Children's developing ability to adjust their beliefs reasonably in light of disagreement

Antonia F. Langenhoff | Jan M. Engelmann | Mahesh Srinivasan

Department of Psychology, University of California, Berkeley, Berkeley, California, USA

Correspondence
Antonia F. Langenhoff, Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704, USA.
Email: afl@berkeley.edu

Abstract
Two preregistered experiments (N = 218) investigated children's developing ability to respond reasonably to disagreement. U.S. children aged 4–9, and adults (50% female, mostly white) formed an initial belief, and were confronted with the belief of a disagreeing other, whose evidence was weaker, stronger than, or equal to participants' evidence. With age, participants were increasingly likely to maintain their initial belief when their own evidence was stronger, adopt the other's belief when their evidence was weaker, and suspend judgment when both had equally strong evidence. Interestingly, 4- to 6-year-olds only suspended judgment reliably when this was assessed via the search for additional information (Experiment 2). Together, our experiments suggest that the ability to respond reasonably to disagreement develops over the preschool years.

From Gram Sabhas, one of the world's largest deliberative institutions in India, to the Kgotla, a traditional community meeting in rural Botswana, people around the world reason together to make collective decisions (Sen, 2003). Public deliberation can be a powerful tool to overcome polarization, counteract populism, and make better judgments—under the right conditions. On the individual level, an important requirement for public deliberation is that people respond to disagreement in reasonable and intellectually humble ways. Sound public decisions can only be reached if people consider points of view that differ from their own, and respond rationally when confronted with differing perspectives (Habermas, 1997).

But what constitutes a rational response to disagreement? The answer depends on what the disagreement is about. When disagreeing about preferences or values (for example, about the best restaurant in town), the ideal thing to do might be to find a compromise or to “agree to disagree” (see Amemiya et al., 2021; Fisher et al., 2016; Foushee & Srinivasan, 2017; Sen, 2003). But people also disagree about issues that have a single, objectively correct answer (such as the number of restaurants in town). In these situations, a reasonable individual should consider which of the two diverging beliefs is supported by stronger evidence. If the individual has good reason to think that the evidence supporting their belief is stronger than that of the person they are disagreeing with, they should maintain their initial belief; if the other person's evidence is stronger, they should adopt the other person's view. Sometimes, however, the evidence supporting the individual's own belief versus the disagreeing other person's belief will be equally strong. By some philosophical accounts, an individual who finds themselves in this kind of disagreement should suspend judgment, and acquire additional evidence before coming to a conclusion (Frances, 2014; Friedman, 2017). Such a willingness to suspend judgment is often considered to be an important intellectual virtue; not only in Philosophy (Frances, 2014; Friedman, 2013, 2017; Turri, 2012), but also in Education (Haney, 1964).

Here, we investigated children's developing ability to respond to disagreement in the above-described three distinct ways across two studies. Research illuminating the developmental origins of responses to disagreement has both theoretical and practical implications. On a
theoretical level, some have argued that the human ability to reason is best conceptualized not as an individual skill, but as a fundamentally social skill, thus defining reasoning as the social practice of giving and asking for reasons in interpersonal discourse (Heyes, 2012; Köymen & Tomasello, 2020; Mercier & Sperber, 2011; O'Madagain, 2019; O'Madagain & Tomasello, 2019; Schleihauf et al., 2022; Tomasello, 2020). By these accounts, the ability to adjust one’s belief appropriately in light of disagreement constitutes a critical hallmark of a rational reasoner. On a practical level, such research can inform interventions aimed at fostering argumentation skills and intellectual humility (see Danovitch et al., 2019; Hagá & Olson, 2017a; Porter & Schumann, 2018). Fostering these skills, in turn, can contribute to facilitating a healthy public discourse.

Much prior work in Developmental Psychology has looked at children's developing ability to understand and evaluate disagreements among third parties (e.g., Amemiya et al., 2021; Birch et al., 2008; Chen et al., 2013; Corriveau et al., 2009; Foushee & Srinivasan, 2017; Heiphetz & Young, 2017; Koenig, 2012; Mercier et al., 2018; Scofield & Behrend, 2008). One central finding from this literature is that, when presented with a disagreement between other agents, even preschoolers are sensitive toward which of two opposing beliefs is supported by stronger evidence (quantitatively or qualitatively), and preferentially adopt that belief. For example, 3- to 4-year-olds preferentially accept a claim that is supported by a unanimous majority over one that is supported by a lone dissenter (e.g., Chen et al., 2013; Corriveau et al., 2009). Around age 4, children also prefer to adopt the beliefs of those who support their claims with strong reasons over the beliefs of those who provide only weak reasons (e.g., Koenig, 2012; Mercier et al., 2018). Thus, when evaluating disagreements between third parties, children respond reasonably from early on.

However, for understanding the development of children's intellectual humility and their ability to become rational contributors to public discourse, it is particularly important how children themselves respond to disagreement. This is what we study here. Prior studies investigating children's responses to first-person disagreements have focused on whether children maintain their initial belief or adopt the conflicting testimonial evidence from another person (Hagá & Olson, 2017b; Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010; Miosga et al., 2020; Schleihauf et al., 2022; Young et al., 2012). Interestingly, these studies show that 2- to 3-year-olds sometimes adopt the belief of a disagreeing other, even if that belief conflicts with an event the child has just witnessed (Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010). Thus, young children appear to often use disagreement as an indicator that they are mistaken, and employ a “when somebody disagrees with you, adopt their view”—strategy (see also Hagá & Olson, 2017b).

Slightly older children, in contrast, appear to consider both the quality of the evidence for their initial belief, as well as the quality of the evidence for the belief of a disagreeing other when deciding whether or not to change their minds after a disagreement—at least in situations where it is easy to see which of the two opposing belief is supported by stronger evidence (Miosga et al., 2020; Young et al., 2012; Schleihauf et al., 2022; see also Bridgers et al., 2016). In one recent study (Schleihauf et al., 2022), 4- to 5-year-old children in one condition were provided with strong evidence regarding which of two boxes contained a reward (e.g., children learned that one box was heavier). These children hardly ever adopted the opposing belief of a disagreeing other when this belief was supported by a bad reason (e.g., “I think the reward is in the other box because it is my favorite color”); but they did so sometimes when the contradictory belief was supported by a good reason (“I think the reward is in the other box, because I saw it in there”). In another condition, when children's initial belief was based on a guess, they almost always adopted the belief of the disagreeing other, regardless of whether the other's belief was supported by a good or a bad reason. Finally, research with even older children suggests that, by age 7—sometimes referred to as the “age of reason” (see Tomasello, 2020)—children reliably compare the strength of the evidence supporting their own claim versus the claim of a disagreeing other, even in more complex contexts. For example, 7-year-olds selectively maintain their initial belief or adopt another person's belief, depending on which belief is supported by first- versus second-hand evidence (Köymen & Engelmann, 2022; Köymen & Tomasello, 2018; see also Morgan et al., 2015).

Taken together, prior research suggests that there appear to be considerable improvements with age in children's ability to respond to first-person disagreements. However, this work leaves two key questions unanswered. First, prior studies have exclusively focused on children’s ability to selectively maintain their initial belief or adopt the belief of a disagreeing other (e.g., Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010; Schleihauf et al., 2022). Yet, as explained above, an additional hallmark of rational reasoning is to suspend judgment when the evidence supporting one’s own belief and a disagreeing other’s belief are equally strong (Frances, 2014; Friedman, 2017). The first goal of our studies was thus to investigate whether, when disagreeing with another person, children would (1) maintain their initial belief when the evidence supporting their own belief was stronger, (2) adopt the other's belief if their own evidence was weaker, and (3) suspend judgment when the evidence supporting the two conflicting beliefs was equally strong. Second, prior work focusing on children's responses to first-person disagreement has rarely compared children of different age groups within the same paradigm. Thus, it is unclear whether the observed age-related increases in reasonable
responses to disagreement reflect genuine developmental change. The second goal of our studies was thus to investigate potential developmental differences in children’s responses to disagreement. Specifically, we compared how children of two different age groups responded to disagreement: 4- to 6-year-olds and 7- to 9-year-olds (we did not test children below age 4 because piloting with 3-year-olds had revealed that our task was too difficult for these children).

To answer the two questions described above, we presented children and adults (for comparison) with a story in which a character's pet had run away, and asked participants to find out where the pet was hiding. Participants first acquired evidence by asking informants, and were then asked to state their belief about where the pet went. Next, participants learned that another agent—who had also consulted informants on the whereabouts of the pet—disagreed with the participant. Finally, we assessed participants' belief for a second time. We were interested in whether participants would maintain their initial belief, suspend judgment by saying they were “not sure” and “needed more information,” or adopt the other agent's belief. Across conditions, we varied whether the evidence supporting the participants' belief was stronger than, weaker than, or equal to the evidence supporting the belief of the disagreeing agent.

Our main indicator of reasonableness was whether participants would maintain their initial belief in the stronger evidence condition; adopt the other agent's belief in the weaker evidence condition; and suspend judgment in the equal evidence condition more often than in the respective other two conditions. We chose to compare participants’ responses in this way because we assumed that there would be baseline differences in how attractive the different response options would be, and that these baseline differences might vary between age groups. Since our main focus was on whether children’s and adults’ responses vary reasonably as a function of the different epistemic conditions despite these underlying baseline differences, our preregistered main analyses compared participants’ responses across conditions rather than within conditions.

Based on prior work (e.g., Köymen & Tomasello, 2018), we expected that by age 7, children would respond to disagreement in these ways, similarly to adults. In contrast, we expected that younger children would diverge from older children and adults in two ways. First, although previous research shows that children begin to compare the quality of the evidence supporting their own belief versus the belief of a disagreeing other when deciding whether to change their minds after a disagreement (Miosga et al., 2020; Schleihaufl et al., 2022; Young et al., 2012), prior work also suggests that they are generally more open to the suggestions of others than older children and adults (see Bridgers et al., 2016; Hagá & Olson, 2017b). We thus expected that, across conditions, 4- to 6-year-olds would be more likely to adopt the belief of the person they were disagreeing with. Second, prior research suggests that the ability to suspend judgment might be difficult for young children, particularly if it requires them to make a metacognitive judgment about their own uncertainty (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013; Perner, 2012). In Experiment 1, suspension of judgment was assessed with an explicit metacognitive judgment; participants suspended judgment by indicating that they were “not sure” and needed “more information.” Accordingly, we predicted that 4- to 6-year-olds in Experiment 1 would be less likely to suspend judgment than the other age groups. Experiment 2 explored whether a more implicit operationalization of suspension of judgment would make it easier for 4- to 6-year-olds to respond rationally to the ambiguous evidence presented to them in the equal evidence condition (Hembacher et al., 2020; Lapidow et al., 2022; Lyons & Ghetti, 2012). Hypotheses, exclusion criteria, analyses, and design for both experiments were preregistered (Experiment 1: https://aspredicted.org/cw9wa.pdf; Experiment 2: https://aspredicted.org/cw9wa.pdf).

**EXPERIMENT 1**

**Methods**

**Participants**

A total of 114 participants completed Experiment 1: 38 younger children aged 4.0–6.11 (16 4-year-olds, eight 5-year-olds, 14 6-year-olds, M<sub>age</sub> = 5.46 years, SD<sub>age</sub> = 0.97 years, 50% female), 38 older children aged 7.0–9.11 (17 7-year-olds, 20 8-year-olds, one 9-year-old, M<sub>age</sub> = 8.10, SD<sub>age</sub> = 0.55, 50% female) and 38 adults (M<sub>age</sub> = 31.45, SD<sub>age</sub> = 13, 50% male, 45% female, 5% agender or non-binary). Children were recruited from a database of families who had signed up to participate in child development studies. Children were predominantly white (45%), followed by Asian (8%), Latinx and white (8%), and Latinx (7%). Sixteen percent of families did not indicate their child's race or ethnicity. Upon study completion, children received a certificate. Two additional children participated but were excluded in accordance with our exclusion criteria because they stopped responding after the second trial. Data for children were collected between June and August 2020.

Adults were recruited via Prolific. Participation was restricted to participants based in the United States with a prior approval rate higher than 90%. Adults were mostly white (66%), followed by African or African American (13%), Asian (11%), Latinx (8%), or white and Latinx (2%). Adults were compensated with $1.60; their data were collected in December 2020.

Our total sample size was determined based on an a priori power analysis. We calculated an estimate of the
sample size required to find a small to medium effect ($f = 0.2$) of the interaction between our predictors (condition and age group) on participants’ responses to the second belief assessment with 90% power. This analysis resulted in an optimal sample size of $N = 111$.

**Design**

The experiment had three within-subjects conditions—corresponding to three different test trials—which differed in the relative strength of evidence that participants were exposed to before encountering a disagreement with another agent. In the *stronger evidence condition*, participants received testimony from three informants who told participants where the pet went. The other agent received testimony from one informant. In the *weaker evidence condition*, participants received testimony from one informant and the other agent received testimony from three informants. In the *equal evidence condition*, both the participant and the other agent received testimony from two informants. The order in which participants were presented with the conditions was randomized.

**Procedure**

Before the experiment, children were asked if they wanted to participate, and only participated if both they and their parents agreed. Adults provided written consent. Testing took place online; experiments were implemented in Qualtrics. Test sessions with children were conducted by an experimenter via Zoom; adults participated self-directed, via Prolific. An overview of the experimental procedure is depicted in **Figure 1**.

**Exposure to evidence**

During the experiment, participants either listened to an experimenter reading (children) or themselves read (adults) a picture book about Emma and her bunny. Participants took part in one practice trial and three test trials. Each trial began as follows: Emma’s bunny ran away and Emma was sad. One of Emma’s friends offered to find the bunny and encouraged the participant to help them, suggesting that the friend and the participant could each ask some people if they saw where the bunny went. There were always two options of where the bunny could have gone, for example toward the house or toward the bridge (presentation of stimuli was counterbalanced). Next, participants were exposed to one, two, or three informants (number varied based on condition), who, one after another, told the participant where the bunny had gone. Informants always provided a consistent response, for example, all informants said the bunny went toward the house. In order to avoid carryover effects between trials, participants “interacted” with each character only once. That is, participants were exposed to a different friend of Emma’s and to different informants in each trial.

**Initial belief assessment**

After being exposed to the statements of the informants, the participant was prompted to state their belief about where the bunny went for a first time (e.g., in

**1. Exposure to Evidence**

“The Bunny Went To the House!”

“Correct”

“Incorrect”

**2. Initial Belief Assessment**

“Correct”

“Incorrect”

**3. Disagreement**

“I don’t Think the Bunny Went to the Bridge. I Asked One Person and the Person said the Bunny Went to the Bridge. So I Think the Bunny Went to the Bridge.”

**4. Second Belief Assessment**

Maintaining Initial Belief

SUSPENDING Judgment

ADOPTING Other’s Belief

**FIGURE 1** Overview of the experimental procedure for Experiment 1. The condition depicted here is the *stronger evidence condition.*
the weaker evidence condition, “So one person told you the bunny went to the house. What do you think? Do you think the bunny went to the house or do you think the bunny went to the bridge?”). Children received feedback on this initial belief assessment. If a child responded differently than what the informant(s) had told them, the experimenter repeated the test question. If a child responded incorrectly for a second time, the experimenter corrected the child (e.g., “remember, one person told you the bunny went toward the house. So I think the bunny went toward the house.”). If a participant responded incorrectly to the first belief assessment (after feedback for children), the rest of the trial was skipped and the participant was directed to the subsequent trial. As specified in our preregistration, these trials were excluded from analyses. Moreover, if a participant responded to the initial belief incorrectly on more than one test trial, their data were excluded completely.

Disagreement and second belief assessment
The first trial was a practice trial that familiarized participants with the structure of the task. In the practice trial, the belief of Emma's friend confirmed the participant's initial belief. In contrast, in the three subsequent test trials, Emma's friend disagreed with the participant, and justified their belief by referring to how many informants they asked. For example, in the weaker evidence condition, Emma's friend said “I don't think the bunny went toward the house. I asked three people and they said the bunny went toward the bridge. So I think the bunny went toward the bridge.”

After this disagreement occurred, the participant's belief about where the bunny went was assessed for a second time. For example, in the weaker evidence condition, the experimenter said “one person told you the bunny went toward the house. And you said the bunny went toward the house. And three people told [friend's name] the bunny went toward the bridge. And [friend's name] said the bunny went toward the bridge. What do you think now? Where do you think the bunny went?”. The response options in this case were “toward the house”, which represented maintaining one's initial belief, “toward the bridge”, which represented adopting the friend's belief, and “I am not sure. I need more information.”, which represented suspension of judgment. The response options were illustrated with pictures, for example, with a picture of the house and a picture of the bridge. The “suspension of judgment” response option was illustrated with an icon of a person shrugging their shoulders in an “I don't know” gesture.

Data analysis
We used logistic linear mixed regression models (lme4 package in R, Bates et al., 2012) with a binomial distribution (logit link) to analyze participants' responses to the second belief assessment. In our primary preregistered analyses, the main predictor was the interaction between condition (stronger vs. weaker vs. equal evidence condition) and age group (adults vs. older children vs. younger children). In addition, in response to a comment from a reviewer, we also investigated the effects of age as a continuous predictor on children's responses. In these models, the main predictor was the interaction between condition and age (in years and months). If possible, all models included gender (female vs. male vs. nonbinary for adults) and trial number (1–3, z-transformed, to avoid convergence issues) as control predictors, random slopes for condition (mean-centered) and trial number (z-transformed), as well as a random intercept for subject. If a model did not converge, we removed correlations between random effects, then removed control predictors (first gender and then trial number), and finally removed random slopes (first trial number and then condition).

For our primary preregistered analyses, as well as for the analyses including age as a continuous predictor, we ran three different types of regression models with binary outcome measures: in the first type of model, we predicted participants' propensity to maintain their initial belief. For these models, responses were recoded as “maintaining one's initial belief” versus “other” (which in this case included suspension of judgment and adopting the friend's belief). In the second type of model, we predicted participants' propensity to adopt the friend's belief (recoding responses to “adopting the friend's belief” vs. “other”), and in the third type of model, we predicted participants' propensity to suspend judgment (recoding responses to “suspension of judgment” vs. “other”). For our primary preregistered analyses, as well as for our analyses including age as a continuous predictor, we included all data from children.

To establish the significance of the full models (Schielethz & Forstmeier, 2009), we compared the deviance of the full models with those of the null models containing only the control predictors, the intercept, and the random slopes, using likelihood-ratio tests (Dobson & Barnett, 2018). To test the significance of the interactions between condition and age, we compared the deviance of the full models with those of corresponding reduced models not containing the interactions, using likelihood-ratio tests. We used the R package emmeans (Lenth et al., 2018) to conduct post hoc analyses (Tukey method), when necessary. All materials for Experiments 1 and 2, including data, analysis scripts, and printouts of the Qualtrics surveys, are available here: https://osf.io/2tbx/?view_only=de6229ef7a2e437a1b9fc9b3e540b9c.
Results

We organize this section according to our three preregistered predictions about how a reasonable person should respond to disagreement. We analyze whether (1) participants maintained their initial belief more often in the stronger evidence condition, (2) adopted the belief of the disagreeing other more often in the weaker evidence condition, and (3) suspended judgment more often in the equal evidence condition than in the respective other two conditions. As specified in our preregistered exclusion criteria, our analyses were based on trials in which participants responded correctly to the initial belief assessment (after feedback for children). In total, participants failed to respond correctly to the initial belief assessment on only five trials (adults: three trials in the weaker evidence condition; children: one trial in the stronger evidence condition, one trial in the weaker evidence condition). An overview of how participants responded in the different conditions in Experiment 1 and Experiment 2 is provided in Table 1.

Did participants maintain their initial belief when their own evidence was stronger?

For analyses involving age group as a categorical variable, we could not fit a model to our whole dataset because there was not enough variance in the adult data: in the stronger evidence condition, adults maintained their initial belief 84% of the time, but they never did so in the other two conditions (see Table 1). A model including the data from children was highly significant when compared with the null model (\(\chi^2[7] = 83.05, p < .001\)). Specifically, we found a significant interaction between condition and age group (\(\chi^2(2) = 8.02, p = .018\)), while gender (\(\chi^2(1) = 1.03, p = .310\)) and trial number (\(\chi^2(1) = 0.84, p = .360\)) had no effect. To further investigate the interaction between condition and age group, we looked at the effects of condition on younger and older children in separate regression models. In both models, condition was a significant predictor (older children \(\chi^2(2) = 49.43, p < .001\); younger children \(\chi^2(2) = 25.99, p < .001\)). Post hoc pairwise comparisons revealed that both age groups were more likely to maintain their initial belief in the stronger evidence condition than in the equal evidence condition (older children \(p = .009\); younger children \(p < .001\)), and in the stronger evidence condition compared to the weaker evidence condition (older children \(p < .001\); younger children \(p = .004\)). Thus, as groups, both younger (i.e., 4- to 6-year-olds) and older (i.e., 7- to 9-year-olds) children responded “reasonably” in the sense that they maintained their initial belief more often when their own evidence was stronger than that of Emma’s friend, compared to when their own evidence was weaker or equally strong. Note that for the above-mentioned, as well as all additional analyses reported throughout the paper, we found the same trends when including data from participants first trials only.

When analyzing the data from children including age as a continuous predictor, we had to take out the control predictors and the random slope for the model to converge. The full model was significant compared to the null model, (\(\chi^2(5) = 35.57, p < .001\)), revealing a significant interaction between condition and age (\(\chi^2(2) = 18.26, p < .001\)). As Figure 2 shows, the condition difference

<table>
<thead>
<tr>
<th>Age group</th>
<th>Response</th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stronger evidence</td>
<td>Weaker evidence</td>
<td>Equal evidence</td>
<td>Stronger evidence</td>
<td>Weaker evidence</td>
<td>Equal evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(f)</td>
<td>SE</td>
<td>(f)</td>
<td>SE</td>
<td>(f)</td>
<td>SE</td>
<td>(f)</td>
<td>SE</td>
<td>(f)</td>
<td>SE</td>
</tr>
<tr>
<td>4–6 years</td>
<td>Maintain</td>
<td>.43</td>
<td>.08</td>
<td>.08</td>
<td>.04</td>
<td>.29</td>
<td>.07</td>
<td>.4</td>
<td>.07</td>
<td>.22</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Adopt</td>
<td>.41</td>
<td>.08</td>
<td>.79</td>
<td>.07</td>
<td>.50</td>
<td>.08</td>
<td>.36</td>
<td>.07</td>
<td>.59</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Suspend</td>
<td>.16</td>
<td>.06</td>
<td>.13</td>
<td>.05</td>
<td>.21</td>
<td>.07</td>
<td>.25</td>
<td>.06</td>
<td>.2</td>
<td>.06</td>
</tr>
<tr>
<td>7–9 years</td>
<td>Maintain</td>
<td>.76</td>
<td>.07</td>
<td>.03</td>
<td>.02</td>
<td>.29</td>
<td>.07</td>
<td>.4</td>
<td>.07</td>
<td>.29</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Adopt</td>
<td>.13</td>
<td>.05</td>
<td>.76</td>
<td>.07</td>
<td>.18</td>
<td>.06</td>
<td>.4</td>
<td>.06</td>
<td>.18</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Suspend</td>
<td>.11</td>
<td>.05</td>
<td>.22</td>
<td>.07</td>
<td>.53</td>
<td>.08</td>
<td>.79</td>
<td>.07</td>
<td>.79</td>
<td>.07</td>
</tr>
<tr>
<td>Adults</td>
<td>Maintain</td>
<td>.84</td>
<td>.06</td>
<td>0</td>
<td>0</td>
<td>.08</td>
<td>.04</td>
<td>.82</td>
<td>.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Adopt</td>
<td>.05</td>
<td>.04</td>
<td>.83</td>
<td>.06</td>
<td>.13</td>
<td>.05</td>
<td>.04</td>
<td>.03</td>
<td>.84</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Suspend</td>
<td>.11</td>
<td>.05</td>
<td>.17</td>
<td>.06</td>
<td>.79</td>
<td>.07</td>
<td>.14</td>
<td>.05</td>
<td>.16</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note: \(f\) indicates relative frequencies, SE indicates the 95% standard error.
CHILDREN RESPOND REASONABLY TO DISAGREEMENT

first emerged in 5-year-olds, and was increasingly clear-cut in older children. It appears that 4-year-olds did not maintain their belief more often in the stronger evidence condition than in the other two conditions.

**Did participants adopt the other's belief when their own evidence was weaker?**

When analyzing the data including age group as a categorical predictor, the full model was significant when compared with the null model ($\chi^2(8) = 140.35, p < .001$). Specifically, the age group x condition interaction was significant ($\chi^2(4) = 12.41, p = .015$). Gender ($\chi^2(3) = 6.03, p = .110$) and trial number ($\chi^2(1) = 1.48, p = .224$) did not have a significant effect on participants' responses. We followed up on this interaction with separate regression models, and found that condition was a significant predictor in all three age groups (adults $\chi^2(2) = 100.72, p < .001$; older children $\chi^2(2) = 78.12, p < .001$; younger children $\chi^2(2) = 26.96, p < .001$). Specifically, all age groups adopted the friend's belief significantly more often in the weaker evidence condition than in the stronger evidence condition (adults $p < .001$; older children $p < .001$; younger children $p < .001$), and in the weaker evidence condition compared to the equal evidence condition (adults $p < .001$; older children $p = .032$; younger children $p = .003$). Thus, as predicted, participants adopted the belief of Emma's friend more often when their evidence was weaker than that of Emma's friend, compared to when their own evidence was stronger or equally strong. Indeed, within the weaker evidence condition, all three age groups were more likely to adopt the other character's belief than to maintain their initial belief or to suspend judgment.

When analyzing children's responses including age as a continuous predictor, we had to take out the control predictor for the model to converge. Compared to the null model, the full model was significant ($\chi^2(5) = 74.06, p < .001$). Specifically, there was a significant interaction between condition and age ($\chi^2(2) = 10.27, p = .006$). As Figure 2 shows, children of all ages adopted the other's belief more often in the stronger evidence condition than in the other conditions, but became slightly more likely to do so with increasing age. In addition, older children

**FIGURE 2** Probabilities with which children in Experiment 1 maintained their initial belief, adopted the other's belief, and suspended judgment in the different conditions.
also became less likely to adopt the other's belief in the other two conditions.

Did participants suspend judgment when their evidence was as good as the evidence of the disagreeing other?

When analyzing participants' responses including age group as a categorical predictor, the full model was highly significant when compared with the null model ($\chi^2(8) = 75.05, p < .001$, note that we had to take out the control predictor gender for the model to converge). Specifically, we found a significant interaction between condition and age group ($\chi^2(4) = 15.96, p = .003$), while trial number did not have a significant effect ($\chi^2(1) = 3.53, p = .06$). Following up on this interaction, we found that in the models for adults and older children, condition was a significant predictor (adults $\chi^2(2) = 81.14, p < .001$; older children $\chi^2(2) = 76.81, p < .001$). Adults and older children were significantly more likely to suspend judgment in the equal evidence condition compared to the stronger evidence condition (adults $p = .001$, older children $p < .001$), and the weaker evidence condition (adults $p < .001$, older children $p = .033$). In contrast, condition was not a significant predictor in the model for younger children ($\chi^2(2) = 1.17, p = .557$). Thus, unlike adults and older children, 4- to 6-year-olds did not show the tendency to suspend judgment in the equal evidence condition; instead, their propensity to suspend judgment was low overall. Consistent with this, within the equal evidence condition, the predominant response of adults and older children was to suspend judgment, while 4- to 6-year-olds were most likely to adopt the other character's belief (see Table 1).

When analyzing children's responses including age as a continuous predictor, we had to take out the control predictor gender and the random slopes for condition. The full-null model comparison was significant ($\chi^2(5) = 27.65, p < .001$), revealing a significant interaction between condition and age ($\chi^2(2) = 11.34, p = .003$). Specifically, Figure 2 shows that 4- and 5-year-old's propensity to suspend judgment was similar across conditions. Around age 6, children began to suspend judgment more often in the equal evidence condition than in the other two conditions, and this tendency was more pronounced with increasing age.

Discussion

In Experiment 1, we investigated whether children and adults responded reasonably to disagreement. In line with our predictions, we found that 4- to 6-year-olds, 7- to 9-year-olds, and adults maintained their initial belief most often when their own evidence was stronger, and adopted the belief of the other agent most often when their own evidence was weaker. However, we also found interesting developmental differences.

A first central difference was that younger children showed an overall tendency to adopt the belief of the disagreeing other. As a group, 4- to 6-year-olds selectively adjusted or maintained their initial belief based on the relative strength of the evidence supporting their own belief versus the belief of the disagreeing other. However, this was mostly driven by the 6-year-olds, as some 5-, and many 4-year-olds adopted the other's belief in both conditions. Thus, a number of our youngest participants may have followed the strategy: “when somebody disagrees with you, adopt their belief.” This would be in line with previous research showing that although 4- and 5-year-old children maintain their initial belief in the face of conflicting testimonial evidence when their initial belief is clearly based on stronger evidence (Miosga et al., 2020; Schleihauf et al., 2022; Young et al., 2012), they often adopt the belief of a disagreeing other when it is not perfectly obvious which of the two opposing beliefs is correct (Bridgers et al., 2016; Hagá & Olson, 2017b).

The scenarios children encountered in our experiment arguably presented such ambiguity, since participants formed their initial belief based on a limited amount of testimonial evidence. Because of this relative ambiguity, some of our youngest participants might have taken disagreement as a general indicator that they were mistaken, leading them to adopt the belief of the disagreeing other without weighing the evidence supporting each of the opposing beliefs, as older children and adults did.

The second central developmental difference we found in Experiment 1 was that, unlike older children and adults, the group of 4- to 6-year-olds did not suspend judgment more often in the equal evidence condition than in the other two conditions. Across conditions, younger children's likelihood to suspend judgment was low, and it did not differ between conditions. Instead, within the equal evidence condition, 4- to 6-year-olds showed a tendency to adopt the other person's belief rather than to suspend judgment. But can we conclude from Experiment 1 that the ability to suspend judgment in situations of ambiguous evidence does not emerge until around age 7? No. There are at least three different reasons for why 4- to 6-year-olds may have refrained from suspending judgment in the equal evidence condition of Experiment 1.

A first possibility is that 4- to 6-year-olds may be altogether unable to suspend judgment. For example, the ability to suspend judgment might require inhibitory control, which develops only gradually over the preschool- and early school-years (Best & Miller, 2010). Children of this age might also struggle with identifying ambiguous evidence. They may not have realized that in the equal evidence condition of our experiment, the strength of their own evidence and that of the disagreeing other were identical. Four- to 6-year-olds might also be unable to suspend judgment because they fail to
identify the uncertainty that situations of ambiguous evidence give rise to, due to limited metacognitive abilities (see Kuhn, 2000).

A second possibility is that while 4- to 6-year-olds are generally able to suspend judgment, they may be unwilling to do so, because children of this age are generally reluctant to embrace ambiguity and prefer holding a belief that is based on ambiguous evidence to not holding a belief at all (see Hagá & Olson, 2017b).

Finally, a third possibility is that 4- to 6-year-olds are generally willing and able to suspend judgment when this is operationalized in a different way. In Experiment 1, suspending judgment required participants to make an explicit verbal judgment of their own uncertainty (“I am not sure, I need more information”). Prior work has shown that children of this age often struggle with explicit uncertainty judgments (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013). Thus, if suspension of judgment were assessed more implicitly, even young children might view this as the most reasonable response in situations of ambiguous evidence. Instead of jumping to the conclusion that the ability to suspend judgment in situations of ambiguous evidence does not emerge until around age 7, we suspended judgment on the issue, and explored it further in Experiment 2.

EXPERIMENT 2

The goal of Experiment 2 was to test one explanation for why 4- to 6-year-olds did not suspend their judgment in a situation of disagreement where the evidence supporting two conflicting beliefs was equally strong. As discussed in the previous section, one possibility is that in general, 4-to 6-year-olds are able and willing to suspend judgment if it is operationalized in a different way. In order to suspend judgment in Experiment 1, participants had to indicate that they were “not sure” and needed “more information,” but they never experienced the actual consequences of suspending judgment; the trial instead ended, and they did not receive any additional information. Previous work from the metacognition literature suggests that when asked to verbally report their own uncertainty, preschoolers struggle with assessing their own knowledge and often overestimate their own abilities (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013). For example, even when their objective chances of correctly answering a question are low, young children often claim that they are “very sure” that they know the answer—an interesting contrast to their general tendency to adopt others’ beliefs (see Hagá & Olson, 2017b). In contrast, studies employing implicit measures suggest that even preschoolers can monitor their uncertainty and draw appropriate consequences from it. For example, preschoolers are more likely to opt out of answering a question (Lyons & Ghetti, 2012) or search for additional information in situations of greater uncertainty (Hembacher et al., 2020).

Thus, even 4- to 6-year-olds might be able to suspend judgment in situations of ambiguous evidence when this is assessed in a more implicit manner. To investigate this possibility, we operationalized suspension of judgment in Experiment 2 as the search for additional information, thus characterizing it as “something one does in order to genuinely inquire” (Friedman, 2017). Specifically, participants were told that if they were not sure, they could ask more people where the bunny went. If participants chose this option, they were exposed to two additional informants, whose statements confirmed either the participant’s initial belief or the belief of the disagreeing other. We were interested in whether, under these circumstances, 4- to 6-year-olds would be more likely to suspend judgment in the equal evidence condition than in the other two conditions.

Methods

Participants

Experiment 2 had 104 participants: 54 children aged 4.0–6.11 (27 4-year-olds, 13 5-year-olds, 14 6-year-olds, $M_{\text{age}} = 5.2$ years, $SD_{\text{age}} = 1.02$ years, 50% female), and 50 adults as a comparison sample ($M_{\text{age}} = 30.54$, $SD_{\text{age}} = 12.10$, 42% male, 58% female, 4% of participants were non-binary and 2% gender non-conforming). Participants were recruited and compensated as in Experiment 1. Children were predominantly white (37%). 19% of parents chose not to indicate their child's race or ethnicity. Fifteen percent of children were Asian and white, followed by 7% Asian, 6% Latinx, 4% Latinx and white, 4% African or African American, 4% Asian and Pacific Islander. The remaining 4% were Latinx and African American, Latinx and Asian, or Latinx, Asian, and white. In line with our pre-registered exclusion criteria, six additional children were excluded because they responded to the initial belief assessment incorrectly in more than one test trial even after feedback ($N = 3$), did not complete the study ($N = 2$), or were older than 6.11 years ($N = 1$). One additional child was excluded due to parental interference. Data for children were collected between November 2020 and March 2021. Adults were mostly white (60%), followed by Asian (16%), African or African American (10%), Latinx (8%), Latinx and white (4%), and African or African American and American Indian or Alaskan Native (2%). Adult data were collected in December 2020.

A simulation-based power analysis (using the R package simr)—based on the effect of our main predictor, condition, on older children's probability to suspend judgment in Experiment 1—estimated that this effect had 61.10% power (1000 simulations). Based on this, we
calculated that a sample size of 50 participants per age group would give us 91.5% power for the effect of condition on our response variable. We tested four additional children because families were recruited simultaneously and testing appointments had already been scheduled when it became clear that we had reached our target sample size.

Design

The design of Experiment 2 was identical to that of Experiment 1, except that participants were able to acquire additional evidence after they decided to suspend judgment. This additional evidence confirmed the participant's initial belief in the stronger evidence condition and the belief of the disagreeing other in the weaker evidence condition. In the equal evidence condition, we counterbalanced between participants whether the additional information confirmed the participant's initial belief or the belief of the disagreeing other.

Procedure

The general procedure for Experiment 2 was similar to that of Experiment 1 but involved some specific changes. First, Experiment 2 contained an additional practice trial. In this practice trial, participants initially asked one informant where the bunny went, but the informant claimed that they had not seen the bunny and thus did not know where the bunny went. The goal of this practice trial was to expose participants to a situation in which suspending judgment was clearly the right thing to do, so that participants would experience what it was like to suspend judgment and ask for additional information before moving onto the test trials. When a participant chose the suspension of judgment response option in this or any of the subsequent test trials, they were exposed to two additional informants who told the participant where the bunny had gone. As in Experiment 1, the informants always provided a consistent response. In the equal evidence condition, we counterbalanced between participants whether the information provided by the two additional informants supported the participant's or the disagreeing other's initial belief.

The main change introduced in Experiment 2 was that the response option representing suspension of judgment was operationalized differently. Instead of being operationalized as “I am not sure. I need more information.”, as in Experiment 1, the suspension of judgment response option was now operationalized as “I am not sure. I would like to ask more people where the bunny went.” If participants had suspended judgment and acquired the additional information, their belief about where the bunny went was assessed for a third time. For example, when a participant had decided to suspend judgment following the disagreement in the stronger evidence condition, the experimenter said (children) /the participant read (adults) “five people told you the bunny went toward the house. And one person told [friend's name] the bunny went toward the bridge. What do you think now? where do you think the bunny went?” In this example, the response options at were “toward the house”, which represented sticking with one's initial belief, and “toward the bridge”, which represented adopting the friend's belief.

An additional minor change was that in Experiment 2, both children and adults received feedback after the initial belief assessment. This was to make the procedure more similar for both age groups (as adults in Experiment 1 did not receive such feedback). If a participant responded to the initial belief assessment incorrectly for the first time, the test question (“Where do you think the bunny went?”) was repeated. If a participant responded to the initial belief assessment incorrectly for a second time, the participant was prompted with, for example: “One person said the bunny went to the house. So, I think the bunny went to the house. What do you think?” If a participant responded to the initial belief assessment incorrectly for a third time in one of the practice trials, the experimenter said (for children) /the participant read (for adults) “Remember, one person said the bunny went to the house. So, I am going to choose the house.”, and the correct response was selected for the participant. In test trials, participants did not receive this third prompt. Instead, they were automatically directed to the subsequent test trial if they responded to the initial belief assessment incorrectly for a third time.

Data analysis

Models were linear mixed regression models (lme4 package in R, Bates et al., 2012) with binomial distributions (logit link). As in Experiment 1, our primary preregistered analyses included the interaction between condition (stronger vs. weaker vs. equal evidence condition) and age group (adults vs. younger children) as the main predictor. However, we could not fit any of the models to our full dataset (i.e., including data from both adults and children) because there was not enough variance in the adult data. We thus report adult data descriptively, and regression results for child data only. The main predictor in these models was condition; when possible, the models included gender (female vs. male) and trial number (1–3, z-transformed) as control predictors, and a random intercept for subject. As secondary preregistered analyses, we investigated the effect of the interaction...
between condition and age as a continuous predictor (in years and months) on children's responses, as in Experiment 1.

Results

As for Experiment 1, we organize this section according to our predictions about how a reasonable person should respond to disagreement. We analyze whether (1) participants maintained their initial belief more often in the stronger evidence condition, (2) adopted the belief of the disagreeing other more often in the weaker evidence condition, and (3) suspended judgment more often in the equal evidence condition than in the respective other two conditions. Again, our analyses are based on trials in which participants responded correctly to the initial belief assessment (after feedback). In total, participants failed to respond correctly to the initial belief assessment on five trials (adults: 1 trial in the weaker evidence condition; children: 1 trial in the stronger evidence condition, 3 trials in the weaker evidence condition).

Did participants maintain their initial belief when their own evidence was stronger?

In the stronger evidence condition, adults maintained their initial belief 82% of the time, while they did so only 10% of the time in the equal evidence condition and never in the weaker evidence condition. In the model including the effect of condition on children's responses, condition was a significant predictor ($\chi^2(2) = 8.65, p = .013$). There was also a significant effect of trial number ($\chi^2(1) = 7.46, p = .006$), such that 4- to 6-year-olds were more likely to maintain their initial belief in the third test trial compared to the first or second test trial. There was no significant effect of gender ($\chi^2(1) = 2.39, p = .122$). Thus, as a group, 4- to 6-year-olds were more likely to maintain their initial belief in the stronger evidence condition than in the other two conditions, as predicted. However, post hoc tests were not significant ($p = .094$ for the stronger vs. equal comparison; $p = .081$ for the stronger vs. weaker comparison). When comparing participants' responses within the stronger evidence condition, both adults and children were most likely to maintain their initial belief. However, 4- to 6-year-olds were almost as likely to adopt the other's belief (see Table 1).

For the model including the interaction between condition and age as a continuous predictor, we had to take out gender and the random slopes. Compared to the null model, the full model was significant ($\chi^2(5) = 20.68, p = .001$), revealing a reliable effect of the interaction between condition and age ($\chi^2(2) = 8.71, p = .013$). The tendency to maintain one's initial belief more often in the stronger evidence condition than in the other two conditions first emerged around age 5, and became increasingly pronounced with age (Figure 3).

Did participants adopt the other's belief when their own evidence was weaker?

Eighty-two percent of adults adopted the other's belief in the weaker evidence condition. In the equal evidence condition and the stronger evidence condition, only 4% and 1% of adults adopted the friend's belief, respectively. In the model predicting children's propensity to adopt the friend's belief, condition was a significant predictor ($\chi^2(2) = 11.15, p = .004$). There was no significant effect of the control predictor trial number ($\chi^2(1) = 0.37, p = .55$) or gender ($\chi^2(1) = 2.16, p = .14$). Specifically, as predicted, children adopted the friend's belief more often in the weaker evidence condition than in the stronger evidence condition, and in the weaker evidence condition than in the equal evidence condition. Again, post hoc tests did not reach significance ($p = .075$ for the weaker vs. stronger comparison; $p = .063$ for the stronger vs. equal comparison). When comparing data within the stronger evidence condition, adopting the other's belief was the predominant response in both age groups, although the difference between adopting the other's belief and the other two response options was more pronounced in adults (Table 1).

Did participants suspend judgment when their evidence was as good as the evidence of the disagreeing other?

Seventy-six percent of adults suspended judgment in the equal evidence condition. In contrast, only 14% did so in the stronger evidence condition, and 16% in the weaker evidence condition. Unlike in Experiment 1, condition was a significant predictor of 4- to 6-year-olds' responses ($\chi^2(2) = 13.94, p < .001$). There was no significant effect of trial number ($\chi^2(1) = 0.75, p = .387$) or gender ($\chi^2(1) = 0.77, p = .380$). Specifically, 4- to 6-year-olds were significantly more likely to suspend judgment in the equal evidence condition than they were in the stronger evidence condition ($p < .001$), and in the weaker evidence condition ($p = .003$). After suspending judgment and acquiring the additional information, the majority of participants...
from both age groups converged onto the belief that was supported by more evidence overall (see supplementary materials). Within the equal evidence condition, the suspension of judgment option was the modal response in both age groups (see Table 1). However, while suspension of judgment was the clear, majority response among adults (85%), it remained a minority response among 4- to 6-year-olds (41%).

For the model including age as a continuous predictor, we had to take out the control predictor gender and the random slopes. The model was significant when compared with the null model ($\chi^2(5) = 17.64, p = .003$). Specifically, there was a significant effect of the condition × age interaction ($\chi^2(2) = 7.36, p = .025$). Children were increasingly likely to suspend judgment in the equal evidence condition with age, and less likely to do so in the other two conditions (Figure 3).

**Figure 3** Probabilities with which children in Experiment 2 maintained their initial belief, adopted the other’s belief, and suspended judgment in the different conditions.

**Discussion**

In Experiment 2, we focused specifically on 4- to 6-year-olds' developing ability to suspend judgment in situations of ambiguous evidence—an ability which has been discussed as a key intellectual virtue in Philosophy (Frances, 2014; Friedman, 2017) and Education (Haney, 1964). Our results show that with increasing age, 4- to 6-year-olds are more willing and able to suspend judgment in situations of disagreement if it is assessed in an implicit manner. Specifically, rather than assessing suspension of judgment as an explicit, metacognitive judgment about participants' uncertainty (as in Experiment 1), suspension of judgment in Experiment 2 was operationalized via the search for additional information.

When comparing 4- to 6-year-olds' propensity to suspend judgment in Experiment 1 and Experiment 2 (see supplementary materials), we found that this procedural change led to a slight overall increase in 4- to 6-year-olds' propensity to suspend judgment (i.e., across conditions). That is, it seems that searching for information was overall more attractive for children compared to making explicit judgments. Although older children were not included in Experiment 2, we speculate that the different operationalization would have had a similar impact on older children, as well. Importantly, however, the increase in suspension of judgment responses among 4- to 6-year-olds was most dramatic in the equal evidence condition.
condition: as a group, 4- to 6-year-olds in Experiment 2 suspended judgment significantly more often in the equal evidence condition than in the other two conditions—this had not been the case in Experiment 1. Thus, our results suggest that assessing suspension of judgment more implicitly made it easier for 4- to 6-year-olds to respond rationally to the ambiguous evidence in the equal evidence condition. These findings are in line with previous work showing that young children show evidence of implicitly monitoring their uncertainty in their information seeking behavior (Coughlin et al., 2015; Hembacher et al., 2020; Lapidow et al., 2022; Perner, 2012).

While 4- to 6-year-olds in Experiment 2 were more likely to suspend judgment in the equal evidence condition than in the other two conditions, and more likely to do so than in Experiment 1, we note two caveats. First, the differences between 4- and 6-year-olds’ propensity to suspend judgment in the equal evidence condition compared to the other two conditions were far less pronounced than in adults. And second, within the equal evidence condition, only 41% of all 4- to 6-year-olds suspended judgment. Thus, the modal response for 4- to 6-year-olds within the equal evidence condition was suspension of judgment (41%). Still, the majority of 4- to 6-year-olds did not suspend judgment, and instead either maintained their initial judgment (24%) or adopted the friend’s belief (35%). What might explain this response pattern? One possibility is that even with the more implicit operationalization, many 4- to 6-year-olds may have failed to realize that there was equal evidence for the two competing claims—and/or that this meant that there was uncertainty about the bunny’s whereabouts—and thus that additional evidence was needed to resolve the disagreement. Alternatively, these children may have understood that there was uncertainty about the bunny’s location but did not opt to suspend judgment because they preferred holding a belief based on ambiguous evidence to not holding a belief at all. Future research could investigate more systematically which of these (or other) factors best explains 4- to 6-year-olds’ limited ability to suspend judgment.

In addition to showing that 4- to 6-year-olds were more likely to suspend judgment in the equal evidence condition than in the other two conditions, Experiment 2 also replicated some of the key findings of Experiment 1. Four- to 6-year-olds, as well as adults, responded reasonably to disagreement in that they were more likely to maintain their initial belief when their own evidence was stronger (compared to weaker or equal), and adopt the belief of the disagreeing other when their own evidence was weaker (compared to stronger or equal). However, unlike in Experiment 1, the post hoc tests comparing the frequency of maintaining one’s initial belief and adopting the friend’s belief across conditions were not significant in children. A look at Table 1 suggests that children in Experiment 2 were more drawn to the suspending judgment option across conditions (perhaps due to how we presented this option), which may have caused the contrasts between the response options to be less pronounced overall compared to in Experiment 1. Yet, taken together, the results of Experiment 2 suggest that as a group, even 4- to 6-year-olds demonstrate intellectual humility in responding reasonably to disagreement.

**GENERAL DISCUSSION**

When disagreeing with others on issues that have a single, objectively correct answer, a reasonable person should adjust their beliefs based on which of the two opposing beliefs is supported by stronger evidence, and suspend judgment if both beliefs are supported by equally strong evidence (Frances, 2014). We investigated the development of this ability across two experiments. In Experiment 1, participants of all age groups responded reasonably in maintaining their initial belief when their own evidence was stronger (stronger evidence condition), and adopting the other’s belief when their own evidence was weaker (weaker evidence condition). However, one central developmental difference was that 4- to 6-year-olds did not reliably suspend judgment when their evidence was as good as that of the disagreeing other (equal evidence condition). In contrast, in Experiment 2, where suspension of judgment was assessed more implicitly—via the search for additional information—4- to 6-year-olds were significantly more likely to suspend judgment in the equal evidence condition than in the other two conditions (although 4-year-olds did still not suspend judgment reliably). This finding is in line with other research showing evidence of implicit uncertainty monitoring in children’s information seeking behavior (Coughlin et al., 2015; Hembacher et al., 2020; Lapidow et al., 2022; see Perner, 2012, for discussion of whether implicit measures of uncertainty are valid measures of metacognition).

Together, our studies demonstrate the presence of social reasoning skills and intellectual humility from at least age 5 onwards. In showing that children begin to respond reasonably to disagreement from early on, our findings align with recent accounts in cognitive science that have emphasized the social dimensions of reasoning as a whole (Heyes, 2012; Köymen & Tomasello, 2020; O’Madagain & Tomasello, 2019; Tomasello, 2020), and intellectual humility, more specifically (Danovitch et al., 2019; Hagå & Olson, 2017a; Lynch et al., 2016; Porter & Schumann, 2018). For such accounts, the ability to respond to disagreement in these ways constitutes a key milestone on the way to becoming a rational reasoner. Such responses to disagreement require that a reasoner understand that others have perspectives that differ from their own, and integrate these perspectives with the objective state of the world—in our case by comparing the evidence supporting each of the two opposing beliefs (see Tomasello, 2020). In addition to illustrating the development of a crucial socio-cognitive
reasoning skill, our research also provides an important practical insight: it suggests that from early in life, children possess the cognitive prerequisites needed to develop into autonomous, open-minded citizens who can contribute to a meaningful public discourse. A meaningful public discourse, in turn, constitutes the backbone of any healthy democracy.

Although as groups, adults, older children, and younger children responded reasonably to disagreement in the current studies, the present findings also demonstrate a significant age-related “increase in reasonableness.” First, the analyses including age as a continuous predictor revealed that 4-year-olds did not actually discriminate between conditions. They were neither more likely to maintain their initial belief when their own evidence was stronger, nor to suspend judgment when their evidence was as strong as that of the disagreeing other. Instead, 4-year-olds almost always adopted the other’s belief. Second, in both experiments, children discriminated more strongly between conditions with increasing age, and adults discriminated most strongly. What might explain these age differences? One option is that with increasing age, children may have simply found it easier to pay attention to the experiment. Another, more interesting possibility is that the ability to respond reasonably to disagreement undergoes significant development throughout childhood. If the latter is the case, what factors are responsible for this development?

One factor that may have played a role in our studies may be children's developing ability to reliably compare how many informants support a specific claim (see Morgan et al., 2015). Importantly, experimentally manipulating how many informants support a claim is only one of many possible ways of manipulating the strength of evidence for a belief. In addition to manipulating the strength of evidence (i.e., the quantity of information), future research could also investigate how children respond to disagreements when opposing beliefs are supported by evidence of different quality, such as stronger or weaker reasons (see, e.g., Koenig, 2012; Köymen & Tomasello, 2018; Mercier et al., 2018).

A capacity that may specifically affect children's developing ability to suspend judgment is inhibitory control (see e.g., Wiebe et al., 2012). As their inhibitory control increases, children might increasingly avoid “jumping to conclusions,” and understand that it is rational to suspend judgment about a given belief until one has acquired sufficient evidence (see Hagà & Olson, 2017a, 2017b). We tested for a potential relation between participants’ inhibitory control skills (assessed via a go/no-go task, e.g., Wiebe et al., 2012), and their propensity to suspend judgment in our Experiment 2 (details on this analysis are reported in our supplementary materials). Although we did not find a relation between participants' inhibitory control skills here, we see this as an interesting avenue for future investigation.

To conclude our discussion of factors that might lead to age-related increases in reasonableness, we want to highlight that what is considered “reasonable” might itself change across development. What constitutes a reasonable response for adults does not necessarily constitute a reasonable response for children. Specifically, we observed across both experiments that younger children showed a general tendency to adopt a disagreeing other’s belief. While this is not in line with the here defined criteria for reasonableness, it is important to note that this strategy need not be considered “unreasonable.” Young children's knowledge about the world is still limited. They frequently experience that they themselves are wrong about things and that others are right, even in situations where the evidence supporting their belief may have seemed strong. For someone who knows little and is used to interacting with knowledgeable adults, being open toward adopting others' conflicting beliefs arguably constitutes an adaptive learning strategy.

Importantly, possessing the ability to respond reasonably to disagreement does not necessarily mean that people actually use that ability. In everyday life (and academic research), examples abound of individuals responding to disagreement in ways that one might perceive as “unreasonable” (at least from an epistemic perspective). People can be hesitant to give up their beliefs, even in light of strong evidence for an alternative view. Prior research with adults shows that this is often the case when an individual's belief carries personal importance, for example, because it is closely connected to their identity (e.g., Kahan, 2012; Lewandowsky & Oberauer, 2016). One explanation for this is that letting go of a belief that is closely connected to one's identity can carry immense social and emotional costs, such as being excluded from one's social group or losing one's sense of belonging. Against this background, it would not be surprising if children—like adults—would forgo pursuing the epistemic goal of “taking on the belief that is best supported by evidence” in favor of protecting their identity (see Kelly, 2003). It is up to future research to show this empirically.

More generally, future studies should focus on how children balance the relative costs and benefits of different responses to disagreement. With regard to suspension of judgment, one study found that, in contrast to older children and adults, young children who overheard speakers talking about ambiguous objects did not favor “intellectually humble” speakers (who acknowledged that they might be wrong), compared to diffluent speakers or intellectually arrogant speakers (Hagà & Olson, 2017a). In other contexts, even older children and adults might negatively evaluate individuals who suspend judgment, because these individuals could be perceived as lacking in self-confidence or as insecure. An interesting avenue for future research would thus be to identify the contextual factors that contribute to a potential shift from perceiving suspension of judgment as “humble” and desirable to perceiving it as hesitant and fickle. With regard to the relative costs and benefits of adopting other people's beliefs, it is well known that both children (Haun & Tomasello, 2011)
CHILDREN RESPOND REASONABLY TO DISAGREEMENT

and adults (e.g., Asch, 1956) are more likely to respond incorrectly to a simple perceptual question after a group of confederates has responded incorrectly. In these situations, participants’ responses do not necessarily represent their true beliefs—when asked individually, participants almost always make the correct perceptual judgment. Nevertheless, this research indicates that conformity- or authority-related concerns might lead people to adopt others’ beliefs even if these beliefs are not supported by stronger evidence. In some communities, respect for authorities plays a more important role than in others, perhaps because the community values integration and social responsibility to the group (Greenfield et al., 2000). This raises the possibility that, in some cultural contexts, maintaining one’s initial belief when disagreeing with older or other higher-status members of the community might simply not be perceived as appropriate, even if the evidence supporting one’s initial belief is stronger. Moreover, parenting styles might affect the extent to which children feel encouraged to speak up to defend their own beliefs as opposed to adopting their parents’ beliefs without questioning (e.g., Baumrind, 1971).

In conclusion, the current experiments present evidence that children from at least age 5 onwards respond reasonably when being confronted with the opposing belief of a disagreeing other. Thus, the ability to respond reasonably to disagreement—an important contributor to a healthy public discourse—is present from at least the preschool years. However, our results also show that individuals become more reasonable with age. Going forward, research should study how children respond to disagreement in a variety of social environments, since we know from research with adults that their responses to disagreement can be heavily influenced by a variety of social and contextual factors.

ORCID
Antonia F. Langenhoff https://orcid.org/0000-0002-9876-4951
Jan M. Engelmann https://orcid.org/0000-0001-5139-0289
Mahesh Srinivasan https://orcid.org/0000-0003-0350-2703

REFERENCES


**SUPPORTING INFORMATION**

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

**How to cite this article:** Langenhoff, A. F., Engelmann, J. M., & Srinivasan, M. (2022). Children's developing ability to adjust their beliefs reasonably in light of disagreement. *Child Development, 00*, 1–16. https://doi.org/10.1111/cdev.13838