: it is (all else equal) more likely that the substance is made of D than that it is made of C. Second, it may license judgments about the informants more generally: informant 2′s epistemic status as reliable and reflective (knowing what count as good evidence and argument) seems more secure than that of informant 1 (who appears to be deeply confused about the evidence and argument structure at hand; or at least about chemistry and/or optics). In fact, informant 1′s giving this particular ill-guided reason actually constitutes reason for not believing her in this particular case; and perhaps for skepticism about her epistemic reliability more generally. Children's sensitivity to the appropriateness of citing evidence and expressing confidence has recently been studies under the rubric of "speaker calibration". This research has found that children from around age 4 begin to prefer speakers who are well calibrated such that they pledge ignorance when appropriate and nuance their expression of confidence as a function of the quality of their evidence (Birch et al., 2020; Tenney et al., 2011).

The rationale of the present study was to investigate the emergence of such forms of evaluation in the context of argumentation: when do young children understand, evaluate and use argument quality in context-sensitive ways in their selective social learning and belief revision. Studies 1a and 1b tested whether children engage in selective belief revision when confronted with an informant whose perceptual access varied across conditions (looking through a clear vs. through a blurred window; analogous to the sunglasses/microscope example). In each case, the agent made an utterance ("…because I have seen it") the quality of which as an argument varied, in turn, as a function of the quality of perceptual access (as in the sunglasses/microscope example). The central question was whether children would selectively revise their beliefs as a function of the quality of the informant's perceptual access and of the quality of her argument.

Study 2 investigated children's selective social learning (belief formation) as a function of argument quality. Children's task was to find out about the color or texture properties of an object hidden in a box ((building on a task by O'Neill et al., 1992). Two informants approached the box which had one side from which one could see (but not feel), and another side from which one could feel (but not see) objects in the box. On a given trial, one informant went to side 1 to look and the other went to side 2 to feel. Then both made conflicting claims about the color/texture of the object and backed them up by uttering "…because I have seen (side 1)/felt (side 2) it" – which constituted good or poor arguments in context-relative ways. The central questions were whether children selectively endorsed claims from the informant at side 1 or side 2 when color or texture were in question, respectively; and whether children would, over trials, keep track of who gave repeatedly sound ("red… because I have seen it") versus disqualifying would-be-arguments ("red…because I have felt it…") and make corresponding inferences about the agents' more general epistemic reliability.

We studied children from age 4 to 5 for various reasons: First, robust nascent skills for selective belief formation and revision have been documented from age 4 on (Harris, 2012; Harris et al., 2018; Rakoczy et al., 2015). Second, basic skills for understanding and evaluating arguments have been uncovered at this age (Corriveau & Kurkul, 2014; Mercier et al., 2014). Third, Study 2 builds on research on children's understanding of modality-relative epistemic access (visual vs. tactile regarding color vs. texture) that has found that children acquire such understanding from around age 4 (O'Neill et al., 1992).

2 | STUDY 1A

Do young children revise their beliefs selectively in response to advice depending on the advisor's perceptual situation and depending on whether or not the adviser supports the advice by arguments? To address this question, we made use of a method that was recently adapted from "judge-advisor" studies in adult social psychology (Sniezek & Buckley, 1995) for the use with children (Miosga et al., 2020). Participants were faced with a perceptual judgment task: to locate visual marks on objects by looking through one of two windows, one clear, the other one blurred. Children acted as judges who made an initial estimation, then received advice by the advisor, and then were allowed to make their final judgment. One factor that varied between conditions was the quality of perceptual access of participant versus adviser (who looked through the clear/blurred window). Previous research has found that children from age 4 revise their initial beliefs more strongly when the advisor was in a better perceptual situation than they were (Miosga et al., 2020). The new variation between conditions in the present study concerned whether the advisor did or did not use an argument ("I know this because I have seen it") to back up her advice.

The questions of interest were, first, whether children would revise their judgments more when the advisor had better perceptual access than they themselves (replicating previous results); and, second, whether children would be sensitive to argument use and quality: If children's understanding of and sensitivity to arguments is superficial, they should revise beliefs more strongly whenever an argument was used (statistically speaking, indicating a main effect of argument use). In contrast, if they understand argument quality in context-sensitive ways, they should revise their beliefs more strongly in response to good arguments (that is, in response to "..because I have seen it" uttered at the

clear window in contrast to the same utterance made at the blurred window; statistically speaking, indicating an interaction effect of perceptual access and argument use).

3 | METHOD

3.1 | Participants

Thirty-nine 4- to 5-year-old children (age range: 48 - 71 months, $M = 59$ months, 22 girls) were included in the final sample. Sample sizes in all studies were determined conventionally on the basis of previous similar studies (e.g. Miosga et al., 2020). Subjects in all studies were native German speakers with mixed socioeconomic backgrounds. Children were recruited from a database of families who had previously given consent to their participation. Participants were tested individually in the lab.

3.2 | Material

We presented participants with a perceptual judgment task in which they estimated the location of a mark on a stick seen through one of two windows in a triangular box. The box had a clear window on one side, a blurred window at another (window measurements: 27.5 cm × 23.5 cm), and an opaque backside.

At the clear window, the black mark was easily visible while at the blurred window it was hardly visible. Wooden sticks of 20 cm length were placed in the box. Participants made their judgment by indicating on an analogue scale (a picture of the stick at a 1:1 scale; identical length) where the mark was. The advisor then pretended to give advice based on her perceptual access (where advice in fact consisted in randomized judgments, so that the de facto accuracy of advices did not co-vary with condition, ranging from one to 200 mm, produced by a random generator and secretly marked on the advisor's scale prior to the test sessions; see Appendix A).

3.3 | Design

Participants were paired with a confederate who animated a puppet in the role of the advisor. What varied between conditions was perceptual access (who sat where) and argument use (whether the advisor used an argument). The design was thus a 2 (perceptual access) × 2 (argumentation) design with perceptual access (*child at clear vs. child at blurred window*) and argumentation (*with* or *without*) as within-subjects factors. All subjects received two consecutive trials of each condition (8 trials in total). Children always received the 4 trials without arguments first, followed by the 4 trials with arguments[3]. Within both blocks, order of perceptual access trials was counterbalanced.

3.4 | Procedure

Introduction phase. In the introduction phase, participants were familiarized with the set-up material to make sure that they understood the design of the three-sided box and the concept of transferring a black mark on a colored stick onto a picture of the stick. Data of the introduction phase were used as manipulation check to ensure that perceptual access was successfully manipulated by having children watch either through the clear or the blurred window.

After entering the test room participants were invited to explore the three-sided box (see Supplement, Figure S1). The two windows were shown and their perceptual quality assessed. To familiarize participants with the stimuli, they looked at one stick placed on a table and were invited to transfer the black mark as precisely as possible onto a picture of the stick. Participants' understanding of the scale as an identical image of the stimuli was ensured. Subsequently,

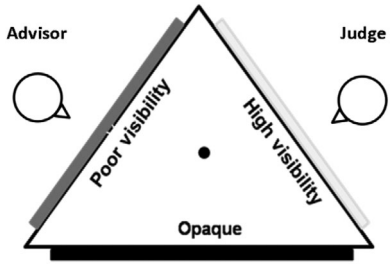

FIGURE 1 Schematic drawing of the experimental setup in a "child at clear window" condition. The child (judge) sits behind the clear window and receives advice from the advisor (puppet) who sits behind the blurred window.

participants judged one stimulus first behind the blurred and second behind the clear window, while it was emphasized that the two judgments differed in quality.

Test phase. Judge and advisor were each placed at a different side of the box (see Figure 1). Then, one stick was placed in the box (in such a way that subjects could not observe the placement), and both judge and advisor were asked to make their initial judgments. Then judges were seated at the opaque side of the box, received advice (accompanied by the scale of the advisor), and were asked to mark their final judgment on a third scale. Advice was presented with or without the argument, "I think it is here (advisor points to mark on her analogue scale) because I looked through here (points to window)" (German: "Ich denke, es ist hier, weil ich hier durchgeschaut habe"). Every two trials judge and advisor switched their starting positions (initial judgment behind the clear or blurred window). All test sessions were video recorded and lasted approximately 25 min.

The dependent measures were as follows:

- *Initial accuracy in the introduction phase*: Initial accuracy was measured as a kind of manipulation check, operationalized as the deviation from the true value, i.e. the distance of the mark drawn by the judge from the true location of the mark on the stick (measured in cm).
- *Advice-taking (AT) in the test phase*: The advice-taking measure is the weight the judge gives the advice, defined by Harvey and Fischer (1997) as

$$AT = \frac{final\ estimate - initial\ estimate}{advice - initial\ estimate} \times 100\,\%$$

If the judge fully endorses the advice, AT would be 100 %, if the judge equally weighs her judgment and the advice, AT would be 50 %, and if she does not consider the advice at all, AT would be 0 %. Other AT scores between 0 % - 100 % represent partial shifts from the initial judgment towards the advice. In order to control for extreme cases and outliers, AT scores are usually truncated. Accordingly, we limited them to 100 % and −100 % (see Schultze et al., 2017 for a similar approach)[4].

4 | RESULTS

4.1 | Preliminary analysis: Initial accuracy

Participants' visibility was successfully manipulated by presenting them with the clear versus blurred window. Children's deviation from the true value when sitting at the table, behind the clear, and behind the blurred window are depicted in Figure 2. An ANOVA for repeated measures with form of visual access (without box, clear or blurred

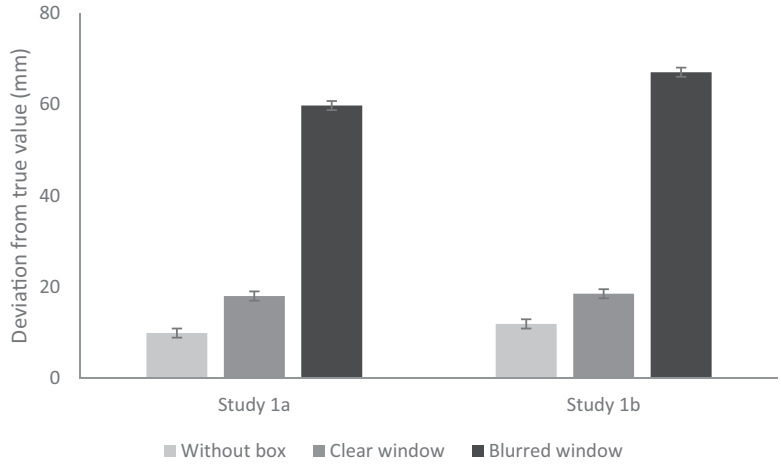

FIGURE 2 Initial accuracy (deviation from the true value) as a function of form of visual access (without box, clear window and blurred window) in Studies 1a/1b. Error bars indicate standard error of the means

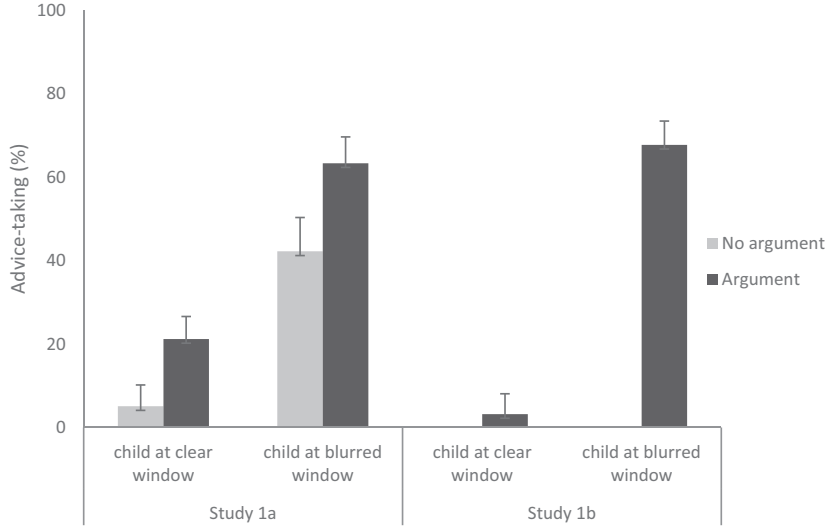

FIGURE 3 Advice-taking as a function of condition and argumentation in Studies 1a and 1b. Error bars indicate standard error of the means

window) revealed a significant main effect of variation of visual access ($F(2, 76) = 81.12$, $p < .001$, $\eta_p^2 = 0.68$). Post hoc paired sample *t*-tests revealed significant differences between all three variations of visual access (without box vs. clear window: $t(38) = -3.46$, $p = .001$, $d_z = -0.56$; without box vs. blurred window: $t(38) = -11.16$, $p < .001$, $d_z = -1.80$; clear vs. blurred window: $t(38) = -7.96$, $p < .001$, $d_z = 1.29$).

4.2 | Main analysis: Advice-taking

Performance in advice-taking as a function of condition and argumentation is depicted in Figure 3. A 2 × 2 ANOVA for repeated measures with perceptual access (*child at clear vs. at blurred window*) and argumentation (*with* or *without*

argument) showed significant main effects of perceptual access ($F(1, 37) = 38.35$, $p < .001$, $\eta_p^2 = 0.51$) and argumentation ($F(1, 37) = 14.67$, $p < .001$, $\eta_p^2 = 0.28$) but no interaction of both factors ($F(1, 37) = 0.17$, $p = .681$, $\eta_p^2 = 0.01$).

5 | DISCUSSION

The present study had three main findings: First, children revised their beliefs more strongly when the advisor was perceptually better situated than they themselves, replicating previous findings (Miosga et al., 2020). Second, children were sensitive to argumentation, revising their belief more strongly when the advisor gave an argument for her judgment. Third, however, there was no evidence for context-sensitive argument evaluation, i.e. no interaction effect of perceptual access and argumentation. Children responded to the utterance "…because I have seen it" in similar ways with increased belief revision in both perceptual access conditions – although arguably this utterance constitutes a good argument only in the condition when the advisor can see well. According to the original rationale of our study, the "…because I have seen it" utterance made at the blurred window should not constitute a convincing argument at all, perhaps even discredit the advisor.

However, maybe children did not really understand the utterance along the lines we had envisaged. Perhaps they understood "…I have seen it" roughly as "I am not just (completely) guessing, but at least trying to see. I'm doing my best (given the circumstances)". Given the present manipulation and results, we cannot tell whether children (mis-)understood the argument in such ways or whether they fail to engage in context-relative argument-evaluation in principle. In addition, two other factors complicate the interpretation of the "argument" conditions of Study 1a: Since the same puppets was used in both conditions (blurred/clear window) there may have been carry-over effects (such that children saw a puppet as competent when she first gave a sound argument at the clear window, and stuck to this impression even in light of her more dubious argument at the blurred window). No such concerns apply to the "no argument" conditions since here the puppet does not make arguments that could discredit her (the quality of the advice is merely a function of the perceptual circumstances and not under the puppet's control). Second, the comparison of "no argument" and "argument" trials is complicated by the fixed order in which they were administered. To overcome these limitations, we implemented a new version of the "argument" conditions in Study 1b.

6 | STUDY 1B

In Study 1b, we tested another sample of children with a new implementation of the two conditions (child at clear/blurred window) with argument. The advisor now not just mentioned having seen, but explicitly justified his judgment and certainty with recourse to perception ("I am absolutely sure because I looked through here"). This utterance should now, again, constitute an unambiguously good argument when made at the clear window, but an unambiguously poor argument when uttered at the blurred window. Children thus only received 4 trials (order of child at clear vs. blurred window counterbalanced) with arguments. In this way, we could compare across the no argument conditions of Study 1a and the new argument conditions in Study 1b in clearer and cleaner ways in a between-subjects design without a confound of this factor with order.

7 | METHOD

7.1 | Participants

Thirty-seven 4- to 5-year-old children (age range: 48 - 71 months, $M = 59$ months, 20 girls) were included in the final sample.

7.2 | Material, design, and procedure

The same stimuli as in Study 1a were used with the exception that every participant judged only four wooden sticks. Children only received 4 trials with arguments, two per perceptual access condition (child at clear/blurred window). The data from Study 1b, taken together with those from Study 1a of the first 4 trials, thus constituted a 2 (perceptual access) x 2 (argument) mixed-factors design, with perceptual access as within-subjects and argument as between-subjects factor.

The procedure was identical to Study 1a with the following exceptions during the test phase: First, the argument was phrased in much stronger and less ambiguous ways: "I really want us to take this one (points to mark on scale). I am absolutely sure because I looked through here (points to window)" (German: "Ich will, dass wir das hier nehmen. Ich bin mir ganz sicher, weil ich hier durchgeschaut habe"). Second, the advice accompanied by the argument was presented by different puppets in the two perceptual access conditions (child at clear vs. blurred window).

8 | RESULTS

8.1 | Preliminary analysis: Initial accuracy

Children's initial accuracy, operationalized as deviation from the true value sitting at the table, behind the clear and behind the blurred window is depicted in Figure 2. Again, visual access of the participants was successfully manipulated as intended. An ANOVA showed a significant main effect of form of visual access ($F(2, 72) = 150.06$, $p < .001$, $\eta_p^2 = .81$). Post hoc paired sample t-tests revealed significant differences between all three forms of visual access (without box vs. clear window: $t(36) = -2.67$, $p = .011$, $d_z = -0.44$; without box vs. blurred window: $t(36) = -17.94$, $p < .001$, $d_z = -2.99$; clear vs. blurred window: $t(36) = -10.67$, $p < .001$, $d_z = 1.78$).

8.2 | Main analysis: Advice-taking

Performance in advice-taking as a function of perceptual access and argumentation is depicted in Figure 3. A 2 × 2 mixed-factors ANOVA with perceptual access as within-subjects and argumentation as between-subjects variables showed a significant main effect of perceptual access ($F(1, 74) = 87.97$, $p < .001$, $\eta_p^2 = 0.54$) and an interaction of perceptual access and argumentation ($F(1, 74) = 4.48$, $p = .038$, $\eta_p^2 = 0.06$). Post hoc tests revealed a significant increase of advice-taking when advice was accompanied by the argument in the *child at blurred window* condition only (*child at clear window* condition: $t(74) = -0.10$, $p = .922$, $d = -0.23$; *child at blurred window* condition: $t(74) = -2.46$, $p = .017$, $d = -0.60$).

9 | DISCUSSION

Study 1b tested for children's context-sensitive argument evaluation in their selective belief-revision with a stronger manipulation involving a less ambiguous utterance. The main finding was that under such circumstances children indeed engaged in context-sensitive argument evaluation, increasing their belief revision in response to advice accompanied by the utterance when it constituted a good argument (at the clear window) only. One potential limitation of the comparison between the argument trials in Study 1b and the no argument trials in Study 1a may be that the conditions were each implemented in appropriate, yet not strictly identical ways (in the argument trials different puppets were used for the different conditions for reasons mentioned above; where this was not necessary for the no argument trials). Ideally, future studies will replicate these findings with conditions even more closely matched.

10 | STUDY 2

Studies 1a and 1b taken together thus suggest that children from age 4 engage in practical forms of context-sensitive argument evaluation: they respond differently and appropriately in their belief-revision as a function of context-relative argument quality. Study 2 addressed the questions whether children this age can also explicitly judge and tell the qualities of arguments (and not just respond differentially to them); and whether children draw wider, dispositional conclusions about advisors from the ways they use arguments. Do they come to see an agent who repeatedly mentions "I am sorry, I have a cold" in non-sequitur ways, or who keeps on claiming to be sure after having had lousy perceptual access as generally epistemically dubious and discredited?

To address these questions, Study 2 tests for children's selective belief formation in a design involving two informants (following the standard paradigm of developmental selective trust research; see Harris, 2012). Building on a task by O'Neill and colleagues (O'Neill et al., 1992), children had to find out about the properties of an object hidden in a box accessible to different sense modalities (color vs. texture). Two informants approached the box either from a side from which one could see (but not feel) the object or from another side from which one could feel (but not see) it. The informants then made conflicting claims about the property in question (color/texture), justifying them by an utterance of the form "…because I have seen it" (good argument regarding color, poor argument concerning texture) versus "…because I have felt it" (vice versa). Children thus always heard the same form of competing claims (e.g. "It is red/blue") accompanied by utterances of the form "because I have seen…" versus "because I have felt…". And the very same utterance constituted good or bad arguments depending on context. The questions of interest were, first, whether children respond appropriately in their selective belief formation to context-relative argument quality (believing the claim justified by a good argument); second, whether children can explicitly judge argument quality; and third, whether they draw more general dispositional conclusions about the agents as a function of their repeatedly giving good versus poor arguments.

11 | METHOD

11.1 | Participants

Forty-one 4- to 5-year-old children (age range: 48 - 71 months, $M = 59$ months, 20 girls) were included in the final sample (one additional girl lost interest during the familiarization phase).

11.2 | Material

We presented participants with a forced choice task in which they were to decide about the color or softness of a hidden object. Six plush toys were used (see Supplement, Figure S3), two of them were always of the same color but different in texture (soft/hard). Toys were placed in a triangular box (edge length of triangle: 57 cm, height: 44 cm). The box was a modified version of the material used in Studies 1a/b. It consisted of a clear window side for seeing, a side with an opening for feeling and an opaque backside (see Figure 4). Little cases were used to move the toys secretly into the triangular box. Assistants presenting the two choices were wearing identical T-shirts with different colors (green/yellow). To allow the children to visualize choice options and thus alleviate working memory demands, the information and the arguments presented were illustrated using memory cards (see Supplement, Fig. S2) Colors were depicted as color patches, texture was depicted as pillow (for being soft) and stones (for being hard). The modalities "seeing" and "feeling" were depicted as eyes or a hand.

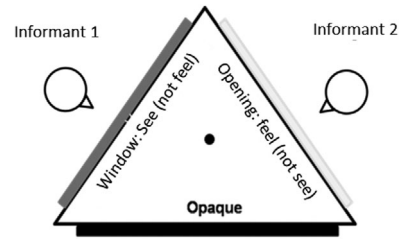

FIGURE 4 Schematic drawing of the box used in Study 2

11.3 | Design

This experiment employed a 2 (argument modality: "…seen it…"/"…felt it…") × 2 (property: color/texture) within subject design. The order in which the informants presented their judgment, on which side each informant was standing when presenting the judgment (left/right), and which of the two assistants was playing the accurate/inaccurate informant was counter-balanced. In a familiarization phase, children received three kinds of practice trials to make sure they understood the general logic of the box and stimuli etc. In the test phase, children received several trials in which their selective trust as a function of argument quality, their explicit judgments, and their generalized dispositional conclusions were tested (see below).

11.4 | Procedure

Test phase. In the beginning of the test phase, children were informed that they were going to participate in a quiz show and answer questions regarding which of two toys was hidden in the box. Two girls would join the game to assist. Then came three kinds of trials in the following order:

1. **Differentiation of arguments**: children were confronted with two informants giving conflicting testimony about color/texture accompanied by arguments of the form "seen it" versus "felt it" (4 trials). The dependent measure was which testimony children would selectively endorse.
2. **Generalization**: children were confronted with (conflicting) testimony by the same two informants in novel trials; the question was to which degree children would generalize trust/skepticism towards these informants to novel trials. (4 trials)
3. **Explicit judgment**: Children were asked explicitly how the informants had performed in their reason-giving and justification. (3 trials)

The overall procedure was similar for all trials. Children sat at a table with an occluder board in front of them hiding the box (see Figure 5). The assistants sat beside the child. The experimenter presented two toys. In color-trials the toys were of the same softness but differed in color and in softness-trials toys were of the same color but differed in softness. Both toys were moved into small boxes and, walking behind the occluder board, the experimenter announced to "secretly hide one of the toys in the triangular box". From behind the occlude, the experimenter called one (or both) assistant(s) behind the occluder. Assistants were instructed to get information from one side of the box. For children, the instruction was hearable but not visible. They thus knew that each assistant could either see or feel inside the box (but not which was the case). The assistants then re-appeared in front of the occluder and presented their information. The specific procedure of the three kinds of test trials was as follows.

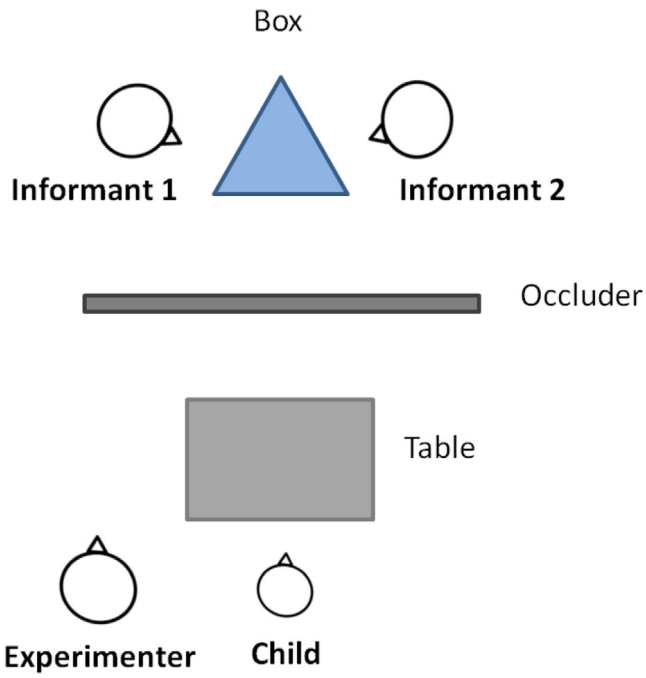

FIGURE 5 Schematic drawing of the setup in Study 2: Informants 1 and 2 obtain visual and tactile information in such ways that it is not visible for the child

In **differentiation of arguments** trials (two color/two softness-trials), two informants presented conflicting claims accompanied by an argument. One informant always gave good arguments ("seen" in color and "felt" in softness trials), and the other poor ones (vice versa). The following phrasing was used to present conflicting information: "I think it is …(e.g., red/soft), because I *looked/felt* into the box". The rationale of these trials was to investigate children's understanding of arguments differing in quality and to establish the two assistant as a high- versus low-quality informants.

There were two types of **generalization trials**: In *single-informant generalization trials*, children were faced with one informant only. Each assistant gave information once in a color- and once in a softness-trial (4 trials total). Behind the occluder board, the single informant went to the box in such a way that the child did not know to which side. Then this informant gave testimony regarding the color/softness: "I think it is …(e.g., red/soft)" without any complementary argument.

In *two-informant generalization trials* (one color/one softness-trial), two informants presented conflicting answers. Behind the occluder board, assistants were instructed by "You go to the window side and you go to the opening side" in such ways that the child did not know who went where. Both assistants then gave diverging testimony without any arguments. For example, in a color trial, one informant stated "I think it is blue" and the second informant stated "I think it is pink".

At the end of the test phase, **explicit judgment trials** were administered. The assistants left the test room and the experimenter asked three explicit forced choice questions.

Q1: Individual explicit judgments:

Q1a/b: "How did the assistant in the green/yellow t-shirt give reasons for her judgment, *good or not so good*?"

Q2: Contrastive explicit judgment: "Who gave *better* reasons, the assistant in the green or the one in the yellow t-shirt?"

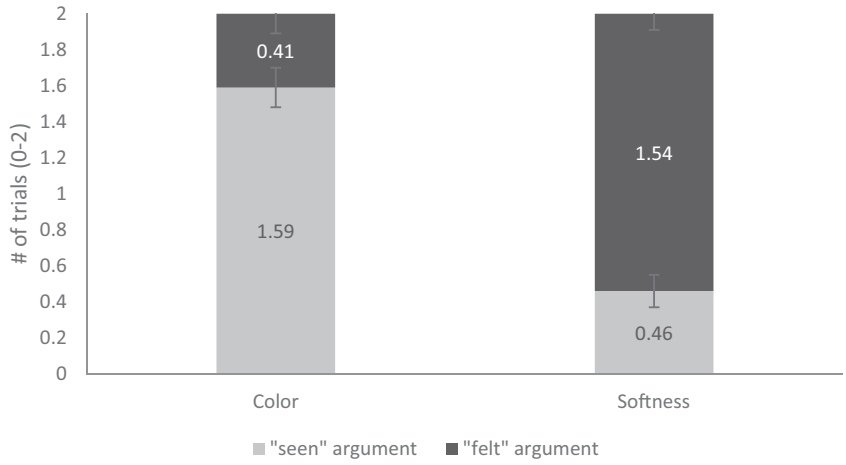

FIGURE 6 Mean number of color/property trials in which children believed the informant with the "seen"/"felt"-arguments (0-2) as a function of property (color/softness) to be found out. Error bars indicate standard error of the means

12 | RESULTS

12.1 | Differentiation of arguments

Children's endorsement of the testimony given by the informants is depicted in Figure 6 for color/texture trials. Paired *t*-tests revealed that children endorsed the testimony by the informant with the high quality argument more often than the testimony by the low-quality informant for both color ($t(40) = 5.31$, $p < .001$, $d_z = 0.84$) and softness ($t(40) = 5.77$, $p < .001$, $d_z = 0.91$). This pattern did not differ in degree between both kinds of properties ($t(40) = 0.35$, $p = .728$, $d_z = 0.05$).

12.2 | Generalization trials

Single-informant generalization trials. The proportions of trials in which children believed the two informants are depicted in Figure 7a. Children responded in testimony-consistent ways in color-trials and softness-trials when presented with the informant who had previously given good arguments (color: binomial, $p = .012$; softness: binomial, $p = .001$). However, when presented with the informant who had previously given poor arguments, children responded in testimony-consistent ways in color-trials only (binomial, $p < .001$), but not in softness-trials (binomial, $p = 1.000$).

Two-informant generalization trials. Children's choices of the informants with the previously good arguments in the color and softness trials are depicted in Figure 7b. In color-trials, children chose at chance level (binomial, $p = .349$), whereas, in softness-trial children preferred the good-argument-informant (binomial, $p = .028$).

12.3 | Explicit judgments

Explicit judgments about single-informants. Explicit judgments of each informant as a function of previous arguments are depicted in Figure 8a. Children judged that the informant with the previously good arguments gave good reasons (binomial, $p < .001$), but did not do so regarding the informant with the previously poor arguments (binomial, $p = .349$).

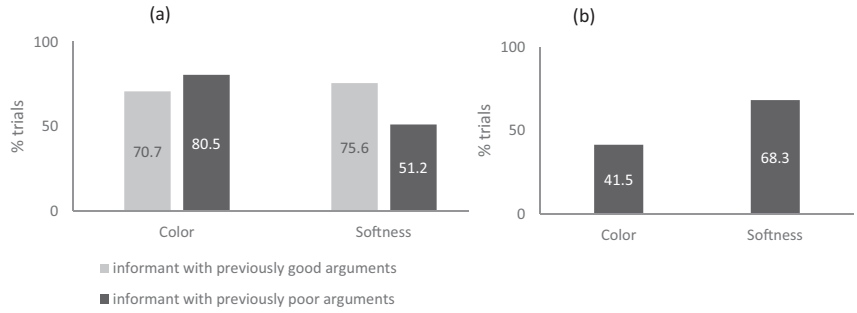

FIGURE 7 Proportions of trials (a) in which participants believed each informant in single-informant generalization trials; and (b) in which participants preferred the informant with the previously good arguments in contrastive two-informant generalization trials

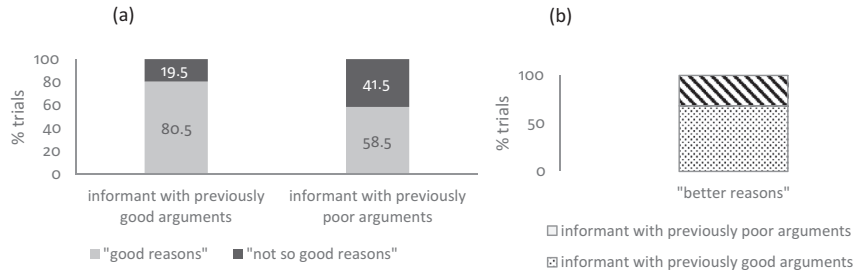

FIGURE 8 Proportion of trials (a) in which participants judged that each informant had "good/not so good reasons" in the individual explicit judgment trials; and (b) in which participants preferred each agent as having had "better reasons" in the contrastive explicit judgment trials

Explicit contrast judgments. The percentage of trials in which children chose each agent in response to the question who gave "better reasons" are depicted in Figure 8b. Children chose the informant with the previously good arguments more often than the informant with the previously poor arguments (Binomial test, $p = .028$).

13 | DISCUSSION

The main findings of Study 2 were the following: First, when confronted with two informants who gave conflicting testimony, children distinguished between them in terms of the appropriateness of the arguments they gave for the testimony. When the question was which color an object had, they selectively believed the testimony of the speaker who justified his claim by "seen it" arguments rather than a speaker who used a "felt it" argument, and vice versa for questions of texture. Second, children distinguished between the informants also in their explicit judgments, claiming that the informant who had given good arguments was the one who "gave better reasons". Third, however, children showed little inclination to draw more generalized conclusions about informants after they had repeatedly and consistently given good or poor arguments. There was a tendency for such generalizations in the softness trials (children did not trust an informant with previously poor arguments invariably, and did prefer the informant with the previously good arguments), but no such pattern was found in color trials. This seems consistent with similar results for better performance in "feel" versus "see" trials in previous studies (O'Neill et al., 1992; Study 1). Whether this is a chance finding or represents some substantial difference between color and softness trials (such that children reliably draw

generalized conclusions in the latter but not in the former) remains unclear from the present study alone and needs to be addressed in future research.

14 | GENERAL DISCUSSION

Do young children understand and evaluate arguments in context-sensitive ways in their selective social learning and belief revision? This was the guiding question of the present research. The main findings were the following: First, children did evaluate the quality of arguments in context-sensitive ways, grasping that one and the same utterance can constitute a good argument in one context, but fail to do so in another. Second, they made these evaluations on the basis of the quality of perceptual access that differed between contexts (Study 1b), and on the basis of the modality of perceptual access (Study 2). Third, children explicitly expressed this context-relative argument evaluation, and practically put it to use in their selective belief formation (Study 2) and belief revision (Study 1b). They learned more from and revised beliefs more in response to an informant who used a given utterance in such ways that it constituted a good argument. Fourth, there was no strong and unambiguous evidence that children would draw dispositional generalizations from informants' argument use to their more general epistemic status.

These findings complement what we knew from previous research on children's developing understanding of arguments. In particular, the present research goes beyond existing results in showing that young children engage in the evaluation of argument quality in flexible and sophisticated ways not tied to superficial properties of the argument utterances: children can distinguish the quality of two arguments even in the absence of any differences in form (syntax, length, narrow informational content, use of connectives etc.).

These new insights raise many new questions for future research: First, what are the ontogenetic origins and early developmental trajectories of sophisticated evaluation of argument quality? Previous work has shown that children from as young as two years of age distinguish between arguments adduced by others according to their quality (Castelain et al., 2018). But in these studies, argument quality was always coupled with some superficial formal properties such as word length, syntactic complexity or narrow semantic content. It remains thus to be seen when the capacity develops to evaluate the quality of arguments as warranting the to-be-supported-conclusions to stronger or lesser degrees in purely context-relative ways, independently from any local properties of the utterances. Relatedly, once children engage in context-relative evaluation of argument quality like in the present studies, how flexible, sophisticated and general is or becomes this capacity? For example, when do children come to understand that subjective argument strength is a matter of background beliefs (Toulmin, 1958)? To illustrate this point, imagine the customer and the male shopkeeper in the Dead Parrot Sketch both have recently read (and mutually know about each other that they have recently read) an article about a new neuroscience discovery: that cold viruses temporarily affect face recognition areas in the brain so that patients often confuse male and female faces. Then "I am sorry, I have a cold" would be a perfectly fine argument to explain the customer's confused address. Or to take another example of sophisticated context-relativity, think of seemingly empty, circular arguments like "It's a fish…because I saw it's a fish" (as used in Castelain et al., 2018). Again, such an utterance may constitute a very good argument in some other context (for example, uttered by a zoologist with extensive training in visually based sortal classification of animals). The question now, to be addressed in future research, is how general children's initial context-relative evaluation of argument quality is, and how it develops: When do children come to understand such subtle context effects related to mutual subjective background beliefs and other contextual factors? Relatedly, the present findings complement other recent results that children from around age 4 track something like speaker calibration (e.g., Birch et al., 2020): from this age, children take into account how well calibrated a speaker is in her expression of confidence, ignorance, hesitation etc. Previous research has documented this for selective learning generally, and the present findings extend this to the case of children's argument evaluation. Many open questions remain for future research. Children's capacity to track speaker calibration in selective trust does not come as an all-or-nothing package, but reveals protracted development through middle childhood. For example, children track the appropriateness of expressing confidence long before they

can judge the appropriateness of hesitating under uncertainty (Birch et al., 2020). Future research thus will need to investigate analogously and in more depth and detail how children's capacity develops to tracking various forms of speaker calibration in the context of argument evaluation.

Finally, when do children begin to engage in more robust dispositional generalizations not only about individual utterances as arguments, but about the individuals making the arguments? Study 2 tested for such generalizations and did not find clear evidence of robust tendencies to draw dispositional conclusions – although children did explicitly remember the episodes of good versus poor argumentation. The fact that children tended to equally believe previously good and poor argument-givers may reflect their general forgivingness, or their known tendency towards naïve trust (Jaswal et al., 2010). Relatedly, children this age may stick to purely episodic representations when it comes to argument-use: they may merely represent individual arguments ("she gave a good argument A, a good argument B etc.") without jumping to general conclusions about the person operating with the arguments ("she is good at giving arguments" or even "she is reliable"). Again, future research will need to address these open questions.

In sum, the present studies show that children from age 4 evaluate arguments in context-relative ways, understanding that one and the same utterance can constitute better or worse arguments as a function of the speaker's perceptual access quality or modality; and they put to use this evaluation in selective learning and belief-revision. How flexible and sophisticated this argument evaluation is, and whether it is connected with dispositional generalizations about others who give arguments remains to be clarified.

ENDNOTES

[1] In fact, even though adults clearly have the competence to critically evaluate argument quality, they often fail to translate this into performance and fall victim to empty and circular ("placebic") arguments (Langer et al., 1978).

[2] https://www.youtube.com/watch?v=vZw35VUBdzo

[3] The reason for always having the trials without arguments at the beginning were the following: First, the trials without arguments served as a kind of baseline against which trials with arguments were to be contrasted. Second, we wanted to test whether the previous findings in these trials could be replicated in the first place. Third, from a pragmatic point of view it makes perfect sense to have trials without arguments followed by trials with arguments, but not vice versa. If previous trials were accompanied by arguments, one would wonder why all of a sudden the informants have stopped giving arguments.

[4] A lower limit of -100% (instead of the more common 0%) for the truncation of AT scores ensures that—in case the judge actually ignores the advice—unsystematic differences between initial and final estimates do not artificially lead to positive AT scores

ACKNOWLEDGMENTS

This work was supported by the German Science Foundation (grant for the Research Training Group 2070 "Understanding Social Relationships"). We would like to thank Stefan Schulz-Hardt for support, cooperation and feedback, and Marlen Kaufmann, Konstanze Schirmer, Marika Reimer, Ariane Bodemeyer, Josefin Johannsen, Lisa Wenzel, Sophia Wild, and Johanna Andratschke for help with recruitment and data collection. We are grateful to Franziska Brugger, Marina Proft, Britta Schünemann, Friederike Schütte, and Lisa Wenzel for helpful discussion and feedback.

DECLARATION OF INTEREST

None.

DATA AVAILABILITY STATEMENT

All the data and relevant analyses will be made available at: https://www.psych.uni-goettingen.de/en/development

REFERENCES

Bernard, S., Mercier, H., & Clément, F. (2012). The power of well-connected arguments: Early sensitivity to the connective because. *Journal of Experimental Child Psychology*, *111*(1), 128–135. https://doi.org/10.1016/j.jecp.2011.07.003

Birch, S. A. J., Severson, R. L., & Baimel, A. (2020). Children's understanding of when a person's confidence and hesitancy is a cue to their credibility. *PLoS ONE*, *15*(1), e0227026. https://doi.org/10.1371/journal.pone.0227026

Castelain, T., Bernard, S., & Mercier, H. (2018). Evidence that two-year-old children are sensitive to information presented in arguments. *Infancy*, *23*(1), 124–135. https://doi.org/10.1111/infa.12202

Castelain, T., Bernard, S., Van der Henst, J. B., & Mercier, H. (2016). The influence of power and reason on young Maya children's endorsement of testimony. *Developmental Science*, *19*(6), 957–966. https://doi.org/10.1111/desc.12336

Corriveau, K. H., & Kurkul, K. E. (2014). Why Does Rain Fall?": Children prefer to learn from an informant who uses noncircular explanations. *Child Development*, *85*(5), 1827–1835. https://doi.org/10.1111/cdev.12240

Dunn, J., & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology*, *23*(6), 791–798.

Fitneva, S. A. (2009). Evidentiality and trust: The effect of informational goals. *New Directions for Child and Adolesc ent Development*, *2009*(125), 49–62. https://doi.org/10.1002/cd.249

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

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*(1), 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

Hermes, J., Behne, T., & Rakoczy, H. (2018). The development of selective trust: Prospects for a dual-process account. *Child Development Perspectives*, *12*(2), 134–138. https://doi.org/10.1111/cdep.12274

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development*, *83*(3), 1051–1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x

Koenig, M. A., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, *52*(3), 423–429. https://doi.org/10.1037/dev0000089

Köymen, B., O'Madagain, C., Domberg, A., & Tomasello, M. (2020). Young children's ability to produce valid and relevant counter-arguments. *Child Development*, *91*(3), 685–693. https://doi.org/10.1111/cdev.13338

Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision-making by preschool peers. *Cognitive Development*, *32*, 74–85. https://doi.org/10.1016/j.cogdev.2014.09.001

Köymen, B., & Tomasello, M. (2018). Children's meta-talk in their collaborative decision making with peers. *Journal of Experimental Child Psychology*, *166*, 549–566. https://doi.org/10.1016/j.jecp.2017.09.018

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*. (pp. 301–326). Hove: Psychology Press/Taylor & Francis (London, UK).

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260. https://doi.org/10.1111/1467-8624.00605

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, *36*(6), 635–642. https://doi.org/10.1037/0022-3514.36.6.635

Mercier, H., Bernard, S., & Clément, F. (2014). Early sensitivity to arguments: How preschoolers weight circular arguments. *Journal of Experimental Child Psychology*, *125*, 102–109. https://doi.org/10.1016/j.jecp.2013.11.011

Mercier, H., Sudo, M., Castelain, T., Bernard, S., & Matsui, T. (2018). Japanese preschoolers' evaluation of circular and non-circular arguments. *European Journal of Developmental Psychology*, *15*(5), 493–505. https://doi.org/10.1080/17405629.2017.1308250

Miosga, N., Schultze, T., Schulz-Hardt, S., & Rakoczy, H. (2020). Selective social belief revision in young children. *Journal of Cognition & Development*, *21*(4), 513–533. https://doi.org/10.1080/15248372.2020.1781127

O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, *63*(2), 474–490.

Rakoczy, H., Ehrling, C., Harris, P. L., & Schultze, T. (2015). Young children heed advice selectively. *Journal of Experimental Child Psychology*, *138*, 71–87. http://doi.org/10.1016/j.jecp.2015.04.007

Robinson, E. J., Butterfill, S. A., & Nurmsoo, E. (2011). Gaining knowledge via other minds: Children's flexible trust in others as sources of information. *British Journal of Developmental Psychology*, *29*(4), 961–980. https://doi.org/10.1111/j.2044-835X.2011.02036.x

Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, *35*(2), 535–546. https://doi.org/10.1037/0012-1649.35.2.535

Robinson, E. J., & Whitcombe, E. L. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74*(1), 48–62. https://doi.org/10.1111/1467-8624.t01-1-00520

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*(4), 1054–1070.

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the inability to ignore useless advice. *Experimental Psychology*, *64*(3), 170–183. https://doi.org/10.1027/1618-3169/a000361

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*(4), 779–797. https://doi.org/10.1037/a0034191

Tenney, E. R., Small, J. E., Kondrad, R. L., Jaswal, V. K., & Spellman, B. A. (2011). Accuracy, confidence, and calibration: How young children and adults assess credibility. *Developmental Psychology*, *47*(4), 1065–1077. https://doi.org/10.1037/a0023273

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rakoczy, H., Miosga, N., & Schultze, T. (2022). Young children evaluate and follow others' arguments when forming and revising beliefs. *Social Development*, *31*, 147–164. https://doi.org/10.1111/sode.12533