Children’s meta-talk in their collaborative decision-making with peers

DOI:
10.1016/j.jecp.2017.09.018

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Journal of Experimental Child Psychology

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Children’s meta-talk in their collaborative decision-making with peers

Bahar Köymen¹ & Michael Tomasello²-³

¹University of Manchester
²Max Planck Institute for Evolutionary Anthropology
³Duke University

Article in press in *Journal of Experimental Child Psychology*

Acknowledgements:

Acknowledgements: We would like to thank Katharina Haberl, Isabelle de Gaillande-Mustoe, Ramona Frickel, Martina Deitrich, and Julia Ohl for their help in recruiting children and collecting data; Pia Fischer, Joanna Buryn-Weitzel, Silvan Holverstein, Alice Dabbagh, and Felix Engelmann for their help in coding; Philipp Brandstädter for the drawings; Susanne Mauritz with her help with the recordings, lovely Wencke Assmann for voicing the clips; Ronny Barr for his help in editing the clips; all daycare centers and children for their friendly cooperation.

Correspondence concerning this article should be addressed to Bahar Köymen, School of Health Sciences, University of Manchester, Oxford Road, M13 9PL, Manchester, United Kingdom. E-mail: bahar.koymen@manchester.ac.uk.
Abstract

In collaborative decision-making, children must evaluate the evidence behind their respective claims and the rationality of their respective proposals with their partners. In the main study, 5- and 7-year-old peer dyads ($N = 196$) were presented with a novel animal. In the key condition, children in a dyad individually received conflicting information about what the animal eats (rocks vs. sand) from sources that differ in reliability (e.g., with first-hand vs. indirect evidence). Dyads in both age groups were able to reliably settle on the option with the best supporting evidence. Moreover, in making their decision children, especially 7-year-olds, engaged in various kinds of meta-talk about the evidence and its validity. In a modified version of the key condition in Study 2, 3- and 5-year-olds ($N = 120$) interacted with a puppet who tried to convince children to change their minds by producing meta-talk. When the puppet insisted and produced meta-talk, 5-year-olds, but not 3-year-olds, were more likely to change their minds if their information was unreliable. These results suggest that even preschoolers can engage in collaborative reasoning successfully, but the ability to reflect on the process by stepping back to jointly examine the evidence emerges only in the early school years.

*Keywords:* Reasoning, collaborative decision-making, information reliability, peer interactions
Children’s meta-talk in their collaborative decision-making with peers

Reasoning – in the sense of explicating reasons for actions – is a fundamentally social skill enabling people to produce and evaluate arguments to reach joint decisions (Mercier & Sperber, 2011; Tomasello, 2014). Sperber and colleagues (2010) argue that in a world of mistrust, where everyone is practicing “epistemic vigilance”, individuals win others over to their point of view by providing good reasons, which will be accepted based not on trust but on logic and evidence. On the other hand, Tomasello (2014) stresses that reasoning is also critical in situations of trust: If the joint decision benefits both parties, then individuals want to make the best decision based on logic and evidence (regardless of who “wins” the argument), so they produce and evaluate reasons cooperatively as a means to that end.

Research on testimony suggests that young children are vigilant social learners and selectively trust certain information sources more than others. Around preschool ages, children prefer to learn new information from informants who accurately label familiar objects more than those who do not (Corriveau & Harris, 2009; Koenig & Harris, 2005); from informants who express their expertise/knowledge more than those who express uncertainty (Sobel & Corriveau, 2010; Sabbagh & Baldwin, 2001); and from informants who produce noncircular arguments more than those who produce circular arguments (Corriveau & Kurkul, 2014; Mercier, Bernard, & Clément, 2014).

Nevertheless, in most of these studies, children were asked to choose the more reliable informant out of the two, rather than evaluating the reliability of a single informant in the absence of a reference point. Moreover, all of these were comprehension studies in which children did not need to justify their decisions to convince a partner. Justifying a decision about information reliability for a partner is
an advanced meta-cognitive skill because speakers have to go beyond the content of the message. First, they need to refer to their information source in their justifications (where or from whom they received the information). Next, they need to express how reliable their information source is and express why they believe this piece of information is therefore reliable (Mahr & Csibra, 2017; Kuhn, 2001). Although preschoolers produce explanations when they have disagreements with others (Dunn & Munn, 1987; Kyratzis & Ervin-Tripp, 1999) and when they encounter information contradicting their prior beliefs (Legare, 2012; Young, Alibali, & Kalish, 2012), to date only much older children (around age 11-12) have been observed to use meta-argumentative strategies, or meta-talk, aimed at assessing the validity of arguments and/or evidence directly (e.g., “Tell us where you got that evidence”, Kuhn, Zillmer, Crowell, & Zavala, 2013, p.466). However, it is possible that young children have so far not been observed in the right situations, that is, not in situations of dispute but rather in situations of collaborative decision-making in which both participants are motivated to get the right answer.

In Study 1, therefore, we created an interactive context in which two peers would have to jointly solve a problem to reach correct joint decisions for a reward and investigated whether/how they produced meta-talk, explanations about the information source and the information reliability, for their decisions. We introduced 5- and 7-year-old peer dyads to a novel animal, called a “selk”, with three unique characteristics (e.g., eating rocks). Each child learned about the novel animal individually from an informant in a clip. In the critical condition, the unequal reliability condition, children within a dyad received conflicting information (e.g., one was told that selks eat only rocks; the other only sand) from sources that differ in reliability (one watched a first-hand report by a selk; the other a second-hand report
by a girl who expressed uncertainty). In the *equal reliability* condition, children received conflicting information from the same source (both children watched first-hand reports or both second-hand reports). In the *same information* condition, children received the same information from different sources (one watched a first-hand report; the other a second-hand report). Later, each dyad had to decorate the home for the novel animal and jointly decide on three items that it needs. We predicted that children would favor the items supported by the first-hand report in the *unequal reliability* condition and in the *same information* condition, whereas they would favor items randomly in the *equal reliability* condition. We also predicted that children would produce justifications and meta-talk more in the two conditions in which children had conflicting information than in the *same information* condition (see Legare, 2012). We explored if there were any age-related changes in children’s argumentation, since studies have shown that preschoolers robustly know whom to trust (Corriveau & Kurkul, 2014; Mercier et al., 2014), but they do not normally talk about information reliability until adolescence (Kuhn et al., 2013).

In Study 2, we explored whether preschoolers, 3-year-olds and 5-year-olds, would be able to choose the correct item, if a puppet partner provided the necessary meta-talk (e.g., asking the children who their source is, whether their source was sure or not) and directed their attention to the reliability of the information, as the literature suggest that preschoolers are able to gauge the information reliability depending on the informants’ accuracy and certainty (Baldwin & Sabbagh, 2001; Carriveau & Harris, 2009).

**Study 1**

**Method**

**Participants**
Ninety-eight 5-year-olds (M = 5;9, Range = 5;6–6;0, 48 girls) and 98 7-year-olds (M = 7;6, Range = 7;1–8;0, 48 girls) in same-age, same-sex dyads participated in the study. The dyads comprised of children who knew each other based their teachers’ recommendations. There were 16 5-year-old and 17 7-year-old dyads in the unequal reliability condition; 16 5-year-old and 16 7-year-old dyads in the equal reliability condition; 17 5-year-old and 16 7-year-old dyads in the same information condition. The children were native speakers of German with various socio-economic backgrounds.

Materials

In the warm-up, there was a 3-room dollhouse: A bathroom, a kitchen, a bedroom. There was a bin and a set of three drawers. Each drawer had two items: one that typically belonged to a room (a toilet, a fridge, a cradle) and the other that did not (a lamp, a ladder, a table). From each drawer, children had to place one item in a room, and the other in the bin.

In the teaching phase, each child watched one of the four clips (a slide show with 4 drawings) narrated by the same child. Two clips presented “strong evidence”, a first-hand report. In the first slide, the novel animal (the selk) introduced itself. In the next three slides, the selk presented their unique characteristics: what they eat, what they drink, and where they sleep (see Appendix A for the full narration in German and its translation in English). The two clips presented the following conflicting information:

- **Strong-evidence clip with Set 1**: A selk stated that they eat rocks, drink blue soda, and sleep on water.
- **Strong-evidence clip with Set 2**: A selk stated that they eat sand, drink green soda, and sleep on flowers.
The other two clips presented “weak evidence”, a second-hand report by a girl called Lisa who expressed uncertainty. In the first slide, Lisa introduced herself. In the next three slides, she presented what selks eat, what they drink, and where they sleep. In each slide, Lisa stated how unsure she was about her knowledge and how strange she finds this (see Appendix B for the full narration in German and its translation in English). The two clips presented the following conflicting information:

- **Weak-evidence clip with Set 1**: Lisa stated that somebody told her that selks eat rocks, drink blue soda, and sleep on water.
- **Weak-evidence clip with Set 2**: Lisa stated that somebody told her that selks eat sand, drink green soda, and sleep on flowers.

In the experimental phase, there was a 3D selk figure, a yellow cloth (selk’s home), three drawers, a box, and a bin. The box had a tree to be placed in selk’s home and a piano in the bin (by the experimenter to demonstrate the game). Children had to choose between rocks and sand in the top drawer; a blue and a green soda in the middle drawer; pictures of a lake and a meadow in the bottom drawer.

**Procedure**

The study took place in quiet rooms of nursery schools. All sessions were videotaped. In the warm-up phase, the experimenter (E1) introduced the two children the dollhouse. She asked them to choose one item from each drawer for a room, throw the other in the bin. E1 asked “why” to encourage children to explain their decision for each item. If children did not answer, E1 provided a reason (“Because one needs a toilet in the bathroom”). At the end, E1 showed a photo of the correctly furnished house and said, “You did it correctly” or “You did not pick the correct item for this room, but the rest was correct”, to highlight that there is a correct choice in the game.

In the teaching phase, a second experimenter (E2) asked one child, Child A,
(randomly chosen) to go outside to play with her. With Child A outside, Child B, watched a clip. Then, Child B went outside; Child A watched a clip. We used clips instead of live informants, because children did not want to contradict the adult informants in our pilot data. In all conditions, each child in a dyad watched a different clip.

- **Unequal reliability** condition: children received conflicting information from different sources. If Child A watched the strong-evidence clip with Set 1, Child B watched the weak-evidence clip with Set 2.

- **Equal reliability** condition: children received conflicting information from the same source. If Child A watched the strong-evidence clip with Set 1, Child B watched the strong-evidence clip with Set 2. Eight dyads in each age group watched strong-evidence clips and eight dyads weak-evidence clips.

- **Same information** condition: children received the same information from different sources. If Child A watched the strong-evidence clip with Set 1, Child B watched the weak-evidence clip with Set 1.

The presentation order of the set of items (Set 1 vs. Set 2) and the type of evidence (strong vs. weak) supporting each set were counterbalanced across dyads.

In the testing phase, E1 said, “In our zoo game, selks live here [placing a selk]. In each of these three drawers, there is one item that selks need. You place one in their home and the other in the bin. For example, the box has a tree and a piano. The selks need a tree so I put the tree in their home, the piano in the bin. Now it is your turn: the top drawer first, the middle next, and the bottom last. If you find the right items, you will both get a surprise gift” and left the room until children were finished. If children placed both or none of the items from a drawer, E1 told them that they had to choose one from each drawer and left the room again. At the end, the
experimenter gave each child a sticker. Children’s conversations in the testing phase were transcribed.

**Coding**

First, we coded which three items were chosen. Second, we coded how children reached joint-decisions. For each of the three items, each dyad received one of the following codes:

- *No disagreement*, if children reached joint decision without disagreeing with one another. They nonverbally agreed on an item, or did not oppose when their partner made a proposal (e.g., “Rocks”) or they simply asked each other questions (e.g., “This one or this one?”, “Are you sure?”);

- *Give-up*, if children disagreed with one another and one of them just gave up without explaining why (e.g., “No rocks!”, “No the blue one!”);

- *Turn-taking*, if children disagreed with one another and resolved their disagreement by playing games like rock-paper-scissors;

- *Justification*, if children disagreed with one another, explained each other their reasons for their disagreement, and one of them gave in. The justifications could be in the form of explanations of what they saw in their clips (e.g., “But in my film, it was rocks”), how reliable the information source was (e.g., “She said she was not sure”), and/or ad-hoc explanations to justify a choice of items (e.g., “Sand is made of rocks”).

Next, we coded children’s meta-talk explaining their choice of items with the following categories from the most advanced to the least:

- *Advanced meta-talk* (score 3), if they talked about the reliability of the information for any item and/or identified the different speakers in each clip (a selk vs. a girl called Lisa), so they realized that they received information from
different sources (e.g., “He said, ‘someone told me ...’ . What one said may not be correct”; “But I heard it from selk/girl”; “She was not sure whether it drinks blue soda”; “Ok fine, I am not so sure about that”).

- **Questions** (score 2), if they did not produce advanced meta-talk but tried to elicit meta-talk from one another by asking their partners to reevaluate their proposals or to justify their proposals (e.g., “Where did you hear that?”, “Are you sure?”, “Why?”).

- **Quotes** (score 1), if their meta-talk was only limited to mentioning their information source (e.g., the speaker, the video, etc.) and focusing only on the content of the message (e.g., “He said he eats rocks there”).

- **No meta-talk** (score 0), if they produced proposals without any justifications (e.g., “No blue soda!”).

If a dyad used more than one of the above strategies, it received the score of the more advanced strategy.

A second-coder, who is blind to the predictions and conditions, recoded 25% of the transcripts (24 dyads: 4 dyads from each age and condition) for the kinds of joint decisions and meta-talk and the agreement was $\kappa = .81$ and $\kappa = .82$ respectively.

**Results**

First, we analyzed whether there were differences in children’s choice of the items across age groups and conditions by comparing the mean number of “correct items” supported by strong evidence to chance. In the equal reliability condition, since there was no correct choice, we compared the mean number of Set 1 items to chance. In the unequal reliability condition, both age groups chose the correct items significantly above chance (5-year-olds: $t(15) = 5.48, p < .001, d = 1.37$; 7-year-olds:...
In the equal reliability condition, both age groups’ preference for Set 1 was at chance (5-year-olds: $t(15) = -1.36, p = .19, d = 0.34$; 7-year-olds: $t(15) = 1.07, p = .29, d = 0.27$). In the same information condition, both age groups chose the correct items significantly above chance (5-year-olds: $t(16) = 9.27, p < .001, d = 2.25$; all 7-year-olds chose the correct items).

![Figure 1](image.png)

**Figure 1.** The mean number of correct items chosen by each dyad across age groups and conditions compared to chance. Since there are no correct items in the condition with equal reliability, the y-axis plots the mean number of Set 1 items for this condition. The error bars indicate the standard error and the numeric values indicate standard deviation.

In the second analysis, we analyzed whether children’s likelihood of expressing and justifying their disagreements changed across conditions and age groups. We used a Generalized Linear Mixed Model (GLMM) with binomial error distribution. The unit of analysis was the decision made for each item. The response variable was a binary measure of whether the children justified their disagreement or not. The full model included the predictors: age group (5 vs. 7), condition (unequal reliability, equal reliability, same information) and their interaction, the order of items (1-3), and the random factor of dyad (as we had repeated observations from each dyad). The null model included the order of items and the random factor of dyad. The full model improved the fit as compared to the null model ($\chi^2 = 60.32, df = 5, p < .001$). To test the significance of the interaction term between age group and
condition, we compared the full model with a reduced model without this interaction term, and the full model did not improve the fit suggesting that the interaction was not significant ($\chi^2 = 1.35, df = 2, p = .51$). Therefore, we dropped this interaction term to get interpretable tests of the main effects. The reduced model revealed two significant main effects. The main effect of condition suggested that children in the same information condition were less likely to express or justify their disagreements than the children in other two conditions ($\chi^2 = 48.53, df = 2, p < .001; z's < -4.66, p's < .001$, see Figure 2a). The unequal and equal reliability conditions did not differ from one another ($z = 0.43, p = .67$). The main effect of age group suggested that 7-year-olds were more likely to express and justify their disagreements than 5-year-olds ($\chi^2 = 13.23, df = 1, p < .001$). There was no significant order effect ($\chi^2 = 0.15, df = 1, p = .69$). We also ran the same GLMM for the equal reliability condition to investigate whether the likelihood of expressing and justifying their disagreements among the dyads who watched strong evidence clips differed from those who watched weak evidence clips. The full model included the predictors: age group (5 vs. 7), type of evidence (weak evidence vs. strong evidence) and their interaction, the order of items (1-3), and the random factor of dyad. The null model included the order of items and the random factor of dyad. The full model improved the fit as compared to the null model ($\chi^2 = 13.95, df = 3, p = .003$). We tested the significance of the interaction term between the age group and the evidence type as in the previous analysis and it was not significant ($\chi^2 = 1.55, df = 1, p = .21$) so we dropped this interaction term from the model. The reduced model revealed two significant main effects. The main effect of evidence type suggested that dyads who saw weak evidence clips were less likely to express or justify their disagreements than those dyads who saw strong evidence clips ($\chi^2 = 7.61, df = 1, p = .006$, see Figure 2b). The
main effect of age group suggested that 7-year-olds were more likely to express and justify their disagreements than 5-year-olds ($\chi^2 = 5.85, df = 1, p = .016$). There was no significant order effect ($\chi^2 = 0.35, df = 1, p = .55$).

Figure 2a. The mean proportion of trials with different kinds of joint decisions across the conditions and age groups.

Figure 2b. The mean proportion of trials with different kinds of joint decisions in the equal reliability condition.

Finally, we looked at children’s use of meta-talk. As Table 1 suggests, there were four 5-year-old dyads and ten 7-year-old dyads that produced advanced meta-talk. All four of the 5-year-old dyads referred to the certainty in their meta-talk. With two of these dyads, the child who saw the weak-evidence clip produced meta-talk (referring to uncertainty). With the other two dyads, the child who saw the strong-
evidence clip produced meta-talk (referring to certainty). Within the 7-year-old dyads, four dyads referred to certainty, three dyads referred to knowledge access, and three dyads referred to both.

We investigated whether dyads’ production of meta-talk differed across age groups and the three conditions, using ordinal regression model because our response variable was children’s meta-talk score, which was in ordinal scale. The full model included the predictors age group (5 vs. 7), condition (unequal reliability, equal reliability, same information) and their interaction. The null model did not include any of these factors. The full model improved the fit as compared to the null model ($\chi^2 = 41.56, df = 5, p < .001$). To test the significance of the interaction term between age group and condition, we compared the full model with a reduced model without this interaction term, and the full model did not improve the fit suggesting that the interaction was not significant ($\chi^2 = 0.34, df = 2, p = .845$). Therefore, we dropped this interaction term to get interpretable tests of the main effects. The reduced model revealed 2 significant main effects. The main effect of condition suggested that children in the same information condition were less likely to produce meta-talk than the children in other two conditions ($\chi^2 = 28.71, df = 2, p < .001; z’s < -3.88, p’s < .001$, see Table 1 and Figure 3). The unequal and equal reliability conditions did not differ ($z = 1.04, p = .298$). The main effect of age group suggested that 7-year-olds were more likely to produce meta-talk than 5-year-olds ($\chi^2 = 14.01, df = 1, p < .001$).

Table 1. Number (proportion) of dyads with meta-talk strategies.

<table>
<thead>
<tr>
<th></th>
<th>Advanced meta-talk</th>
<th>Questions</th>
<th>Quotes</th>
<th>None</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unequal Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year-olds</td>
<td>2 (0.13)</td>
<td>2 (0.13)</td>
<td>6 (0.38)</td>
<td>6 (0.38)</td>
<td>16 (1.00)</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>6 (0.35)</td>
<td>3 (0.18)</td>
<td>7 (0.41)</td>
<td>1 (0.06)</td>
<td>17 (1.00)</td>
</tr>
<tr>
<td>Equal Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year-olds</td>
<td>2 (0.13)</td>
<td>2 (0.13)</td>
<td>3 (0.19)</td>
<td>9 (0.56)</td>
<td>16 (1.00)</td>
</tr>
</tbody>
</table>
The following excerpt shows a conversation about the food item between two 7-year-olds in the unequal reliability condition. Andi saw the strong-evidence clip with rocks (set 1), and Sam saw the weak-evidence clip with sand (set2). They first disagreed with one another (lines 7-8) and discovered that they watched different videos (lines 11-14), after Sam asked Andi how he knows that the bowl of rocks is the correct decision (line 11). Then, they both described what they saw in their videos. In lines 14 and 16, Andi justified his proposal by saying that they have a selk (the 3-d selk figure) and he saw the selk itself who said “selks eat rocks”. In line 11, Sam agrees to take the bowl of rocks. Eventually, they choose the correct item.

**Example 1**

[6 lines skipped in which they disagreed about whether it is sand or rocks]

7 Sam: *Nur Sand.*
Only sand.

8 Andi: *Nur Steine.*
Only rocks

9 Sam: *Hmm, also was soll nun richtig sein?*
Hmm but what is right thing then?

10 Andi: *Das ist richtig.* [pointing at the rocks]
This is right. [pointing at the rocks]

*11 Sam: *Woher willst du das wissen?*
How do you know?
The following excerpt shows a conversation about the food item between two 5-year-olds in the unequal reliability condition. Tom saw the strong-evidence clip with sand (set2), and Nico saw the weak-evidence clip with rocks (set1). As can be seen, there is no disagreement about the choice between rocks and sand. Nico immediately agrees with Tom in line 2 and then says, “Hopefully that was right”. Tom reassures him and justifies his proposal of sand by saying that this is what he saw in the video. Eventually, they chose the correct item.

**Example 2**

1. **Tom**
   
   *Guck mal. Was braucht der? Sand!*
   
   Look, what does it need? Sand!

2. **Nico**
   
   *Ja, Sand. ... Nein, in die Schublade- der [: rocks] kommt in den Karton! Hoffentlich war das richtig, Sand.*
   
   Yes, sand … no, into the drawer- it [: rocks] goes into the bin. Hopefully that was right, sand.

3. **Tom**
   
   *Ja, wir haben doch gerade eben das Video gesehen.*
   
   Yes, we did see that in the video.
Discussion

The results of Study 1 showed that in the critical condition, the unequal reliability condition, when children received conflicting information from different sources, both age groups endorsed the information supported by strong evidence in their joint decisions. When they received conflicting information from equally reliable sources (equal reliability condition), they chose the items randomly. These results suggest that by age 5, children are able to reach the “correct” joint decisions and endorse the information that contradicted their prior knowledge, when their prior knowledge was from a less reliable source.

While reaching these joint-decisions, the communicative strategies differed across age groups and conditions. The two conditions in which children received conflicting information (equal and unequal reliability conditions) elicited significantly more justifications for disagreements as well as meta-talk than the condition that children received the same information (same information condition). Moreover, 7-year-olds produced more meta-talk than 5-year-olds (see Examples 1 and 2). Despite this age difference in the communicative strategies, there was no age difference in their choice of items especially in the main condition, the unequal reliability condition. Yet, how 5-year-olds decided on the correct items in the unequal reliability condition is unclear. Thus, in Study 2a, we modified the unequal reliability condition to better understand how 5-year-olds reached correct decisions. Instead of interacting with a peer, we had individual children interact with a puppet partner because we wanted to control what the puppet partner said, especially how much the puppet disagreed, insisted, and how much meta-talk the puppet produced. We investigated whether individual children would be more likely to change their minds, depending on the reliability of the information they receive and depending on the
reliability of the information their partner reports. In addition, we tested younger preschoolers, 3-year-olds to see whether they would be able to choose the items in the unequal reliability condition in a simplified procedure as this one.

Just like in the unequal reliability condition in Study 1, each child watched a clip and the puppet pretended to have watched a different clip. The child and the puppet received conflicting information. In the strong-evidence condition, children have seen a strong-evidence clip and the puppet pretended to have seen a weak-evidence clip. In the weak evidence condition, children have seen a weak-evidence clip and the puppet pretended to have seen a strong-evidence clip. In both conditions, the puppet disagreed with the child and produced the same amount of meta-talk to draw children’s attention to the information reliability (e.g., “Who did you hear this from? I heard it from a girl/selk”, “Was the girl/selk in your film sure? The girl/selk in my film was [not] sure”). We investigated whether 3- and 5-year-olds would be able to comprehend meta-talk and favor the items supported by the strong evidence, since the literature has shown that 5-year-olds, and 3-year-olds to some extent, could evaluate an informant’s certainty and the argument quality (Baldwin & Sabbagh, 2001; Carriveau & Kurkul, 2014; Mercier et al., 2014).

Study 2a

Method

Participants

Forty 3-year-olds ($M = 3;9$, $Range = 3;6–4;0$, 20 girls) and forty 5-year-olds ($M = 5;9$, $Range = 5;5–6;1$, 20 girls), who did not participate in Study 1, participated in the study. There were 3 additional children (two 3-year-olds and one 5-year-old), who could not be included in the analyses, because they either refused to give any
answers or gave irrelevant answers to questions. The children were native speakers of German and had various socio-economic backgrounds.

Materials

The materials were the same as Study 1.

Procedure

The warm-up was the same as Study 1, except that a puppet called “Maxi” replaced one of the children and acted like a partner of the child. While decorating the dollhouse, the puppet asked the child which item they should choose. If the child responded with the correct item (e.g., choosing toilet for the bathroom), the puppet agreed with the child. If the child responded incorrectly (e.g., choosing lamp for the bathroom), the puppet proposed to place the correct item and let the child decide eventually. Then the experimenter said, “Now I have a new game for you two. Now you [the child] will watch a film inside and Maxi will watch a film outside”. Then the puppet left the room.

In the teaching phase, each child watched one of the four clips twice. Then, each child additionally heard E1 commenting on the clip to draw children’s attention to information sources. The experimenter said, “Wow you saw the selk itself”, if the child had seen a strong-evidence clip; and said, “Hmmm, Lisa was not sure at all, right?”, if the child had seen a weak-evidence clip. In the weak-evidence condition, children watched a weak-evidence clip (e.g., with set 1), and the puppet pretended to have seen a strong-evidence clip (e.g., with set 2). In the strong-evidence condition, children watched a strong-evidence clip (e.g., with set 1), and the puppet pretended to have seen a weak evidence clip (e.g., with set 2). The set of items (set 1 vs. set 2) supported by weak vs. strong evidence was counterbalanced.
Like the testing phase in Study 1, E1 instructed the child and the puppet jointly decide and choose one item that the selks need from each of the three drawers and left the room. Alone with the child, the puppet led the conversation and said, “Did you see the film? I saw the film too. Let’s open the top drawer”. After the child took out the bowl of sand and the bowl of rocks, the puppet asked, “What was the food in your film? Was it sand like this? Or was it rocks like this?”. Then the puppet said, “In my movie, the food was sand/rocks. Who said that in your film? Was it a selk like this [pointing at the animal] or someone else?”. Then the puppet reported her own source “In my movie, it was the selk like this” or “In my movie, it was a girl called Lisa” depending on which clip the puppet was pretending to have seen. Next, the puppet said, “Was the selk/girl sure?”. Finally, the puppet said, “In my film, the selk was sure” or “In my film, the girl was not sure at all”. The rates of correct responses for these questions were the following:

a. “What was the [food/drink/sleep] item in your clip?”: 3- and 5-year-olds responded to this question correctly 86% and 98% of the time respectively.

b. “Who said that?”: 3- and 5-year-olds responded to this question correctly 80% and 90% of the time respectively.

c. “Was she sure?: 3- and 5-year-olds responded to this question correctly 58% and 80% of the time respectively. If we only look at the responses in the weak-evidence condition, 3- and 5-year-olds responded correctly 30% and 60% of the time respectively.

At the end, the puppet asked the child which item they should choose and asked the child why they chose that item to give them a chance to explain their decision.
Coding

We first coded which items children chose. Next we coded how children justified their choices with the following categories from most advanced to least:

- **Meta-talk**: reference to information reliability such as “The selk was sure”, “Lisa was not sure”;
- **Quotes**: reference to the content of the message without meta-talk such as “Because it is their food”, “In my film, they eat rocks”;
- **No/incomplete/irrelevant** justifications such as “Because I like green”.

A second coder recoded the justification types of 20%\(^1\) of the children (24 children in all conditions and ages) and the agreement was \(\kappa = .86\).

Results

We first analyzed whether children chose the correct items supported by strong evidence and compared the number of correct items chosen by children to chance. In the strong-evidence condition in which the correct items corresponded to what children had seen in their videos, both age groups were able to choose the correct items significantly above chance (3-year-olds: \(t(19) = 4.27, p < .001, d = 0.95\); 5-year-olds: all children choose the correct items; see Figure 4). In the weak-evidence condition in which the correct items did not correspond to what the children had seen in their videos, 3-year-olds were marginally below chance and chose mostly what they have seen (\(t(19) = -1.85, p = .080, d = 0.41\), see Figure 4); whereas 5-year-olds were at chance and did not show a preference between the items they have seen in the videos and the items that the puppet suggested (\(t(19) = -0.76, p = .456, d = 0.17\), see Figure 4).

---

\(^1\) Because the coding in Study 2a and Study 2b was the same, the sample in the reliability coding also included justifications from Study 2b.
Next, we analyzed whether children’s preference for the items that they had seen in the videos (“own” items) changed across conditions and age groups, using ANOVA. The response variable was the number of “own” items (0-3). The between-subjects factors were age group (3, 5) and condition (weak-evidence, strong-evidence). There was a significant interaction between condition and age group ($F(1,76) = 6.15$, $p = .015$, $\eta^2_p = .06$). The post-hoc comparisons suggested that 3-year-olds favored their own items equally in both conditions (TukeyHSD, $p = 0.571$); whereas 5-year-olds favored their own items significantly less often in the weak-evidence condition than in the strong-evidence condition (TukeyHSD, $p < 0.001$). Moreover, there was no age difference in the way children favored their own items in the weak-evidence condition (TukeyHSD, $p = 0.881$), but 5-year-olds favored their own items significantly more than 3-year-olds did in the strong-evidence condition (TukeyHSD, $p = 0.035$).

In their justifications, children often used quotes of what they heard without meta-talk “Because it is their food” or “In my film, they eat rocks” (3-year-olds: 43%; 5-year-olds: 68%). 3- and 5-year-olds produced meta-talk in their justifications such as “She wasn’t sure”, “The selk was sure” only in 3% (4 out of 120 questions) and 7% (8 out of 120 questions) of the time respectively.
Discussion

The results of Study 2a showed that 3-year-olds mostly chose the items that they have seen in their clips, regardless of the reliability of the information they receive and regardless of what the puppet partner said; whereas 5-year-olds chose their own items significantly less often when their information source was unreliable than when it was more reliable. However, 5-year-olds’ choice of correct items was still at chance in the weak-evidence condition, just like 3-year-olds’. The age difference between 3- and 5-year-olds was apparent in the strong-evidence condition rather than the weak-evidence condition.

The results of Study 2a clarified the findings with 5-year-olds in Study 1. The results of Study 2a suggest that it is the children who saw the strong-evidence clips seem to have done more of the interactive work in Study 1. In Study 2, when 5-year-olds saw the strong-evidence clips and the puppet stated that its informant was unsure, 5-year-olds were confident about their knowledge. Even if their partner disagreed with them (sometimes in Study 1, always in Study 2), they never changed their minds and stuck to their own items. Moreover, when 5-year-olds saw weak-evidence clips, they were more likely to change their minds than when they saw strong-evidence clips. They preferred their own items and the items suggested by the puppet equally frequently. Thus, the combination of these two factors (the confidence by the children who saw strong-evidence clips and the willingness to be convinced by the children who saw weak-evidence clips) seemed to have led the 5-year-olds choose the correct items in Study 1 without much disagreement.

3-year-olds’ choice of items, on the other hand, was not influenced by the reliability of the information they received or by what the puppet said. In both conditions, 3-year-olds favored their own items and changed their minds equally
frequently upon hearing a conflicting view from the puppet partner. One remaining concern in Study 2a was the high rates of incorrect responses to the control question about certainty: “Was Lisa sure in your film?” In the weak-evidence condition, 3- and 5-year-olds answered this question correctly by saying “no” only 30% and 60% of the times respectively. Despite hearing the girl saying, “I am not sure” three times in the weak evidence clips, and E1 repeating this after each weak-evidence clip, children mostly answered this question by saying “yes”. We suspect that this was due to “yes-bias” in children’s answers to yes-no questions (Fritzley & Lee, 2003; Okanda & Itakura, 2010). Thus, we ran Study 2b, in which we asked additional questions. We first asked two yes-no questions about the certainty of informant: “was she sure?” and “did she say she was sure?”. Finally, we asked a forced-choice question “Did she say ‘I am sure’ or ‘I am not sure’?” and gave children two options to choose from to see whether children could accurately report what the informant said. We also simplified the instructions, especially for 3-year-olds, emphasizing that in there is a correct item and an incorrect item in each drawer and that the puppet and the child will have to find the correct items together.

**Study 2b**

*Participants*

Twenty-one 3-year-olds ($M = 3;10$, $Range = 3;7–4;0$, 10 girls) and nineteen 5-year-olds ($M = 5;9$, $Range = 5;6–6;0$, 10 girls), who did not participate in Study 1 or Study 2a, participated in the study. The children were native speakers of German and had various socio-economic backgrounds.

*Materials*

The materials were the same as Study 1 and Study 2a.
Procedure

The procedure was the same as Study 2a, except for the two following changes:

1) When giving instructions to the child and the puppet, E gave simplified instructions and emphasized that there is one correct and one incorrect item in each drawer (“In each drawer, there are two things. But be careful! In each drawer there is one correct item and there is one incorrect item. If you two find the correct items together, you will both get a surprise gift at the end”).

2) When the puppet and the child were deciding on the item, the puppet asked the same set of 3 questions, as she did in study 2a:

   a. “What was the [food/drink/sleep] item in your clip?”: 3- and 5-year-olds responded to this question correctly 79% and 98% of the time respectively.
   b. “Who said that?”: 3- and 5-year-olds responded to this question correctly 57% and 74% of the time respectively.
   c. “Was she sure?`: 3- and 5-year-olds responded to this question correctly 30% and 67% of the time respectively.

For each item, the puppet asked 2 more questions about the informant certainty:

   d. “Did she say she was sure?”: 3- and 5-year-olds responded to this question correctly 35% and 70% of the time respectively.
   e. “Did she say ‘I am sure’ or ‘I am not sure’?”: 3- and 5-year-olds responded to this question correctly 52% and 91% of the time respectively.

At the end the puppet asked the child why she/he chose what she/he chose to elicit some justifications.
Coding

The coding was the same as Study 2a: we first coded which items children chose and how children justified their choices.

Results

We analyzed whether children chose the correct items supported by strong evidence and compared the number of correct items chosen by children to chance. The results replicated the weak-evidence condition in Study 2a. 3-year-olds were marginally below chance and mostly chose what they have seen, rather than the ones supported by strong evidence ($t(20) = -1.97, p = .063, d = 0.43$, see Figure 5); whereas 5-year-olds were at chance and did not show a preference between the items they have seen in the videos and the items that the puppet suggested ($t(18) = -0.96, p = .350, d = 0.19$, see Figure 5). There was also no age difference in children’s choice of the correct items ($t(38) = 0.32, p = .75, d = 0.10$).

![Figure 5](image)

**Figure 5.** The mean number of correct items chosen by each age group in weak evidence condition. The error bars indicate the standard error and numeric values indicate the standard deviation.

Similar to Study 2a, when children were asked to justify their choices, they usually answered the last question by saying things like “Because it is their food” (3-year-olds: 25%; 5-year-olds: 26%). Similar to Study 2a, in response to “why”-questions, 3-year-olds never used meta-talk in their justifications and 5-year-olds did so only 10% of the time.
Discussion

The results of Study 2b replicated the results of Study 2a in terms of children’s choice of items and their justifications for these items. Although there was no significant age difference, 3-year-olds preferred their own items more than the items suggested by the puppet; whereas 5-year-olds preferred their items and the items suggested by the puppet equally frequently. More importantly, however, with the addition of forced choice control question (“Did she say ‘I am sure’ or ‘I am not sure’?”), the rate of correct answers increased to 52% for 3-year-olds and 91% for 5-year-olds. Thus, the high number of incorrect responses to the certainty question (“was she sure?”) in Study 2 was due to the “yes-bias” for 5-year-olds.

3-year-olds continued to have difficulty with the certainty question. One explanation for this difficulty could be that unlike other studies with younger preschool children (e.g., Baldwin & Sabbagh, 2001; Corriveau & Harris, 2009) in which children witnessed both informants, children in our study witnessed the testimony of only one informant (without a reference point) and heard the verbal evaluation of the other informant (or a meta-testimony) by the puppet. This required 3-year-olds to comprehend sentences about a third-person’s mental states or a person’s certainty (e.g., “She was not sure”). This difficulty by 3-year-olds is in line with the literature which shows that children’s understanding of other’s mental states and their comprehension of third-person references to mental states develop in later preschool ages (Brandt, Buttelmann, Lieven, & Tomasello, 2016; see also Wellman, Watson, & Cross, 2001; Wimmer & Perner, 1983).

General Discussion
Overall, our results showed that when children and their partners received conflicting information from sources that differ in reliability, 5- and 7-year-olds, but not 3-year-olds, trusted the information supported by the strong evidence rather than the weak evidence in their joint decisions. In Study 1, around half of the 7-year-old peers engaged in various kinds of meta-talk in reaching their joint decisions. They verbally evaluated the reliability of the sources (e.g., “She was not sure whether it drinks blue soda”, “He said, ‘someone told me ...’ What one said may not be correct”). They tried to resolve their disagreement by “publicly” and jointly examining the reliability of the information that each of them had (see Example 1). They, in fact, tried to elicit information about the credibility/reliability of the information from one another by asking questions like “How do you know that?”, “Are you sure?”. As discussed earlier, talking about the credibility of information is an advanced meta-cognitive skill, which had previously only been observed in adolescents and after intervention (Kuhn et al., 2013). Our findings suggest that as early as 7 years of age, children begin to talk about these to justify their decisions spontaneously in their conversations with peers. The frequency of these meta-argumentative strategies by children potentially increases with age during school ages.

5-year-olds, on the other hand, limited their meta-talk to the content of the messages such as “In my story, they ate rocks”, “We did see that in the video” by mentioning the information source without mentioning its reliability. This was true in their spontaneous justifications in Study 1 and elicited justifications in Study 2. However, their choices of items in Study 1 and Study 2 suggest that 5-year-olds were sensitive to the reliability of the information source. They were less likely to change their minds, when they received reliable information (from strong-evidence clips).
than when they received unreliable information (from the weak-evidence clips). In fact, as the equal reliability condition in Study 1 showed, when both children in a dyad watched weak-evidence clips, they were less likely to express or justify their disagreements than children who watched strong-evidence clips. This suggests that children privately judged that their information source was unreliable so they did not contradict their peers’ choices. Nonetheless, 7-year-olds overall expressed their disagreements more than 5-year-olds did, regardless of whether the information was supported by weak or strong evidence, suggesting that 7-year-olds were more likely to evaluate the information reliability more publicly than 5-year-olds.

The results of Study 1 (together with examples 1 and 2) and Study 2a suggest that in the unequal reliability condition, both children contributed to the making correct decisions. The children who saw the strong-evidence clips insisted on their items. The children who saw the weak-evidence clips, on the other hand, were more willing to change their minds. Thus, our findings illustrate that joint decisions made in cooperative peer interactions (or in contexts of trust), in which getting it right benefits both parties, motivates and facilitates young children’s reasoning (see also Köymen, Rosenbaum, & Tomasello, 2014; Köymen, Mammen, & Tomasello, 2016). In Study 1, two peers, despite their conflicting views on what the correct items were in the unequal reliability condition, had a joint goal of making the right decision and both age groups were able to reach the correct decisions. More importantly, the two conditions in which the children began with conflicting views seemed to elicit meta-talk in a way that arguing only to win almost certainly would not (as each arguer would then only be seeking to confirm her own point of view). Hearing a conflicting view in the context of collaborative decision-making not only reminds one to question the reliability of one’s own knowledge, but also to demand that the partner justify
his/her line of reasoning and any evidence or reasons for it more fully, perhaps with meta-talk about the overall reasoning process (Kuhn, 2015; Tomasello, 2014).

There is one important question that remains unanswered. In our procedure, the weak evidence clips involved different markers of unreliability such as limited knowledge access (a second-hand report as opposed to first-hand report), expressions of uncertainty (e.g., “I am not sure”; “This is strange”). 5-year-olds only referred to the informant’s uncertainty because it was perhaps easier for them to simply repeat what the informant said. However, overall 5-year-olds did not produce enough meta-talk to reach any conclusion. On the other hand, 7-year-olds referred to knowledge access and uncertainty equally often. This might suggest that 7-year-olds judged both knowledge access and certainty of the informant as important cues to information reliability. Still, we cannot tell from the current data what markers of unreliability mattered more for children. Thus, future research should pit these different markers of unreliability against one another to disentangle what markers of unreliability children pay attention to more.

To conclude, when working towards a joint goal, young children display sophisticated reasoning skills. They are able to disregard their own prior knowledge, especially when they acquired this knowledge from an unreliable source, and accept the alternative proposed by their peer. They also begin to use meta-talk in their spontaneous peer conversations to justify their proposals. Our results support the view that children's joint reasoning is a fundamentally cooperative enterprise aimed at making jointly rational decisions.
References


Appendix A. The strong evidence clips.

**Set 1**

Hallo Ich bin ein Selk.
‘Hello I am a Selk’.

‘We, the selks, eat rocks/sand. Look how I eat rocks/sand here. We eat only rocks/sand, nothing else.’

‘We drink blue/green soda. Look how I drink blue/green soda here. We drink only blue/green soda, nothing else.’

‘We sleep in water/on flowers. Look how I sleep in water/on flowers here. We sleep only in water/on flowers, nowhere else.’
Appendix B. The weak-evidence clips.

Set 1

‘Hello I am Lisa. I heard about an interesting animal. I believe they are called selks.’

Set 2

‘Someone told me that they eat only rocks/sand like these here. Weird, perhaps they can eat only rocks/sand, but I am not sure.’

Jemand hat mir erzählt, dass sie nur blaue/grüne Limonade wie diese hier trinken. Komisch, vielleicht können sie nur blaue/grüne Limonade trinken, aber ich bin mir nicht sicher.
‘Someone told me that they drink only blue/green soda like this here. Weird, perhaps they can drink only blue/green soda, but I am not sure.’

Jemand hat mir erzählt, dass sie nur im Wasser wie diesem hier schlafen. Komisch, vielleicht können sie nur im Wasser schlafen, aber ich bin mir nicht sicher.
‘Someone told me that they sleep only in water/on flowers like this here. Weird, perhaps they can sleep only in water/on flowers, but I am not sure.’