Journal of Experimental Child Psychology 270 (2026) 106547

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Reflective belief revision before the age of reason

Kirsten H. Blakey [a,b,c,*], Chloe L. Dow [a,b,d], Ariane Veit [e], Brina Recelj [a,b,d], Zsófia Virányi [e], Giacomo Melis [a], Eva Rafetseder [b]

[a] Philosophy, Faculty of Arts and Humanities, University of Stirling, Stirling, FK9 4LA, UK
[b] Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, UK
[c] Psychological & Brain Sciences, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada
[d] Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, SE5 8AF, UK
[e] Messerli Research Institute, University of Veterinary Medicine, Vienna, Medical University of Vienna, University of Vienna, Veterinaerplatz 1, 1210 Vienna, Austria

ARTICLE INFO

Keywords:
Belief revision
Reflective thinking
Rational thinking
Language development
Epistemic defeaters

ABSTRACT

Young children can revise beliefs in light of new evidence, yet some philosophers argue that this is not true rationality. They define rationality as the capacity to reflect on and re-evaluate one's reasons for beliefs, tying it to language and proposing that it emerges only around the so–called "age of reason" (age 6). In challenge to this, we tested whether children aged 3 to 6 ($N = 93$) were capable of a basic form of reflective responsiveness to reasons: acquiring and responding to an undermining defeater—evidence that challenges the grounds for a belief—through exposure to overriding defeaters. In each trial, before choosing where to search, children observed a Reliable informant (100% accurate) and an Unreliable informant (50% accurate, 50% misleading) indicating one of two reward locations. Each misleading trial served as an overriding defeater, directly contradicting children's initial belief about the reward location. Age and language predicted children's tendency to follow the Reliable over the Unreliable informant suggesting they may have acquired an undermining defeater related to informant reliability. However, neither age nor language predicted preferences when the informants were pitted against each other in a new context. These findings suggest some basic capacity for reflective responsiveness to reasons may emerge earlier than the "age of reason", more independently of language than some philosophical accounts have assumed, and may be grounded in capacities (overriding defeaters) already present at the unreflective level.

Our beliefs are constantly challenged and updated in response to new evidence, whether gathered from personal experience or from others via testimony, news, or social media. Yet not all evidence warrants belief revision—particularly when it is mixed or weaker than the original evidence. Adults can reflectively assess incoming evidence by evaluating the reliability of its source, which allows them to recognise when evidence may be misleading (e.g., an unreliable source), though they are not immune to being misled.

Young children also acquire information from others: they distinguish between reliable and unreliable informants and selectively trust those they deem reliable (Chow et al., 2008; Harris et al., 2018; Rakoczy et al., 2009; Tummeltshammer et al., 2014), and they revise their beliefs in response to new evidence (Kimura & Gopnik, 2019; Király et al., 2018, 2023; Schleihauf et al., 2022). A key

* Corresponding author at: Psychological & Brain Sciences, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada.
E-mail address: kirstenhblakey@gmail.com (K.H. Blakey).

https://doi.org/10.1016/j.jecp.2026.106547
Received 3 October 2025; Received in revised form 7 May 2026;
Available online 30 May 2026

question is whether performing these tasks requires children to identify and assess reasons for their beliefs, which would indicate that they are reflecting on those reasons *as* reasons, or whether children instead rely on unreflective associations or heuristic biases without reflecting on their reasons (Blakey et al., 2021). This ambiguity arises partly because evidence remains consistent across the familiarisation and test trials in most tasks, potentially allowing children to form unreflective associations or use responses based on heuristics rather than actively evaluating the reliability of a source (Blakey et al., 2021; Heyes, 2016). Incorporating changes in context across trials could help prevent children from relying on associations formed in preceding trials (Blakey et al., 2025b; Melis & Blakey, 2025). Kidd et al. (2013) implemented this approach by introducing a change in context between establishing the researcher's reliability and assessing children's response to that reliability. Children aged four were given the choice to use a set of well-worn crayons immediately or wait for the researcher to return with new art supplies. The *unreliable* researcher returned without the promised art supplies, whereas the *reliable* researcher fulfilled their promise. Children were then presented with a single marshmallow and told they would receive two marshmallows if they waited until the researcher returned. Children who had experienced the reliable researcher waited considerably longer than those who experienced the unreliable one, suggesting that their wait times were influenced by reasoned expectations about the researcher's reliability. However, children in the reliable condition may have been in a more positive mood, having received the promised art supplies and not experiencing the unreliable researcher. This mood difference may have influenced their waiting time without reasoning about reliability, leaving open the question of whether children were genuinely reflecting on the researcher's reliability.

Some philosophers who characterize rationality as the capacity to respond to reasons or evidence attest that rationality requires the capacity to reflect on and re-evaluate one's reasons for beliefs (Boyle, 2018; Korsgaard, 2018; Marcus, 2021; McDowell, 1994). They tie this capacity to language, specifically to the ability to reply to requests for reasons, such as "why do you believe that P?" (e.g., Boyle, 2018; Korsgaard, 2018). A key consequence is that non-linguistic or minimally verbal subjects, including young children who are still developing language skills and animals, would not be considered capable of rational belief formation or revision. These accounts of rationality have been developed, to the best of our knowledge, independently of empirical research in developmental or comparative psychology—a gap we intend to address. By contrast, other philosophers propose that *unreflective* belief revision or responsiveness to evidence is sufficient for rationality (Conee et al., 2004; Kornblith, 2012; Williamson, 2000), and available to non-linguistic populations (Danón & Kalpokas, 2023; Dretske et al., 2006; Glock, 2019; Hurley, 2006; Mercier & Sperber, 2018). Empirical research in developmental psychology and comparative cognition similarly supports the view that young children and animals are capable of rationality in a variety of forms (Buttelmann et al., 2008; Gergely et al., 2002; Kimura & Gopnik, 2019; O'Madagain et al., 2022). There are thus two conflicting characterizations of what it takes to be a rational subject: one that entails being able to reply to verbal requests for reasons and one that is independent of language. Bridging the gap between them requires investigating how unreflective belief revision—which appears to be present both in adults and non-linguistic populations—relates to the reflective belief revision that some view as language-dependent. For example, it is important to investigate empirically how the capacity to identify and assess reasons relates to the capacity to answer "why" questions. If one can identify and evaluate reasons and yet be unable to respond appropriately to requests for reasons—or without possessing language altogether—this would call into question the view that language is necessary for rationality. Central to this investigation is understanding how exposure to counterevidence, initially processed unreflectively, can support the acquisition of a basic form of reflective thinking.

Given the importance of the notion of responding to evidence, it is surprising that the debate has so far overlooked relevant contributions from analytic epistemology. In particular, *epistemic defeaters* (Pollock, 1974) offer a promising tool for examining the transition from unreflective to basic reflective thinking. Normally, beliefs are formed on the basis of evidence, whether from testimony or first-hand experience, that provides a reason for believing a proposition. Epistemic defeaters are simply counterevidence that give reasons against believing a proposition, and can be divided into two types: overriding defeaters and undermining defeaters. *Overriding defeaters* are reasons to give up an original belief < P > and adopt its negation < not P >. They can be acquired unreflectively by simply accepting the proposition supported by the counterevidence, without the need to identify and assess the reasons. Responding to overriding defeaters is a basic skill of any subject capable of forming beliefs in response to new evidence (e.g., changing environment). For example, a child can believe that the toy is in the drawer but, upon opening the drawer and finding no toy, revise their belief to < the toy is not in the drawer >. In contrast, *undermining defeaters* give one a reason to give up a belief without necessarily suggesting that one should replace it with belief in its negation. Responding to undermining defeaters is typically more demanding than responding to overriding ones. Some undermining defeaters are acquired through testimony suggesting that something was wrong with how the belief was formed, such as an unreliable source of evidence—thereby prompting reflection on the process through which the belief was formed. Importantly, undermining defeaters can also be acquired through first-hand experience, by making an inference about a source of information after experiencing multiple overriding defeaters from the same source. For example, if a colleague frequently fails to uphold promises, each instance acts as an overriding defeater, whereby one's belief changes from < colleague does X > to < colleague does NOT do X >. After experiencing several overriding defeaters coming from the same colleague, one might infer that this colleague is unreliable. Making such an inference indicates that an undermining defeater has been acquired, the response to which should be reduced confidence or suspension of judgment in the proposition that the colleague will deliver on any future promises. Acquiring and responding to an undermining defeater in this sense involves identifying the evidence and assessing it as potentially misleading (Melis & Monsó, 2023). If so, responding to undermining defeaters is a form of reflective responsiveness to reasons. Importantly, responding to undermining defeaters does not appear to require the capacity to reply to verbal requests for reasons. Thus, testing whether non-verbal or minimally verbal subjects are capable of responding to undermining defeaters is a way to test empirically the theory that language is necessary to respond to reasons.

Recent empirical studies suggest that young children and animals might possess the reflective abilities needed to process undermining defeaters. For example, O'Madagain et al. (2022) examined whether great apes or 3- and 5-year-old children could assess their

reason for a belief against new physical or social evidence. In Study 2, subjects watched an experimenter place a reward in one of two boxes. The boxes were then rotated so a social partner could see the location of the reward and pick that location, which was either consistent with or conflicted with the subject's expectation of the reward location. Subjects were then allowed to seek additional information by peeking inside the boxes from the top. Both 3- and 5-year-olds peeked more often before choosing a box in the conflicting condition. In Study 1, in which conflicting evidence was based on a physical manipulation, great apes and 5-year-olds, but not 3-year-olds, more often sought additional information. Increased information seeking in response to conflicting evidence was taken by the original authors as an indication that subjects examined the reasons for their belief. We suggest however, that a different view needs to be considered. The conflicting evidence may have acted as an overriding defeater with the same strength as their initial belief, and as the initial and new evidence were of comparable strength, the new evidence did not compel immediate belief revision. Instead they may in fact have been responding to being in a state of uncertainty rather than recognising that they were uncertain (Perner, 2012), which would also lead to information seeking. Note that this need not involve explicit assessment of reasons.

Schleihauf et al. (2022, Study 3) offer a clear example of responding to undermining defeaters acquired via testimony in 4- to 5-year-old children. In this study, seeing that footprints led to one location but not to another, children initially formed a belief that an animal was hiding in location A rather than B. The reason for their belief that the animal was hiding in location A was subsequently either confirmed or disconfirmed by an agent. The confirming and disconfirming reasons related to the evidence that children used to form their initial belief—confirming reasons supported the relevance of the evidence (i.e., confirmed the footprints as coming from the animal being searched for); while disconfirming reasons challenged the relevance of the evidence (e.g., *"These don't look like bird footprints. Look! Bird footprints look like this. These footprints don't look like that"*). In both conditions, a second agent then provided a plausible alternative reason to believe the animal was hiding in B. Children were more likely to revise their initial belief when they had been presented with a disconfirming reason beforehand. Encountering this disconfirming reason could serve as an undermining defeater, suggesting that something was wrong with how their initial belief was formed. This may have prompted them to suspend judgement regarding the animal's location, rather than replacing their initial belief with its negation. These findings suggest that 4- and 5-year-old children may be able to respond to undermining defeaters. However, because the use of testimonial evidence relies on language skills, this procedure cannot be used to assess whether younger children with less developed language abilities can acquire and respond to undermining defeaters. This is why we focus on cases in which undermining defeaters are acquired through first-hand experience rather than testimony.

Two recent studies have explored the acquisition of and response to undermining defeaters through inference over repeated overriding defeaters. These studies used a non-verbal task with 2-year-old children, dogs, and pigs (Blakey et al., 2025b), as well as adults (Blakey et al., 2025a). They aimed to determine whether—after experiencing multiple overriding defeaters—subjects could infer that evidence from a particular source was misleading and thereby acquire an undermining defeater like < the source is unreliable >. In addition to testing whether non-verbal subjects are capable of responding to undermining defeaters, these studies investigated whether the acquisition of an undermining defeater may be grounded in capacities already available at the unreflective level; namely, the capacity to respond to overriding defeaters and the capacity to make simple generalisations. If so, there is reason to think that reflective responsiveness to reasons both emerges from, and is continuous with, the unreflective one (see Melis & Blakey, 2025, for a detailed discussion).

In both Blakey et al. (2025a) and Blakey et al. (2025b), subjects watched informants use different actions to indicate one of two locations. A *Reliable informant's* actions consistently indicated the location of a reward, while an *Unreliable informant's* actions were independent of the reward location, indicating the reward in only 50% of trials. Subjects then had to choose whether to follow the evidence provided by the informant. If subjects followed the evidence in non-misleading trials, they received a reward. In misleading trials, following the evidence led to no reward and the chance to see that the reward had really been in the other location. If subjects did not follow the informants' evidence, they also received feedback on the reward location, either by obtaining it directly (in misleading trials) or by seeing where it was hidden (in non-misleading trials). Results of the first study revealed that neither 2-year-olds nor animals differentiated between Reliable and Unreliable informants. Rather, they showed a reduction in following the evidence of both informants across actions and did not show a preference for either informant in subsequent (Transfer) trials in which the informants were pitted against each other (Blakey et al., 2025b). One explanation for this is that subjects may have processed an undermining defeater related to both informants, inferring that both informants are unreliable. Such a response may have occurred because the cumulative strength of the overriding evidence was not sufficient for subjects to differentiate between the Reliable and Unreliable informants. The second study conducted a tablet-based version of the task with adults, investigating how varying the strength—in relation to quantity and quality—of the evidence available after making a choice affected adults' ability to infer informants' reliability (Blakey et al., 2025a). In a weak feedback condition, choosing the correct location revealed the reward, while choosing the incorrect location resulted in no reward. In a strong feedback condition, participants received additional cues: a green check mark indicated a correct choice, and a black X marked an incorrect choice, with the reward appearing in the alternative location after a brief delay. Unlike 2-year-olds and animals, adults differentiated between the informants, following the Unreliable informant's evidence less often and consistently preferring the Reliable informant in subsequent Transfer trials, but only in the strong feedback condition. These results confirmed that acquiring and responding to an undermining defeater is indeed dependent on the strength of the evidence and its counterevidence. Despite having been provided with similarly strong feedback in Blakey et al. (2025b), 2-year-old children and animals did not show clear evidence of this ability (though they may not have accessed the feedback despite the reward behind the unchosen location being visible). Therefore, the question remains whether minimally verbal children and animals are capable of acquiring and responding to undermining defeaters, or whether doing so does require more developed language skills.

The current study aimed to address this question by running the strong feedback condition of the tablet-based task, used previously with adults (Blakey et al., 2025a), with 3- to 6-year-old children. We were interested in age-related changes in children's ability to

acquire and respond to undermining defeaters, specifically the age at which children begin to discriminate between reliable and unreliable sources, as adults do. Our focus was on 3- to 6-year-olds since 2-year-olds did not differentiate between the Reliable and Unreliable informants in the physical version of the task (Blakey et al., 2025b), and some philosophers argue that children acquire reflective abilities when they can respond to why questions at the "age of reason"—typically around 6 years old. If children can acquire and respond to an undermining defeater related to an informant's reliability, they should show a particular pattern of results across the different actions used by the informants. Initially, children would follow the evidence of both the Reliable and Unreliable informants, since they have no prior experience to suggest unreliability. After experiencing repeated instances of being misled (i.e., they are exposed to overriding defeaters), they would be less willing to follow the Unreliable informant's evidence while continuing to follow evidence from the Reliable informant in later actions, thus indicating that they may have inferred the informants' reliability and assessed that the Unreliable informant could also be misleading in a new context. Children's responses on the *first trial* of each action with each informant were expected to be particularly informative, as they could reveal whether children inferred the informants' reliability without needing to relearn that a new type of evidence presented by the same Unreliable informant could also be misleading. Similarly, we expected that if children had acquired an undermining defeater, they would show a preference for the Reliable informant in the subsequent *Transfer trials* in which the informants were pitted against each other.

To the best of our knowledge, advocates of the view that language is necessary for the identification and evaluation of reasons have not referenced research in developmental psychology. Given this gap, to examine the relationship between language and children's performance, we also assessed their receptive and expressive language skills. Developed linguistic skills are likely sufficient for acquiring and responding to undermining defeaters and for engaging in reflective thought, but the question is whether they are necessary. As outlined earlier, they may not be, as undermining defeaters can in principle be acquired from repeated exposure to overriding defeaters, which are available at the unreflective level. To explore the connection between language and reflective skills, we used a non-verbal task with children aged 3 to 6 years old. This age range thereby includes children who are younger than the so-called "age of reason". If developed language is necessary to respond to undermining defeaters, we would expect that—even in this non-verbal task—only children with higher language scores would show the predicted pattern of differentiating between reliable and unreliable informants. If developed language facilitates without being necessary, children with lower language scores may still acquire and respond to an undermining defeater, suggesting that the connection between language and reflective thinking may not be as strong as typically assumed.

Method

Participants

The final sample included 93 children aged 3 to 6 years (44 female, 49 male, *M* age = 63 months, SD = 13.1 months), recruited in Scotland. The sample was predominantly British, as identified by parents/guardians. Eighty-two of these children were included in the language analysis (39 female, 43 male, *M* age = 63.8 months, SD = 13.2 months). An additional 27 children were excluded from all analyses due to not reaching the criterion of evidence following ($n = 19$, see Procedure) or missing data based on preregistered criteria ($n = 8$; completing fewer than 25 trials: $n = 3$, timed out on more than 50% of completed trials: $n = 5$). Thirty-eight children participated in a quiet area in their school or nursery, while 55 participated in a lab at the University of Stirling. This study was approved by the General University Ethics Panel at the University of Stirling (GUEP 2023 3029 10554).

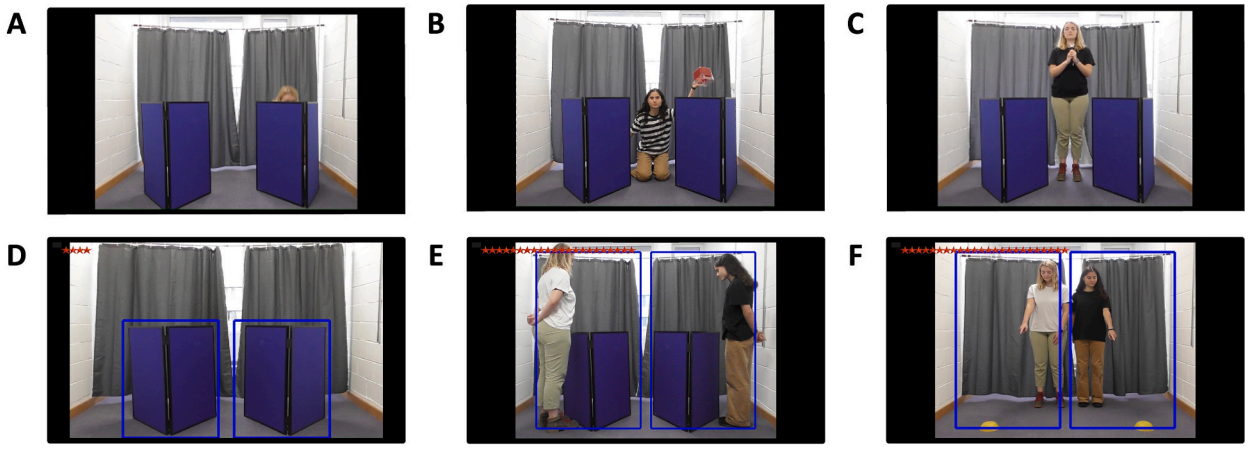

Fig. 1. Panels **A–C** depict the different action types presented during the Demonstration trials: **A.** Crouching, **B.** Lifting, and **C.** Sound. Panels **D–F** show the different choice screens used in: **D.** Demonstration trials. **E.** Screen Choice Transfer trials, and **F.** Pointing Transfer trials.

Materials

Belief revision task

The task was run on a Microsoft Surface tablet, using the touchscreen to make choices. Participants also wore Klugmia over-ear headphones. The task was the same as that used with adults in the strong feedback condition in (Blakey et al., 2025a). We used PsychoPy v2023.2.2 (Peirce et al., 2019) to run the original stimuli and task code (available on OSF: https://osf.io/xhe8t/; Blakey et al., 2024).

The task included 44 Demonstration trials and 6 Transfer trials (for order see supplementary Table S3). In Demonstration trials, children watched a video of one of three female informants holding a reward (red star) and then using an action (crouching, lifting, or sound) to indicate one of two locations (blue V-shaped screens). The actions involved crouching behind one of the two screens (Fig. 1A), lifting a box above a screen while crouching between the screens (Fig. 1B), and ringing a bell that could be heard through one side of the headphones while standing between the screens (Fig. 1C). The three informants played the roles of the Demonstrator, the Reliable informant, and the Unreliable informant (wearing differently coloured shirts: black, white, and striped). The Demonstrator and the Reliable informant always indicated the reward location, while the Unreliable informant indicated the reward location in only 50% of trials. The Reliable informant trials came before the Unreliable informant in the crouching and lifting trials, but to mitigate order effects, the order of informants was reversed in the sound trials.

After watching the video, children could choose which location to search for a reward (Fig. 1D). Transfer trials presented the Reliable and Unreliable informants together, each indicating a different location. In the four Screen Choice trials, the informants stood beside the outer edges of the screens, looking down behind them (Fig. 1E). In the two Pointing trials, informants stood next to each other pointing and gazing at upturned bowls (Fig. 1F). Action and reward locations were pseudorandomized to avoid more than two consecutive trials with the same action side or reward side.

Language task

The Test of Early Language Development 3 (TELD-3; Hresko et al., 1999) was used to assess receptive and expressive language skills. The test includes two subtests that provide separate scores for receptive and expressive language. American English terms were changed to British English equivalents for this study.

Procedure

Belief revision task

The experimenter read aloud the on-screen instructions, explaining that children would be looking for rewards and had five seconds to make a choice. A practice trial was conducted to ensure children understood how to make a choice. During this practice trial, an image of one of the reward locations appeared within a blue rectangular box, and the experimenter said, *"Let's practice how to look for a reward. Swipe inside the blue box to look for a reward"*. The experimenter also demonstrated the swiping motion on the table.

Before the first trial, an image of the Reliable informant, the Unreliable informant, and the Demonstrator (in the centre) was shown for five seconds. In each trial, children received feedback about their choice or, if they did not make a choice within five seconds, saw a timeout message, "You ran out of time" (Fig. 2A). When children chose the rewarded location, the reward (a red star) and a small green

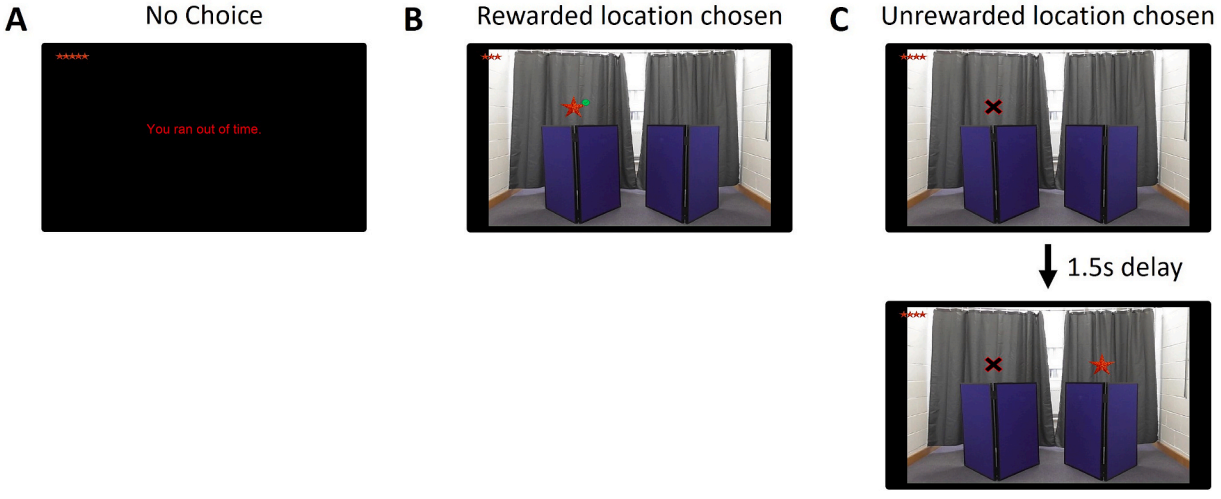

Fig. 2. Illustration of the feedback children received. **A.** If no choice was made within five seconds, the trial timed out. **B.** Choosing the rewarded location revealed the reward and a small green check mark. **C.** Choosing the unrewarded location revealed an X, and after a 1.5-second delay, the reward appeared above the other screen. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

check mark appeared (Fig. 2B). Choosing the unrewarded location resulted in a black X with a red outline, followed by the reward appearing (without a check mark) above the unselected location after a 1.5 s delay (Fig. 2C). The task lasted approximately 12 min, and children could track their success via a reward bar at the top of the screen. When necessary, the experimenter used verbal prompts to direct the child's attention to the task.

Analyses included only the trials with the Reliable and Unreliable informants. To be included in the analysis children had to choose the location indicated by the Demonstrator in at least three of the first four crouching trials. Although not preregistered, this criterion ensured that participants were initially willing to follow the evidence, establishing a comparable starting point from which to assess subsequent differentiation between informants. The same criterion was applied in a previous study using this paradigm with younger children and animals (Blakey et al., 2025b); adults completing the same task required no exclusions, as all met the criterion (Blakey et al., 2025a).

Language task

The experimenter administered the TELD-3 according to the test instructions. Each subtest began at an age-based entry point, and children were required to answer three consecutive questions correctly to establish a basal point. If this criterion was not met, testing proceeded backwards until it was achieved. After the basal point was established, children continued answering questions until they made three consecutive errors. Breaks were offered between the subtests or whenever deemed necessary. Raw language scores were calculated separately for the receptive (maximum 37) and expressive (maximum 39) subtests and used separately in analyses rather than the preregistered composite score, which sums age-standardised quotients across subtests (see Table S2).

Data analysis

The analyses for this study were preregistered on the OSF (https://osf.io/vbku3 & https://osf.io/9mk36). Deviations from the preregistrations are identified in the text and reported in supplementary Table S2.

All analyses were performed using R, version 4.3.2 (R Core Team, 2024), using generalised linear mixed models (GLMMs) with binomial distribution and logit link function (*lme4* package; Bates et al., 2015). For a description of the model fitting process see the supplementary materials. We additionally performed a hierarchical cluster analysis based on principal components (HCPC; *FactoMineR* package: Husson et al., 2010; Lê et al., 2008; *factoextra* package: Kassambara & Mundt, 2020).

For Demonstration trial analyses we fitted binomial GLMMs with the binary dependent variable of whether children followed the informants' evidence. Timed out trials were excluded because not responding within the time limit is ambiguous—it does not clearly indicate a choice to "not follow". This is particularly relevant given the age of the participants (3–6 years), for whom a five second response window may sometimes have been too short regardless of their intended choice. We fitted six such models, two each for the effect of age, receptive language, and expressive language, with either the full data set or only first trial data. Age and language were examined in separate models to avoid issues of multicollinearity. The models included fixed effects of 'informant', 'action number' (Crouching –1, Lifting 0, Sound 1), 'age in days', the interactions between these three variables, and a control fixed effect of 'action side' (side indicated by the informant). In models examining association with language skills, age was replaced with receptive or expressive language scores, due to strong positive correlations between age and these scores (Receptive: $r(80) = .68$, $p < .001$, Expressive: $r(80) = .75$, $p < .001$). 'Age in days' and the receptive and expressive language scores were scaled and centred before including them into the model. To account for repeated observations, 'participant ID' and 'colour-dyad' (pairing of informants' colour assignment and reliability) were included as random intercepts, each with random slopes for all fixed effects, keeping random effects structures 'maximal' (Barr et al., 2013). The fixed effects 'informant' and 'action side' were dummy-coded and centred before including them as random slopes.

For Transfer trial analyses, we fitted three GLMMs (one each for 'age in days', 'receptive language', and 'expressive language') with 'preferred informant' as the dependent variable (Reliable, Unreliable). Due to a high proportion of Timed out responses in the first Transfer trial, data for this trial was removed altogether. We included the fixed effects of age (or language score), 'transfer task number' (2 to 6), the interaction between these variables, and a control fixed effect of 'Reliable informant side' (left, right). Note that 'transfer task number' replaced the two preregistered terms of 'transfer task' and 'task trial number' (see Table S2). All fixed effects except for the control predictor were scaled and centred before including them in the model. The random effects structure was set up similarly to the previous GLMMs, with 'participant ID' and 'colour-dyad' as random intercepts and all fixed effects as random slopes. The fixed effect 'Reliable informant side' was dummy-coded and centred before including it as a random slope. Note that these analyses differ from those reported in the original submission which used CLMMs rather than the preregistered GLMMs; see Table S2 in the supplementary materials for a brief comparison of results.

In an exploratory analysis we used the Demonstration trial data for a hierarchical cluster analysis on principal components (HCPC) to identify groups of children based on children's evidence following behaviour. First, a principal component analysis (PCA) was applied using six variables describing the proportion of trials with evidence following per informant and action. Timed out trials were included in the total number of trials to extract a more conservative measure of the proportion of trials in which children followed each informant. Since all variables were proportions (0 to 1), data was not scaled before running the PCA. Clusters were identified using the *HCPC* function in *FactoMineR* applying 'Euclidean' distance and Ward's criterion. The optimal number of clusters was determined by the *HCPC* function.

Finally, to examine whether older children, who were more likely to discriminate between the informants, responded similarly to adults on the same task, we conducted an additional analysis comparing children with adults from Blakey et al. (2025a); this analysis is reported in the supplementary materials.

Results

Age effects

Demonstration trials

Full data set. The binomial GLMM for evidence following in the Demonstration trials was significantly better than the null model ($\chi^2(7) = 44.18$, $p < .001$). There was no evidence of a three-way interaction between informant, action number, and age ($\chi^2(1) = .143$, $p = .706$). Examination of all two-way interactions showed no significant interactions involving action number ($p \geq .174$), so these interactions were removed. The final model showed a significant interaction between informant and age ($\chi^2(1) = 8.89$, $p = .003$), indicating that with age children were significantly more likely to follow the evidence of the Reliable informant compared to the Unreliable informant (Fig. 3A). Children began to significantly differentiate between informants at approximately 4 years and 8 months of age (4.71 years). There was also a significant main effect of action number ($\chi^2(1) = 18.34$, $p < .001$) indicating lower evidence following in later actions (Fig. 3B), irrespective of informant role (see also supplementary Fig. S1) and children's age.

First trials. As expected, children followed both informants at similarly high rates in the first Crouching trials, with only three children not following the evidence (two in Reliable trials, one in an Unreliable trial). The binomial GLMM for evidence following in the first trials with each informant in each action in the Demonstration trials was significantly better than its null equivalent ($\chi^2(7) = 33.98$, $p < .001$). There was no evidence of a three-way interaction between informant, action number, and age ($\chi^2(1) = .267$, $p = .606$). A reduced model looking at all two-way interactions also revealed no significant interactions ($p \geq .060$). The final model without interaction terms revealed a significant effect of action number ($\chi^2(1) = 20.30$, $p < .001$) indicating lower evidence following in later actions, and a significant effect of informant ($\chi^2(1) = 4.74$, $p = .030$) indicating lower evidence following with the Reliable informant compared to the Unreliable informant. There was no significant effect of age ($\chi^2(1) = 3.21$, $p = .073$).

Transfer trials

The GLMM constructed to examine the effect of the children's age on their probability to choose the Reliable informant in the Transfer trials (excluding trial 1 due to a high number of timed out trials, see Fig. 4) did not reveal a significantly better fit than the equivalent null model ($\chi^2(3) = 4.66$, $p = .199$). An intercept model without fixed effects revealed no significant difference from chance level in choosing the Reliable informant ($z = .93$, $p = .354$). Across all Transfer trials, children chose the Reliable informant in 52.2% of trials and the Unreliable informant in 47.8% of trials.

Language effects

Demonstration trials

Full data set. The two binomial GLMMs for evidence following in the Demonstration trials examining the effect of receptive and expressive language will be described together as the patterns of results are the same. Both models were significantly better than the null model (Receptive: $\chi^2(7) = 21.61$, $p < .001$; Expressive: $\chi^2(7) = 23.98$, $p = .001$). There was no evidence of a three-way interaction between informant, action number, and language score (Receptive: $p = .726$; Expressive: $p = .844$). Two-way interactions involving action number were removed after a reduced model showed no significant interactions (Receptive: $p \geq .301$; Expressive: $p \geq .086$). The

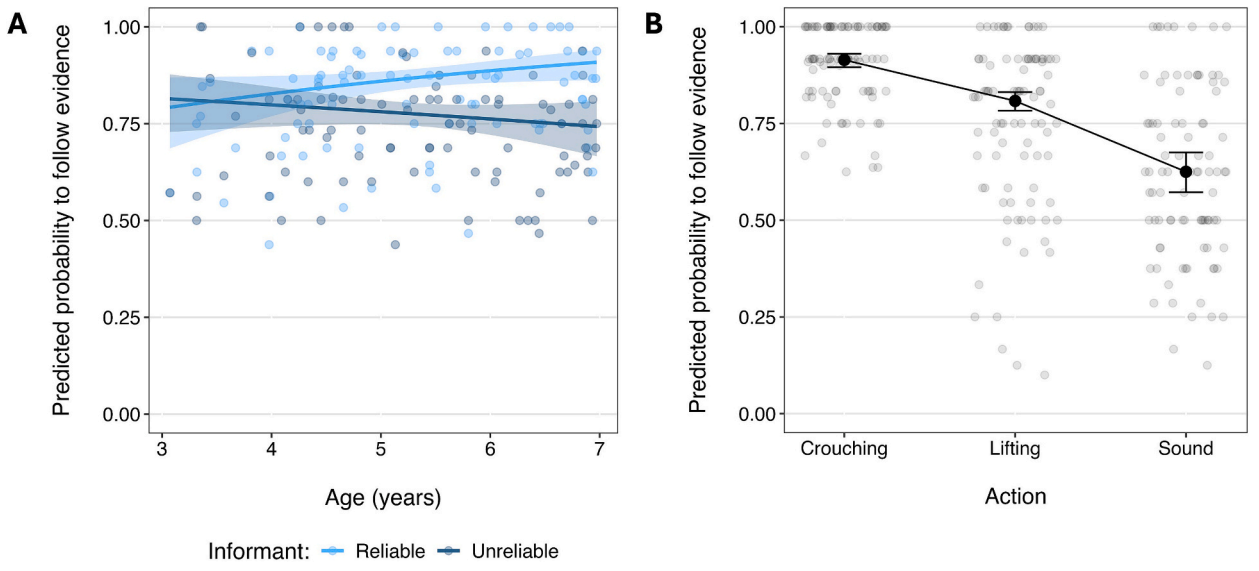

Fig. 3. Predicted probabilities to follow the evidence of informants. **A.** Across age (results averaged over action and action side). **B.** Across actions (results averaged over age, informant, and action side). Confidence intervals are given at the 95% level.

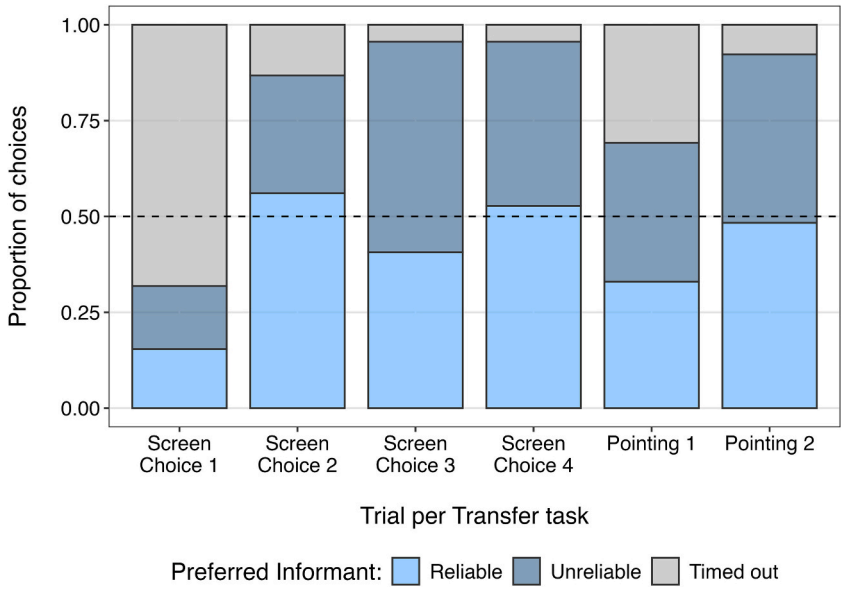

Fig. 4. Proportion of subjects choosing the respective informant or timing out (NA) in each Transfer trial (Screen choice 1–4, Pointing 1–2).

final models showed a significant interaction between informant and language score (Receptive: $\chi^2(1) = 6.51$, $p = .011$; Expressive: $\chi^2(1) = 4.27$, $p = .039$; see supplementary Fig. S2), indicating that as their language score increased children were significantly more likely to follow the evidence of the Reliable informant compared to the Unreliable informant. Children began to significantly differentiate between the informants when their receptive and expressive language scores reached 27. These raw scores correspond to an age equivalent of 5 years and 4 months for receptive language and 4 years 1 month for expressive language (see supplementary materials for an explanation of this discrepancy). In both models, evidence following significantly reduced in later actions (Receptive: $\chi^2(1) = 14.00$, $p < .001$; Expressive: $\chi^2(1) = 13.75$, $p < .001$).

First trials. The binomial GLMMs for evidence following in the first trials in the Demonstration trials were both significantly better fits than the null model (Receptive: $\chi^2(7) = 37.27, p < .001$; Expressive: $\chi^2(7) = 35.24, p < .001$). There was no evidence of a three-way interaction between informant, action number, and language score (Receptive: $p = .144$; Expressive: $p = .136$), nor any two-way interactions in a reduced model (Receptive: $p \geq .065$; Expressive: $p \geq .241$). The final model revealed a significant main effect of language score (Receptive: $\chi^2(1) = 5.23$, $p = .022$; Expressive: $\chi^2(1) = 5.08$, $p = .024$) indicating that children with better language scores were more likely to follow the informants' evidence. There was also a significant main effect of action number (Receptive: $\chi^2(1) = 19.86, p < .001$; Expressive: $\chi^2(1) = 20.22, p < .001$) indicating lower evidence following in later actions. In the expressive language model, informant produced a significant main effect (Expressive: $\chi^2(1) = 3.87$, $p = .049$), while in the receptive language model, the

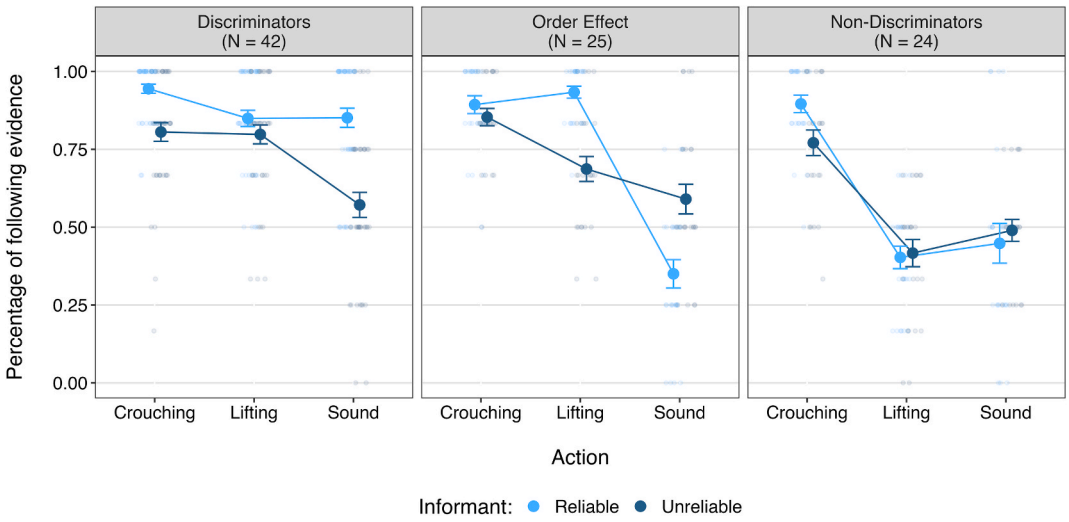

Fig. 5. Response patterns during the Demonstration trials for each cluster. Mean percentages (±SE) are provided per informant and action.

effect was marginally significant (Receptive: $\chi^2(1) = 3.80$, $p = .051$). In both cases, children were less likely to follow evidence from the Reliable informant compared to the Unreliable informant.

Transfer trials

The GLMMs constructed to examine the effect of the children's language scores on their probability to choose the Reliable informant in the Transfer trials did not reveal a significantly better fit compared to their respective null models (Receptive: $\chi^2(3) = 3.35$, $p = .341$; Expressive: $\chi^2(3) = 3.29$, $p = .350$).

Categorising response patterns

To explore whether distinct patterns of evidence following behaviour could indicate that some children were responding to an undermining defeater, we examined response patterns in the Demonstration trials. This allowed us to see whether preferences in the Transfer trials were explained by their evidence following behaviour in the Demonstration trials.

In a previous study using the same paradigm (Blakey et al., 2025a), adults had shown two distinct patterns: (1) *Discriminators*, who differentiated between the informants, and (2) *Followers*, who followed both informants' evidence. For children, we used a hierarchical cluster analysis on principal components (HCPC) to identify response patterns. The HCPC revealed three distinct clusters of response patterns: 1) *Discriminators* ($n = 42$): children who discriminated between informants, particularly in the sound trials; 2) *Order Effect* ($n = 25$): children who appeared to discriminate between informants in the lift trials but were influenced by the change in informant presentation order in the sound trials; and 3) *Non-Discriminators* ($n = 24$): children who did not discriminate between informants, and rather showed an overall decrease in following behaviour after the crouching trials; see Fig. 5.

Age and language scores across clusters

To examine whether the clusters differed in age or language abilities (see Table 1 for means and standard deviations), we conducted linear models followed by *emmeans* post hoc comparisons. This analysis revealed marginal evidence for age differences across clusters, $F(2, 88) = 3.04$, $p = .053$. Post hoc comparisons suggested that children in the Non-Discriminator cluster tended to be younger than those in the Order Effect cluster, $t(88) = 2.31$, $p = .06$, but no other pairwise differences were significant ($p \geq .112$), providing only weak evidence for age differences between clusters. No significant differences were found between clusters for receptive language, $F(2, 79) = 2.20$, $p = .117$, or expressive language, $F(2, 79) = 1.09$, $p = .342$.

Transfer trials

After obtaining clusters, a binomial GLMM ('logit' link) was fit for 'preferred informant' (Unreliable 0, Reliable 1) in the Transfer trials, with 'cluster group' (3 levels) as the predictor of interest, in an interaction with 'trial number' (2 to 5, z-transformed; the first screen choice trial was excluded due to a high proportion of Timed out trials). 'Reliable informant side' (left, right) was included as a control variable. The random effects structure included random intercepts of 'participant ID' and 'colour-dyad', with all fixed effects (dummy-coded and centred) as random slopes.

The constructed GLMM did not reveal a significantly better fit compared to the null model ($\chi^2(5) = 8.02$, $p = .155$), indicating that the response patterns in the Demonstration trials did not influence the probability of choosing the Reliable informant across the Transfer trials.

Discussion

Using a non-verbal task, this study investigated whether children's ability to acquire and respond to undermining defeaters such as < the source is unreliable > is related to age and developing language skills. Specifically, we examined whether 3- to 6-year-olds could identify and assess evidence coming from an unreliable source as misleading (i.e., acquire an undermining defeater from a series of overriding defeaters), and infer that subsequent evidence provided by the same source may also be misleading (i.e., respond to the undermining defeater). Evidence of this ability in a non-verbal task would challenge the hypothesis that language is required for identifying and assessing reasons. In the framework outlined in the introduction, successfully doing so would constitute a non-linguistic form of reflective responsiveness to reasons. We expected this to manifest as reduced reliance on evidence from the unreliable source, while continuing to follow equivalent evidence from a reliable source after a change in context. A similar pattern on the first trial with each source in each context was expected to indicate that children had genuinely inferred informant reliability rather than relearning it within each action. Additionally, we expected that children who had acquired an undermining defeater in Demonstration trials would prefer the Reliable over the Unreliable informant when given a choice between the two in the Transfer trials. Previous research using the same non-verbal paradigm found that adults (Blakey et al., 2025a)—but not 2-year-old children or

Table 1
Mean (SD) for age and language scores by cluster.

Cluster	Age in months	Receptive language score	Expressive language score
Discriminators	64.4 (12.2)	28.1 (5.8)	29.9 (6.7)
Order Effect	66.3 (13.3)	29.3 (4.0)	30.0 (4.9)
Non-Discriminators	57.7 (13.8)	26.0 (5.2)	27.7 (5.8)

animals (Blakey et al., 2025b)—show this type of discrimination.

Regarding the effect of age across the Demonstration trials, we found that with increasing age, children were more likely to discriminate between the Reliable and Unreliable informants. Specifically, starting from around 4 years 8 months, children followed the evidence of the Reliable informant more often than that of the Unreliable informant—mirroring the pattern previously observed in adults (Blakey et al., 2025a). By contrast, 3-year-olds and younger 4-year-olds were less likely to discriminate between the informants, aligning more closely with the behaviour of 2-year-olds and animals (Blakey et al., 2025b). This suggests that the ability to discriminate between informants may undergo a meaningful development between ages four and five. However, in the Transfer trials, there were no age differences in children's preference for the Reliable informant when given a choice between the two.

Despite the discrimination between informants in the Demonstration trials, older children did not show the predicted *selective* reduction in following the evidence. That is, they did not reduce their reliance specifically on the Unreliable informant's evidence while continuing to follow the Reliable informant. Instead, both older *and* younger children's tendency to follow the evidence declined across actions for *both* informants—a pattern again consistent with findings from adults (albeit with a small yet significant reduction; Blakey et al., 2025a) and 2-year-olds and animals (Blakey et al., 2025b).

The observed increase in informant discrimination with age may indicate that *older* children were more likely to have acquired an undermining defeater such as < *this* source is unreliable>, based on exposure to misleading evidence (i.e., overriding defeaters) from the Unreliable informant. This interpretation aligns with adults' responses in the same paradigm (Blakey et al., 2025a). Importantly, it cannot be attributed to any pre-existing preference for the Reliable informant, as almost all children followed the evidence of both informants in the first crouching trial. If an undermining defeater was acquired, as few as two or three misleading trials appear to have been sufficient for the Unreliable informant to be judged as such, given that the discrimination between informants did not increase in later actions. This is consistent with adult data, and also with findings from Kidd et al. (2013), where a single instance of unreliability influenced children's behaviour in a delayed gratification task.

This account, however, does not fully explain the observed decline in evidence following across actions (regardless of informant identity). One possibility is that while the decline was independent of age, the underlying reasons for this decline may have differed between older and younger children. For example, older children, despite acquiring an appropriate undermining defeater related to the Unreliable informant, may have become less engaged with the task over time, whereas younger children may have grown uncertain about whether to rely on the evidence at all. In older children, the decline in evidence following may also reflect reduced motivation or a shift in strategy as the task progressed—mirroring a similar though more modest pattern observed in adults (Blakey et al., 2025a). A direct comparison between older children and adults confirmed that both groups showed a reduction in evidence following across actions, though the reduction was significantly greater in children (see supplementary materials). This suggests that their decline may not reflect a developmental limitation but a broader tendency to disengage from evidence across repeated trials. For younger children, the growing uncertainty about whether to rely on the evidence at all, particularly after encountering more inconsistent evidence (Blakey et al., 2025b), may have stemmed from difficulty interpreting the reliability of the informants: while the Reliable informant consistently indicated the correct location, the Unreliable informant did so only half the time. For some children repeated exposure to inconsistent cues may not have led to a clear inference of unreliability but instead fostered confusion about the task more generally. As a result, they may have approached the evidence with increasing caution or disengaged altogether, leading to the observed reduction in evidence following across actions.

Another possibility, not mutually exclusive with the previous account, is that some children (especially younger children) may have processed an undermining defeater like < the source is unreliable > but applied it indiscriminately, generalising mistrust to both informants. Third, some children may have responded in the expected way while others based their responses on more superficial features of the task such as the order in which informants were presented. That is, children's responses may reflect a mixture of reflective and unreflective strategies, with only some responding to an undermining defeater.

Given the wide range of explanations regarding the variability of children's responses, we conducted an exploratory cluster analysis to explore whether some children may have acquired and responded to an undermining defeater in the expected way. Specifically, we investigated whether children's patterns of evidence-following in the Demonstration trials could be grouped into distinct behavioural profiles that, in turn, predicted children's informant preferences in the Transfer trials. A hierarchical cluster analysis revealed three distinct clusters: 1) Discriminators, who selectively followed the Reliable informant's evidence more than the Unreliable informant's evidence, particularly in the sound trials; 2) Order Effect children, who showed discrimination in the lift trials but appeared to have been influenced by the change in the informant presentation in the sound trials—possibly attributing unreliability to the last informant to provide evidence rather than to the individual informant, and thus continuing to follow the second informant's evidence even when that person was now the Unreliable informant; and 3) Non-Discriminators, who did not distinguish between informants in the Demonstration trials and showed a general reduction in evidence following after the crouching trials.

We examined whether the three clusters differed by age and found only marginal evidence of an age difference across clusters. Although the Non-Discriminators tended to be younger than the Order Effect group, the substantial overlap in ages across all three clusters indicates that age alone does not fully account for children's ability to discriminate between informants. Notably, some younger children were nonetheless classed as Discriminators, suggesting that this capacity emerges variably during early childhood rather than following a particular developmental trajectory.

Contrary to expectations, we did not find evidence of a difference in Reliable informant preferences across Transfer trials between clusters. This suggests that the subset of children—Discriminators—who may have been the most likely to have acquired and responded to an undermining defeater concerning informant reliability were not able to apply it consistently across contexts. This would require a basic form of reflective thinking—demanding not only the identification and assessment of evidence coming from a particular source as misleading, but also generalisation and application to other contexts (i.e., the same source may be misleading in

another context). Given the complexity of this capacity, it is unlikely to be evenly distributed across early childhood, nor to emerge consistently even in children who appear to discriminate based on informant reliability.

Although this paradigm was designed to be non-verbal, we were interested in the relationship between children's receptive and expressive language skills and their ability to acquire and respond to undermining defeaters. Children with stronger receptive and expressive language skills were more likely to discriminate between the informants in the Demonstration trials, following the evidence of the Reliable informant more often compared to the Unreliable informant. This mirrors the age-related pattern and supports the same interpretations. As with age, higher language scores were also associated with a decline in following the evidence of both informants across actions, suggesting a similar combination of disengagement and uncertainty. Strong correlations between language ability and task success could provide initial support for the claim that language is needed to identify and assess reasons, though correlation does not imply causation, and may nevertheless indicate the presence of some shared mechanism linking language and reflection that remains to be discovered. By contrast, the absence of such a correlation would challenge the hypothesis that language is necessary for identifying and assessing reasons. The current findings are more consistent with the former possibility, suggesting that linguistic skills may support the ability to acquire and respond to undermining defeaters, even in a non-verbal task. Specifically, children with higher language scores appeared to have acquired an undermining defeater related to the Unreliable informant like < the source is unreliable >. Doing so led them to follow the evidence of the Unreliable informant less often than the Reliable informant. The absence of a significant relationship between language scores and preferences in the Transfer trials suggests that while language may support sensitivity to informant reliability in the Demonstration trials, it does not appear to extend to the ability to generalise that judgment to a new context. Interestingly, language scores also did not differ between the response-based clusters, further suggesting that while language supports some aspects of sensitivity to informant reliability, it does not fully account for individual differences in response patterns across the task.

It is important to note that we cannot fully tease apart the effects of age and language, as these measures were highly correlated. We need to distinguish a *facilitating* role of language, whereby linguistic skills accelerate or enhance the ability to respond to undermining defeaters, from a *determinant* role, whereby linguistic skills make responses to undermining defeaters possible in the first place. Our findings are consistent with language playing a facilitating role, though they do not rule out the possibility that undermining defeaters are accessible without it. Undermining defeaters may still be accessible at a non-linguistic level through repeated exposure to over-riding defeaters. Notably, however, whichever role language plays, at least some children in this study appeared to acquire such defeaters before the so-called "age of reason". This suggests an early-emerging capacity for reflective evaluation of sources, albeit one that may be context-bound rather than fully generalised. It is perhaps unsurprising that older children and those with more advanced language skills may be better equipped to acquire and respond to undermining defeaters. Language is known to enhance cognitive capacities broadly, including those involved in reflective thinking. For instance, there is a strong link between children's language development and their theory of mind abilities (Astington & Jenkins, 1999), suggesting that language supports the emergence of reflective understanding of others' mental states.

We additionally examined children's behaviour on the first trial with each informant in each new context. These first trials were expected to be especially informative: if children had inferred that an informant was unreliable, they might treat new evidence coming from the same source with suspicion without needing to re-evaluate that evidence in the new context. The results of the first trials were unexpected and inconsistent with the full trial analysis. Children were actually less likely to follow the evidence provided by the Reliable than the Unreliable informant, with no age or language related differences. This cannot be attributed to a pre-existing preference for the Unreliable informant, as only three children chose not to follow an informant in the initial crouching trials. This pattern is difficult to reconcile with the interpretation that children had acquired and responded to an undermining defeater, as it suggests they may not have generalised the defeater across contexts. However, there are a few reasons why this test may have been too stringent. First, there were only 16% of trials in which children did not follow the evidence, suggesting a strong tendency to follow that may have obscured discrimination on individual first trials. Even if children had acquired an appropriate undermining defeater, they might still have followed the Unreliable informant occasionally. Across multiple trials, this would yield chance-level responding (around 50%), but for a binary measure on the first trials, such variability could obscure underlying patterns. Moreover, lapses in attention or uncertainty could lead even children who had identified the Reliable informant to fail to follow them consistently, especially on a single trial. These sources of noise may reduce the likelihood of observing a clear contrast between informants in such a small number of trials. Thus, while this result challenges some of our earlier interpretations, we caution against placing too much weight on this comparison as it only draws on two sets of two trials, making it a difficult test to pass.

Similar to our results, Schleihauf et al. (2022) reported that children from 4 years could reflectively revise their beliefs in response to undermining defeaters acquired via testimony. In the current study we aimed to assess this ability without relying on verbal testimony, instead using a non-verbal task to explore whether children with less developed language skills also have the capacity for basic reflective belief revision. However, as noted, while linguistic ability may be sufficient to support the acquisition of undermining defeaters, the more pressing theoretical question is whether it is necessary. These findings add to a growing body of work suggesting that the capacity for reflective belief revision emerges earlier in childhood than traditionally assumed by some philosophers. Even in a non-verbal task, some children appeared to acquire and respond to undermining defeaters related to informant reliability. These results challenge the view that the 'age of reason' marks the onset of rationality, showing that some children exhibit reflective and rational thinking earlier—and prompting a reconsideration of this developmental timeline.

CRediT authorship contribution statement

Kirsten H. Blakey: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration,

Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chloe L. Dow:** Writing – review & editing, Writing – original draft, Investigation. **Ariane Veit:** Writing – review & editing, Formal analysis. **Brina Recelj:** Methodology, Investigation. **Zsófia Virányi:** Writing – review & editing, Methodology, Conceptualization. **Giacomo Melis:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Eva Rafetseder:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful all those who participated in this study. We especially thank the University of Stirling students who assisted with data collection. This work was supported by a UKRI Future Leaders Fellowship (grant # MR/T042249/1) awarded to Giacomo Melis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2026.106547

Data availability

The data and analysis code have been made available in an OSF repository [https://doi.org/10.17605/OSF.IO/NBTVQ].

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 1–43. https://doi.org/10.1016/j.jml.2012.11.001

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

Blakey, K. H., Melis, G., Virányi, Z., & Rafetseder, E. (2024). *The effect of counterevidence strength on adults' reflective belief revision [Materials Folder]. Strong_feedback_condition.py* [Experiment Code]. Open Science Framework. Doi: 10.17605/OSF.IO/XHE8T.

Blakey, K. H., Melis, G., Virányi, Z., & Rafetseder, E. (2025a). Adults show selective responses to unreliability based on the strength of counterevidence. *PLoS One 1, 20*(11), Article e0331480. https://doi.org/10.1371/journal.pone.0331480

Blakey, K. H., Rafetseder, E., Atkinson, M., Renner, E., Cowan-Forsythe, F., Sati, S. J., & Caldwell, C. A. (2021). Development of strategic social information seeking: Implications for cumulative culture. *PLoS One 1, 16*(8). https://doi.org/10.1371/journal.pone.0256605

Blakey, K. H., Rafetseder, E., Melis, G., Veit, A., Amelung, K., Freudensprung, F., Kovacs, K., & Virányi, Z. (2025b). Nonverbal rationality? 2-year-old children, dogs, and pigs show unselective responses to unreliability but to different degrees. *Child Development, 96*(6). https://doi.org/10.1111/cdev.70020

Boyle, M. (2018). A different kind of mind? In *The Routledge Handbook of Philosophy of Animal Minds*.

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2008). Rational tool use and tool choice in human infants and great apes. *Child Development, 79*(3), 609–626. https://doi.org/10.1111/j.1467-8624.2008.01146.x

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science, 11*(5), 761–770. https://doi.org/10.1111/j.1467-7687.2008.00726.x

Conee, E. B., & Feldman, R. (2004). *Evidentialism: Essays in epistemology* (R. Feldman, Ed.). Oxford University Press.

Danón, L., & Kalpokas, D. E. (2023). Doxastic revision in non-human animals: The first-order model. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-023-00693-x

Dretske, F. I. (2006). Minimal rationality. In S. Hurley, & M. Nudds (Eds.), *Rational Animals?* (pp. 107–116). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528272.003.0003.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), Article 6873. https://doi.org/10.1038/415755a

Glock, H.-J. (2019). Agency, intelligence and reasons in animals. *Philosophy, 94*(4), 645–671. https://doi.org/10.1017/S0031819119000275

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-%2520011710

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences, 20*(3), 204–213. https://doi.org/10.1016/j.tics.2015.12.007

Hresko, W., Reid, D., & Hammill, D. (1999). *Test of early language development: 3rd ed. (TELD-3)*.

Hurley, S. (2006). Making Sense of Animals. In S. Hurley & M. Nudds (Eds), *Rational Animals?* Oxford University Press.

Husson, F., Josse, J., & Pages, J. (2010). *Principal component methods—Hierarchical clustering—Partitional clustering: Why would we need to choose for visualizing data?*.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (Version 1.0.7) [Computer software]. https://cran.r-project.org/web/packages/factoextra/index.html.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*(1), 109–114. https://doi.org/10.1016/j.cognition.2012.08.004

Kimura, K., & Gopnik, A. (2019). Rational higher-order belief revision in young children. *Child Development, 90*(1), 91–97. https://doi.org/10.1111/cdev.13143

Király, I., Oláh, K., Csibra, G., & Kovács, Á. M. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences, 115*(45), 11477–11482. https://doi.org/10.1073/pnas.1803505115

Király, I., Oláh, K., & Kovács, Á. M. (2023). Can 18-month-olds revise attributed beliefs? *Open Mind, 7*, 435–444. Doi: 10.1162/opmi_a_00087.

Kornblith, H. (2012). *On Reflection*. Oxford University Press.

Korsgaard, C. M. (2018). *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software, 25*, 1–18. https://doi.org/10.18637/jss.v025.i01

Marcus, E. (2021). *Belief, Inference, and the Self-Conscious Mind*. Oxford University Press.

McDowell, J. (1994). *Mind and World*. Harvard University Press.

Melis, G., & Blakey, K. H. (2025). Epistemic rationality begins unreflectively. *Erkenntnis, 91*, 1989–2012. https://doi.org/10.1007/s10670-025-00977-x

Melis, G., & Monsó, S. (2023). Are humans the only rational animals? *The Philosophical Quarterly*. https://doi.org/10.1093/pq/pqad090

Mercier, H., & Sperber, D. (2018). The Enigma of Reason. In *The Enigma of Reason*. Harvard University Press. Doi: 10.4159/9780674977860.

O'Madagain, C., Helming, K. A., Schmidt, M. F. H., Shupe, E., Call, J., & Tomasello, M. (2022). Great apes and human children rationally monitor their decisions. *Proceedings of the Royal Society B: Biological Sciences, 289*(1971). https://doi.org/10.1098/rspb.2021.2686

Peirce, J. W., Gray, J. R., Simpson, S., Macaskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perner, J. (2012). MiniMeta: In search of minimal criteria for metacognition. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds), *Foundations of Metacognition* (pp. 94–116). Oxford University Press. Doi: 10/m4c9.

Pollock, J. L. (1974). *Knowledge and Justification*. Princeton University Press.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development, 24*(1), 61–69. https://doi.org/10.1016/j.cogdev.2008.07.004

Schleihauf, H., Herrmann, E., Fischer, J., & Engelmann, J. M. (2022). How children revise their beliefs in light of reasons. *Child Development, 93*(4), 1072–1089. https://doi.org/10.1111/cdev.13758

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science, 25*(9), 1730–1738. https://doi.org/10.1177/0956797614540178

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.