Running Head: META-TALK

Preschool children's use of meta-talk to make rational collaborative decisions

Kirstie Hartwell[1], Silke Brandt[2], Laura Boundy[1], Grace Barton[1], Bahar Köymen[1]

[1]University of Manchester,

[2]Lancaster University

Corresponding author: Kirstie Hartwell

School of Health Sciences, University of Manchester, Oxford Road, Coupland 1 Building

M13 9PL, Manchester, United Kingdom.

E-mail: kirstie.hartwell@manchester.ac.uk

Acknowledgements: We would like to thank the children, their parents, the nurseries and the schools for their friendly cooperation. We would like to thank Manuel Bohn for his feedback on statistical analyses and Anna Theakston for her general support for Study 2. This publication was made possible through the support of a grant from Templeton World Charity Foundation, Inc., awarded to Bahar Köymen.

Abstract

In collaborative decision-making, partners compare reasons behind conflicting proposals through meta-talk. We investigated UK-based preschoolers' (mixed SES) use of meta-talk (Data collection: 2018-2020). In Study 1, 5 and 7-year-old peer dyads (*N*=128, 61 girls) heard conflicting claims about an animal from two informants. One prefaced her claim with "I know"; the other with "I think". Dyads identified the more reliable informant through meta-talk ("She said she knows"). In Study 2, 3- and 5-year-olds (*N*=64, 34 girls) searched for a toy with an adult partner making incorrect proposals. Children refuted this through reporting what they had witnessed (It cannot be there because "I saw it move", "she moved it"). In preschool period, children start using meta-talk to make rational collaborative decisions.

Keywords: reasoning, collaborative problem-solving, meta-talk, peer interactions

Preschool children use meta-talk to make rational collaborative decisions

Recent accounts of reasoning have emphasized its social dimension, where speakers exchange their reasons for their beliefs to come to agreement (Mercier and Sperber, 2011; Tomasello, 2019). The ability to reason with another mind emanates from children's species-unique inclinations for cooperation and shared intentionality, which involves *cooperative motives* through which individuals are after finding a good solution benefitting all parties and *cooperative cognition* through which individuals compare the relative merits of competing reasons to arrive at the correct decision (Köymen & Tomasello, 2020). During collaborative problem-solving, partners often engage in a special kind of reason-giving, termed as "meta-talk" (Köymen and Tomasello, 2018, 2020), through which they talk about standards of evidence (e.g., first-hand observation tops indirect observation), compare the validity of competing reasons, or provide reasons for reasons (Iordanou, 2010; Kuhn, 2001; Mahr & Csibra, 2018; Tomasello, 2019).

Around age 3, children can identify valid and invalid reasons (Castelain, Bernard, & Mercier, 2018; Domberg, Köymen, & Tomasello, 2019; Koenig, 2012). Around age 4, when children master the concept of belief (Wellman, Cross, Watson, 2001), they comprehend meta-talk involving mental state terms (Moore, Bryant, & Furrow, 1989; Sabbagh & Baldwin, 2001). For example, Brandt and colleagues (2016) presented children with informants providing conflicting information about a sticker's location. The reliable informant said, "*I know* it is in the blue box"; the other said, "*I think* it is in the red box". Four-year-olds, but not 3-year-olds, could identify the sticker's location.

Preschoolers are also adept at producing reasons when their partners are unaware of some information (Köymen, Mammen and Tomasello 2016); and adjust the informativeness of their reasons depending on their common ground with their partners (Mammen, Köymen,

& Tomasello, 2018). Yet the ability to produce meta-talk has mostly been observed in adolescents, who could use reasons that challenged their partner's assertions through counterarguments (Kuhn & Udell, 2003; Kuhn, Zillmer, Crowell, & Zavala et al., 2013; Papathomas & Kuhn, 2017).

Two recent studies investigated younger children's use of meta-talk, but provided limited evidence. Köymen and Tomasello (2018) presented 5- and 7-year-old peer dyads with a collaborative problem-solving task. Each child separately received conflicting information about what a novel animal eats (rocks or sand). One child heard a first-hand report by the animal itself; whereas the other child heard a second-hand report expressing uncertainty. When collaboratively deciding what the animal eats, children from both age groups endorsed the item suggested by the first-hand report, but only the 7-year-olds produced meta-talk, where they justified their proposals by talking about the reliability of their information source (e.g., "she was not sure"). The crucial point was that the 7-year-olds were not just talking about whether the animal eats rocks or sand, but engaged in meta-talk, through which they evaluated the relative merits of different types of evidence and referred to the standards of reasoning (e.g., first-hand evidence tops hearsay). However, in this study, the task burden to produce meta-talk mostly fell on the child who received the second-hand report (e.g., to say "she said she heard it from someone", "she was not sure"); whereas the other child who received the first-hand report only saw the animal stating they eat rocks and did not have any information about the reliability of the informant. Thus, this imbalance in the amount of information each child received might have reduced the overall frequency of meta-talk, particularly by the 5-year-olds.

A simpler form of meta-talk could involve counterarguments, through which speakers explain why their partner's reasons are not valid. Counterarguments do not evaluate a proposal per se, but they evaluate the validity of the reason supporting that proposal. The second study reported that when collaboratively deciding which box contained the item a bear needed, 5-year-olds, and even 3-year-olds after training, could successfully refute their partners' reasoning (e.g., "let's pick this box, because it contains honey") through meta-talk in the form of counterarguments (e.g., "no the honey is mouldy"; Köymen, O'Madagain, Domberg, & Tomasello, 2020), indicating that preschoolers might be capable of producing meta-talk. However, in this study children were provided with two counterarguments, one valid and one invalid, and they simply had to selectively repeat what they heard without spontaneously providing these counterarguments.

Therefore, in the current studies, we investigated whether 3-, 5-, and 7-year-olds could produce spontaneous meta-talk about information reliability during collaborative problem-solving. In Study 1, we modified the procedure of Köymen and Tomasello (2018). Five- and 7-year-old peer dyads heard two informants give information about the needs of a novel animal (e.g., its food) together, rather than individually, so that the contrast in the informants' testimony was more salient and each child would be equally likely to produce meta-talk. One informant was more reliable and confident, and prefaced her proposals with, "I know"; the other was less reliable and less confident, and prefaced her proposals with "I think". In the experimental condition, the two informants gave conflicting information (e.g., one stated it eats rocks, the other stated sand); whereas in the control condition, informants gave the same information (e.g., both stated rocks). As part of our confirmatory hypotheses, we predicted that children would choose the item suggested by the reliable and confident informant in both conditions but would justify their proposals with meta-talk (e.g., "she said she knows") more in the experimental condition than in the control condition. In Study 2, we investigated whether younger children, 3- and 5-year-olds, could produce meta-talk in a simpler task, which did not require comprehension or production of mental state verbs or complex sentences.

Study 1

Method

Participants

Sixty-four 5-year-olds (*M* = 5;5, *Range* = 5;0 – 5;11, 34 girls) and 64 7-year-olds (*M* = 7;6, *Range* = 6;11 – 8;0, 27 girls) in 64 same-age dyads (9 mixed-gender dyads) participated in this study. Data collection took place between March 2018 and May 2019. The sample size was decided prior to data collection based on studies using similar procedures (e.g., Köymen & Tomasello, 2018). The dyads were formed based on the teacher's recommendations and were randomly assigned to the experimental or control condition (16 dyads in each condition and age group). An additional two 5-year-old dyads were excluded from the analyses because they did not finish the game. Children were recruited from various schools in North West England. They were all native speakers of English and had various socio-economic backgrounds.

Materials

The warm-up game used a set of three drawers and a dollhouse which had a bathroom with a bathtub and a sink, a bedroom with a bed and a closet, and a kitchen with an oven and a sink. Each drawer included two items: one that typically belonged to a room and needed to be placed in the corresponding room (a toilet, a cot, a fridge), and one that did not and so was to be thrown away (a lamp, a table, a ladder).

In the teaching phase, each dyad watched two videos together, narrated by different female informants, about a novel creature named "selk" on two iPads. Each video involved a slideshow with 4 drawings. In slide 1, the informant introduced herself. In the next three slides, the informant listed what the animal eats (rock or sand), drinks (blue or green soda), and where it sleeps (water or grass). Rocks, blue soda, and water were grouped as Set 1 items; sand, green soda, and grass as Set 2 items. One informant was reliable and used the verb "know" for each item (e.g., "I *know* they eat rocks because I always give them their food") whereas the other informant was less reliable and used the verb "think" for each item (e.g., "I *think* they eat sand, but I never give them their food"; see Appendix A for the pictures and the narration).

In the experimental phase, dyads decorated the selks' home with the items in the three drawers. There were rocks and sand in the top drawer, a blue soda and a green soda in the middle drawer, a lake and grass pictures in the bottom drawer.

Procedure

The study took place in quiet areas of children's schools. All sessions were videotaped. In the warm-up, the experimenter (E) introduced the dollhouse game, "Each drawer contains a correct item that should go to one room and an incorrect item that should go to the bin." E asked the dyads to name the first room (bathroom) and the items in the top drawer (a toilet, a lamp). Next, E asked them to choose the correct item and asked "why" to prime children to justify their choices. If they provided no reason, E provided one ("because one needs a toilet in the bathroom"). This procedure was repeated for the bedroom and kitchen. Finally, E showed a picture of the correctly furnished house and gave dyads feedback ("All correct!", "You were wrong for this room, but the rest is correct!") to emphasize that there were correct choices to prevent them from choosing the items randomly.

In the teaching phase, E said, "Now you will watch two videos about an animal called Selk. Each video has a different girl and only one is telling the truth". One video, shown on one iPad, always had the reliable informant who used the verb "know" whereas the other video, shown on a second iPad, always had the less reliable informant who used the verb "think". In the *experimental condition*, the informants gave conflicting information (one informant presented set 1 items, the other set 2 items). After the first video, E placed a picture of set 1 items listed in the video in front of the first iPad and repeated what the reliable informant had said, for instance, “She said she knows selks eats rocks, she knows they drink blue soda, and she knows they sleep on water”. After the second video, E placed the picture of set 2 items mentioned in the video in front of the second iPad and said, for instance, “She said she thinks selks eat sand, she thinks they drink green soda, and she thinks they sleep on water”. In the *control condition*, the procedure was the same except that both informants suggested the same information (e.g., both suggested set 1 items or set 2 items).

In the experimental phase, E instructed the dyads, “Each drawer has one correct item and one incorrect item. You should place the correct ones in selks’ home. Only one girl is right. You need to decide together which girl. You will get a surprise gift, if you choose the correct items. Start with the top drawer, then the middle drawer and finally the bottom drawer”. E then left the room. When the dyads finished, E asked, “You thought this girl was right as all/most of these items were hers. Why her?” so children could report their reasoning to E. Regardless of the items chosen, E gave children a picture as a gift.

The identity of the reliable informant, the items that each informant mentioned (Set 1, Set 2) and the order in which the reliable information was presented in the videos were counterbalanced. The procedure lasted around 15 minutes.

Coding

First, we coded the number of items suggested by the more reliable informant that the dyads chose (0-3). Then we coded whether children produced meta-talk in two contexts: 1) *spontaneous meta-talk* in their peer conversations, including when they were watching the videos (e.g., children occasionally produced meta-talk during the pause when E switched from one iPad to the other to play the videos); 2) *elicited meta-talk* in their responses to E’s *why*-question after the task. Meta-talk included the reasons that referred to the informants’ reliability, particularly how they know what they are suggesting. This could involve references to the informant's mental states (e.g., "she said that she actually knows", "because that one said, 'I know' and that one said, 'I think'") and the references to whether the informants interacted with the animal or not (e.g., "that girl says that she feeds them", "Because at the end she says, she's never put them to sleep."). A second coder went over 25% of the transcripts (four dyads in each condition and age) to code whether a dyad produced meta-talk. The agreement was $\kappa = .88$.

Results

Peer conversations varied in length. The following is an example of a conversation in the experimental condition by 7-year-olds who heard that the more reliable informant suggesting rocks, blue soda and water.

B: So rocks,
A: And sand. I think are both.
B: Probably rocks. Because that girl said "I know it" but that girl said she's not sure. So yeah that [sand] in the bin. Second draw,
A: Green soda or blue soda?
B: I reckon blue soda.
A: Yeah.
B: Because she-
A: Yeah.
B: Now last is flowers and- flowers in the bin.
A: Definitely this [water].

We carried out four sets of analyses. First, we analyzed children's choice of items. The mean number of items chosen by the peer dyads that were suggested by the reliable informant was significantly above chance in both conditions for each age group (see Figure 1; in the experimental condition, $t(15) = 23$, $p < .001$, $d = 5.75$ for 5-year-olds; $t(15) = 2.86$, $p = .012$, $d = 0.71$ for 7-year-olds; in the control condition, $t(15) = 6.46$, $p < .001$, $d = 1.61$ for 5-year-olds, $t(15) = 4.14$, $p < .001$, $d = 1.04$ for 7-year-olds).

Second, we analyzed whether children's production of *spontaneous* meta-talk during peer conversations differed across conditions and age groups, using Generalized Linear

Model (GLM) using binomial error distribution. The unit of analysis was each dyad. The response variable was the binary measure of whether a dyad produced meta-talk. The full model included the predictors age (5, 7), condition (experimental, control), and their interaction; as well as gender composition of the group (female, male, mixed-gender). The full model improved the fit as compared to the null model without these predictors (χ^2 = 11.98, *df* = 3, *p* = .007). The interaction between age and condition was not significant (χ^2 = 2.43, *df* = 1, *p* = .119). The reduced model without this interaction revealed significant main effects of condition (χ^2 = 4.50, *df* = 1, *p* = .034, 95% CI [0.09 2.37]) and age (χ^2 = 5.41, *df* = 1, *p* = .019, 95% CI [0.20 2.55], see supplementary materials A for the model summary). Both age groups produced meta-talk significantly more often in the experimental condition than in the control condition; 7-year-olds produced meta-talk significantly more often than did 5-year-olds (See Figure 2, left panel).

Third, we analyzed whether the production of spontaneous meta-talk predicted choosing the items suggested by the reliable informant in the experimental condition (we did not include control condition because very few children produced meta-talk in the control condition). We ran a between-subjects ANOVA with production of meta-talk, age, their interaction, and gender composition of the dyad as predictors. We found significant main effects of production of meta-talk ($F(1, 26) = 4.62$, $p = 0.041$, $\eta_p^2 = 0.12$) and age ($F(1, 26) = 5.31$, $p = 0.029$, $\eta_p^2 = 0.13$). Those dyads who produced meta-talk chose significantly more items suggested by the reliable informant than those dyads who did not produce meta-talk (*M* = 2.88, *SD* = .50 vs. *M* = 2.38, *SD* = 1.09). Five-year-olds chose significantly more items suggested by the reliable informant than 7-year-olds (*M* = 2.94, *SD* = .25 vs. *M* = 2.31, *SD* = 1.14). The interaction between age and production of meta-talk or the gender composition of the group was not significant ($F(1, 26) = 2.53$, $p = 0.124$, $\eta_p^2 = 0.06$; $F(2, 26) = 0.76$, $p = 0.479$, $\eta_p^2 = 0.04$).

Finally, we analyzed whether children's production of *elicited* meta-talk (their responses to E's question) differed across conditions and age groups, using the same models as in the third set of analyses. The full model improved the fit of the null model only marginally significantly ($\chi^2 = 7.53$, $df = 3$, $p = .057$, see supplementary materials A for the model summary), suggesting that there was no robust age or condition difference in children's elicited meta-talk (see Figure 2 right panel).

Discussion

Our results suggest that to reach rational collaborative decisions, both 5- and 7-year-olds spontaneously produced meta-talk more often in the experimental condition than in the control condition in their peer conversations and those dyads who produced meta-talk followed the suggestions of the reliable and confident informant more often than the dyads who did not produce meta-talk. Thus, our study is the first to document that children, as young as age 5, can produce meta-talk involving complex sentences with mental state verbs, a skill so far observed only with children of age 7 and older (Köymen & Tomasello, 2018; Papathomas & Kuhn, 2017). Thus, around age 5, when children master the concept of belief (Wellman et al., 2001): someone might have a false belief despite having good reasons for this belief (e.g., I heard this from an informant who might have a false belief), they also gauge and refer to the trustworthiness of the informants based on the mental state verbs used by these informants in their meta-talk. Unlike the study of Köymen and Tomasello (2018), in our study when each child had access to the testimony of both informants, and when the task burden did not fall onto one child, 5-year-olds were able to produce meta-talk involving mental states.

A notable developmental pattern was that the condition difference was overall clearer with 5-year-olds than with 7-year-olds. That is, 7-year-olds produced meta-talk almost equally frequently in both conditions. One explanation for this pattern is that the task was easy for the 7-year-olds and they focused on working out why one girl could be trusted more than the other and produced meta-talk (e.g., "she said she knows") across both conditions. In contrast, the 5-year-olds may have found this assessment of truthfulness more cognitively demanding, so tended to produce meta-talk only when it was necessary for choosing items (e.g., in the experimental condition). Interestingly, however, 5-year-olds chose the items suggested by the reliable and confident informant more than did 7-year-olds. We have observed that the 7-year-olds who were occasionally using their world knowledge and second-guessing their decisions (e.g., "animals do not eat rocks/sand") ended up choosing the items suggested by the less reliable and less confident informant. Nonetheless, each age group ended up choosing the items suggested by the reliable and confident informant significantly above chance in each condition.

Another notable finding was that age and condition did not have a robust effect on children's elicited meta-talk (when the experimenter interviewed them after their decision); whereas the pattern was clearer in spontaneous meta-talk. From the perspective of the children, the task was to choose the right items. Once they were done with their decisions, children were probably more reluctant to talk when E asked further questions. This is also in line with the literature which suggests children do not often answer open-ended why-questions (Mammen, Köymen, & Tomasello, 2021). Spontaneous meta-talk, as compared to elicited meta-talk, was more informative in terms of what actually guided children's reasoning and decision-making.

Although we have documented that 5-year-olds could produce meta-talk, our task was still quite linguistically demanding, as it required children to master the concept of belief, comprehend, and produce complex sentences with mental state verbs, and would certainly not be suitable for younger age groups.

Study 2

In Study 2, we provided children with a new task, which was less linguistically demanding. First, instead of watching videos of second-hand reports, which requires comprehension of mental state verbs, children witnessed events live and could rely on their first-hand observation rather than someone else's report of their first-hand observation. Second, instead of producing meta-talk with mental state verbs, children could produce meta-talk through simple sentences (e.g., "I saw it"). The experimenter (E1) tasked 3- and 5-year-olds with finding where a toy was hiding together with an adult partner (E2). One house had the toy's footprints in front of it. In the experimental condition, after the child and E2 decided that the toy must be in the house with the footprints, E2 left the room and E1 moved the toy to a new house. E2's belief of the toy's location was outdated, and the child had to convince E2 using meta-talk ("I saw it there"). In the control condition, E1 moved the toy to a new house then back into the original house so E2's belief remained correct. As part of our confirmatory hypotheses, we predicted that children would produce meta-talk more often in the experimental condition than in the control condition. We also predicted that 5-year-olds would produce meta-talk more often than 3-year-olds.

Method

Participants

32 3-year-olds (*M* = 3;9, *Range* = 3;6 – 4;0, 18 girls) and 32 5-year-olds (*M* = 5;5, *range* = 5;1 – 5;11, 16 girls) participated in the study. Data collection took place between December 2018 and March 2020. The sample size was decided prior to data collection based on studies using similar procedures (e.g., Köymen et al., 2020). Each child participated in both conditions (experimental and control). An additional six children were excluded from

the analysis due to experimenter error (3), refusal to play (2), and poor attention (1). Children were recruited from various nurseries/schools in North West England. They were all native speakers of English and had various socio-economic backgrounds.

Materials

There were three toys: an elephant, a doll, and a truck. In each trial (one warm-up, two critical trials), a toy was hidden in one of the three houses (cardboard boxes), and one house had the path with the footprints that belonged to one of the three toys.

Procedure

The study took place in quiet areas of the nurseries and schools. All sessions were video taped. The child and the two experimenters (E1, E2) first did some coloring for familiarization.

In the warm-up, E1 told the child and E2 that they were going to play a hide-and-seek game. E1 then said, “The toy elephant will hide in one of the three houses. You two will find which house”. E1 then said that the elephant liked to play in paint, dipped it into paint and showed it leaving footprints behind on a sheet of paper. E1 commented, “Look, the elephant left footprints!” With E2 and the child turned away, E1 hid the elephant in the middle box, which now had the path with footprints. E1 asked the pair to turn around. E2 asked the child, “so which house are we choosing?” If the child picked the correct house, E2 asked why. If the child chose the incorrect box, E2 said, “I think it’s in another house” until they had picked the correct one. If the child did not refer to the footprints, E2 did (“because of the footprints, right?”). Then E1 revealed the elephant.

The order of conditions was counterbalanced. When the experimental condition was first, the procedure was the same as the warm-up trial, except that the toy was a doll, who left footprints and was hidden in the left box, which now had the path with the footprints. E2 and the child decided on a house the same way they did in the warm-up trial. When E1 was about to reveal the toy, E2 left the room saying she needed to make a phone call. With E2 outside, E1 took the doll out and wiped the paint off its feet. E1 moved the toy by walking it first up the path to the middle house and then up the path to the right house (see Figure 3). Thus, E2's belief was now outdated. E1 asked, "Where is the doll now?" to make sure the child knew where the toy was now located. Most children answered this question correctly. If the child answered incorrectly, E1 said "I think it is in one of the other houses" until they selected the correct house. E1 then left the room and E2 entered the room. E2 reminded the child of the original house they chose, "We said this house [pointing at the original, and now incorrect, house with the footprints]. Are we choosing that house?" If the child agreed, E2 said, "because of the footprints, right?". If the child disagreed and pointed to the right house, E2 challenged this by saying, "but there are no footprints there?" (See Table 1 for E2's prompts). After the child responded (or after five seconds, if there was no response), E2 said, "Which house are we choosing? We need to get this right". After the child pointed at a house, E2 asked why they should choose that house and then agreed with the child. They called E1 back into the room and announced which house they chose. If it was the correct box, E1 said, "It is in this box [showing the inside]"; if incorrect, E1 said "It's not in this box" in a neutral tone.

E1 moved on to the control condition, and said, "Let's play again with a truck." The procedure was the same as the experimental condition. E2 left the room again by saying that she forgot her keys outside. With E2 outside, the child saw E1 move the toy up the path to the middle box, then back to where it was originally, so E2's belief was still correct. When E2 returned, she again said, "We said this house. Are we choosing this house?" and provided the same prompts as she did in the experimental condition, depending on whether the children agreed or disagreed with her (see Table 1). They called E1 back into the room to check the correct box.

The order of conditions, the toys in each condition, and position of the path with traces (left or right) were counterbalanced.

Coding

In the conversations between the child and E2, we identified children's meta-talk operationalized as reasons involving children witnessing the experimenter's act on the toy and/or the toy's relocation, thereby proving their better knowledge access as compared to E2 (e.g., "because I saw that she moved it", "I saw the baby in there", "She's in that house … and I saw that", "she moved it to that house", "because she wiped the feet."). There were some minor experimenter errors with 10 children (all 5-year-olds, 8 in the experimental condition, 2 in the control condition) where E2 repeated the prompts "because of the footprints, right?" or "but there are no footprints there?". If the children provided meta-talk in response to these extra prompts, we coded these children as those who did not produce meta-talk. A second coder coded the 25% of the transcripts (eight children in each age group) for meta-talk. The agreement was $\kappa = .89$.

Results

Eight children (five 3-year-olds and three 5-year-olds) did not correct E2's incorrect proposal in the experimental condition and two children (one 3-year-old and one 5-year-old) suggested to choose the incorrect house in the control condition. In total, ten children ended up choosing the incorrect house in one condition (10 out of 128 trials).

Children received a total of three prompts: 1) But/because there are no footprints there? 2) Which house are we choosing? 3) Why are we choosing this house? (see Table 1) For those children who produced meta-talk, the mean number of prompts they received was 1.71 ($SD = 1.25$) for 3-year-olds and 1.43 ($SD = 1.24$) for 5-year-olds in the experimental condition. It was 0 (*SD* = 0) for 3-year-olds and 1.33 (*SD* = 1.23) for 5-year-olds in the control condition.

We analyzed whether children's production of meta-talk differed across age groups and conditions, using Generalized Linear Mixed Model (GLMM). The unit of analysis was each condition. The response variable was the binary measure of whether children produced meta-talk in one condition. The full model included age (3, 5), condition (experimental, control), and their interaction; condition order (experimental condition first, control condition first), gender, and the random factor of child id (as we observed each child in both conditions)[1]. The null model included condition order, gender, and the random factor of child id. The full model improved the fit as compared to the null model ($\chi^2 = 36.69$, $df = 3$, $p < .001$). However, the interaction between age and condition was not significant ($\chi^2 = 0.21$, $df = 1$, $p = .645$). We then removed the interaction term to be able to interpret the main effects. The reduced model without the interaction term revealed significant main effects of condition ($\chi^2 = 11.43$, $df = 1$, $p < .001$, 95% CI [0.66 3.08]) and age ($\chi^2 = 25.62$, $df = 1$, $p < .001$, 95% CI [-5.09 -1.61], see supplementary materials B for the model summary). Both age groups were more likely to produce meta-talk in the experimental condition than in the control condition, and 5-year-olds were more likely to produce meta-talk than 3-year-olds (see Figure 4).

Discussion

Our results suggest that children were able to produce meta-talk (including simpler forms which we termed meta-talk) from as young 3 years of age in a less demanding task. Moreover, when their proposals were challenged by their partner (E2) in the experimental

[1] We also ran a GLMM which included the 3-way interaction between age, condition, and condition order, but this model had convergence problems. Thus, we are reporting the model above which included the main effects of age and condition, as these were our main manipulations.

condition ("But there are no footprints there?"), children produced counterarguments refuting the reasoning of their partner again through meta-talk ("*I saw* that it is in that house"). This finding particularly challenges the findings by Köymen and colleagues (2020) which suggest that 3-year-olds were only able to produce meta-talk in the form of counterarguments after receiving discourse training prior to the task and being primed during the task (children selectively repeated one of the two counterarguments presented to them). Our findings demonstrate that preschool children can formulate meta-talk in the form of counterarguments *spontaneously*, without any prior training or priming.

We also found that 5-year-olds produced meta-talk more often than did 3-year-olds in the control condition and our prediction was that children would not produce meta-talk in the control condition. The 5-year-olds, especially those who had the experimental condition first, tended to produce more meta-talk in the control condition. These 5-year-olds often referred to how E1 was naughty and produced meta-talk in the control condition following the experimental condition such as "She moved it again".

A few children did not correct E2 or refute her incorrect proposal when she returned to the room in the experimental condition and even fewer children corrected E2 when E2's proposal was actually correct in the control condition. We believe these are unlikely to be memory problems as we checked whether children knew the final location of the toy prior to their interaction with E2. A more likely explanation would be that children did not want to contradict the adult in the experimental condition or were not concerned about making the correct decision in the control condition.

General Discussion

Although the kinds of meta-talk observed in Study 1 and Study 2 were structurally different (Study 1 involved complex sentences, whereas Study 2 involved simple sentences), the discursive function of the meta-talk in the two studies was the same. It involved references to standards of reasoning or ranking of evidence. In Study 1, children alluded to one standard of reasoning "mentioning first-hand knowledge tops mentioning a lack of first-hand knowledge". In Study 2, children alluded to a similar standard of reasoning "witnessing an event tops not witnessing that event."

Another difference between the two studies was that in Study 1, children had to compare the knowledge access of two independent sources in their meta-talk, whereas in Study 2 they had to compare their knowledge access to their partner's. It is well-documented that by age 3 and 4, children place more weight on their own first-hand experience than someone else's report (Robinson, Champion, & Mitchell, 1999) but evaluating other people's knowledge states or comparing the knowledge states of two independent sources (comparing two subjective perspectives) emerges around age 5 (Tomasello, 2018; Wellman et al., 2001). Thus, when children had to report what they witnessed ("I saw the toy move") to refute their partner's reasoning in Study 2, preschoolers were able to produce meta-talk.

Our findings have implications for children's Theory of Mind and language development. Study 1 was cognitively and linguistically demanding because it required the understanding of mental states and comprehending and using the associated complex complement-clause constructions with mental state verbs. Children's comprehension of complement-clause constructions with mental state verbs (e.g., "I think that ...", "I know that ...") has been highlighted to be an indicator of and a linguistic tool for children's ability to represent mental states of themselves and others, including false beliefs (e.g., de Villiers & de Villiers, 2000; Lohmann & Tomasello, 2003). Although around age 3, children start using complement-clause constructions in their spontaneous speech, these are almost always with first-person pronouns such as "I think X" which are considered to be propositionally empty and function as attention getters or epistemic markers and mitigators such as "maybe", rather than references to mental states (Brandt el al., 2010; Diessel, 2004; Diessel & Tomasello, 2001; Köymen, Lieven, & Brandt, 2016). Yet comprehending and flexibly using these constructions with third-person pronouns (e.g., "She knows they eat rocks") are considered to be better indicators of understanding mental states of others (Brandt et al., 2016). Our study is the first experimental study to show that by age 5 children use these constructions with mental state verbs with third-person pronouns after hearing them with first-person pronouns. In addition, we also show that 5-year-olds compare mental state verbs to justify a claim for their partners through meta-talk.

Study 2, on the other hand, required neither an understanding of mental states nor production of complex sentences, even if some children did use complex sentences (e.g., "I saw it move"). Although the experimenter had a false belief and challenged the child's proposal of choosing a house without the footprints ("but there are no footprints there"), the child had to respond to E2's objection and explain why his or her reasons are better than the experimenter's reasons by referring to their own knowledge access (e.g., "I saw it in that house").

Overall, our findings demonstrate that children can produce meta-talk in increasingly complex social contexts as their socio-cognitive skills improve with age. By age 3, children begin producing meta-talk when they themselves have access to information that their partner does not. By age 5, children can produce meta-talk to decide which of two informants that they heard from is better informed. As other studies have shown (e.g., Köymen & Tomasello, 2018), by age 7, children can produce meta-talk to decide which of the two informants is better informed: one that they heard from and the other that they did not hear from, but their partner heard from, when they and their partner individually received conflicting information from informants that differed in reliability. Thus, as children's mental state understanding and linguistic knowledge advances around age 5, their meta-talk also becomes more complex, and children start emphasizing the mental state verbs in their meta-talk to justify their claims about the trustworthiness of independent informants.

Limitations and Future Directions

One potential criticism could be that the meta-talk in Study 2 involved reasons with simple sentences without mental state terms, such as “she moved it”, which is different from complex sentences that clearly express the knowledge access of the speaker in Study 2 (e.g., “*I saw* it move”) or the informant as in Study 1 (e.g., “She said she knows”). However, we argue that even simple sentences such as “she moved it” implicitly express children witnessing the relocation of the toy, their knowledge access and justifies how they know the toy’s location. This can in fact be an artifact of the linguistic structure of English, in which evidential marking or marking the information source is optional. Unlike in English, in some languages such as Turkish, Japanese, etc., the use of evidential marking is grammatically obligatory. In reporting past events, speakers of these languages must choose one of the two evidential markers: one that denotes a witnessed event or a first-hand observation and the other that denotes unwitnessed event or a second-hand observation (Aksu-Koc, 1988; Ozturk & Papafragou, 2016). Because evidential marking is not obligatory in English, some children expressed that they themselves saw the toy move and some stated that someone else moved it. Future research could compare preschool children’s use of meta-talk in different languages such as the comparison of Turkish- and English-speaking children to see some structural properties of a language would provide children certain advantages in their reasoning or social learning.

Conclusion

Producing meta-talk requires speakers to reflect on their partners' reasons, as well as their own reasons, and explain which set of reasons are better, a crucial ability in making rational collaborative decisions (Köymen & Tomasello, 2020; Mercier & Sperber, 2011). In two studies, we have demonstrated that starting around age 3, more reliably at age 5, children spontaneously produce meta-talk to convince their partners to reach rational decisions during collaborative problem-solving, even when their partner challenged them with valid reasons. Overall, our findings highlight that collaborative problem-solving, in which conflicting perspectives are evaluated, presents an important context to observe young children's reasoning.

References

Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge, England: Cambridge University Press.

Brandt, S., Buttelmann, D., Lieven, E., Tomasello, M. (2016). Children's understanding of first- and third-person perspectives in complement clauses and false-belief tasks. *Journal of Experimental Child Psychology*, 151, 131-143. https://doi.org/10.1016/j.jecp.2016.03.004

Brandt, S., Lieven, E., & Tomasello, M. (2010). Development of word order in German complement-clause constructions: effects of input frequencies, lexical items, and discourse function. *Language, 86*, 583-610. https://doi.org/10.1353/lan.2010.0010

Castelain, T., Bernard, S., & Mercier, H. (2017). Evidence that two-year-old children are sensitive to information presented in arguments. *Infancy,* 23(1), 124-135. https://doi.org/10.1111/infa.12202

de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and the understanding of false belief. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, UK: Psychology Press.

Diessel, H. (2004). *Acquisition of complex sentences*. Cambridge, UK: Cambridge University Press.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics, 12*, 97–141. https://doi.org/10.1515/cogl.12.2.97

Domberg, A., Köymen, B., & Tomasello, M. (2019). Young children choose to reason with partners who submit to reason. *Cognitive Development, 52*, 100824. https://doi.org/10.1016/j.cogdev.2019.100824

Iordanou, K. (2010). Developing argument skills across scientific and social domains. *Journal of Cognition and Development, 11*, 293– 327. https://doi.org/10.1080/15248372.2010.485335

Koenig, M.A. (2012). Beyond Semantic Accuracy: Preschoolers Evaluate a Speaker's Reasons. *Child Development*, 83(3), 1051-1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x

Köymen, B., Lieven, E., & Brandt, S. (2016). Syntactic and semantic coordination in finite complement-clause constructions: A diary-based case study. Journal of Child Language, 43(1), 22-42. https://doi.org/10.1017/S0305000914000853

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

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

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

Köymen, B. and Tomasello, M. (2020). The early ontogeny of reason giving. *Child Development Perspectives*. https://doi.org/10.1111/cdep.12384

Kuhn, D. (2001). How do people know? *Psychological Science, 12*, 1–8. https://doi.org/10.1111/1467-9280.00302

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

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition and Instruction, 31*, 456–496. https://doi.org/10.1080/07370008.2013.830618

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*, 1130–1144. https://doi.org/10.1111/1467-8624.00597

Mahr, J., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences, 41*, E1. https://doi.org/10.1017/S0140525X17000012

Mammen, M., Köymen B., and Tomasello, M. (2018). The reasons young children give to peers when explaining their judgements of moral and conventional rules. *Developmental Psychology,* 54(2), 254-262. https://doi.org/10.1037/dev0000424

Mammen, M., Köymen B., and Tomasello, M. (2018). Young children's moral judgments change depending on the relationship between the agents. *Cognitive Development, 57*, 100973. https://doi.org/10.1016/j.cogdev.2020.100973

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*, 167–171. http://www.jstor.com/stable/1131082

Mercier, H. and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioural and Brain Sciences*, 34, 57-111. https://doi.org/10.1017/S0140525X10000968

Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*, 199-230. http://dx.doi.org/10.1080/15475441.2015.1024834

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in Language and Cognition.

Cognition, 103, 253-299. https://doi.org/10.1016/j.cognition.2006.04.001

Papathomas, L. & Kuhn, D. (2017). Learning to argue via apprenticeship. *Journal of Experimental Psychology, 159*, 129-139. http://dx.doi.org/10.1016/j.jecp.2017.01.013

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

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070. https://doi.org/10.1111/1467-8624.00334

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Cambridge, MA: The Belknap Press of Harvard University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta- analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684. https://doi.org/10.1111/1467-8624.00304

Table 1. E2's responses to child, after E2 says, "We said that house [Pointing to the original house they chose] Are we choosing that house?"

Child disagrees with E2 (usually in the experimental condition)	**Child agrees with E2 (usually in the control condition)**
1. But there are no footprints there? 2. So which house do we choose? Because we really want to get this right. 3. Why are we choosing this house? 4. Okay, so we're choosing this house.	1. Because there are no footprints there? 2. So which house do we choose? Because we really want to get this right. 3. Why are we choosing this house? 4. Okay, so we're choosing this house.

Figure 1. Mean number of items suggested by the reliable informant chosen by dyads in each group and condition in Study 1.

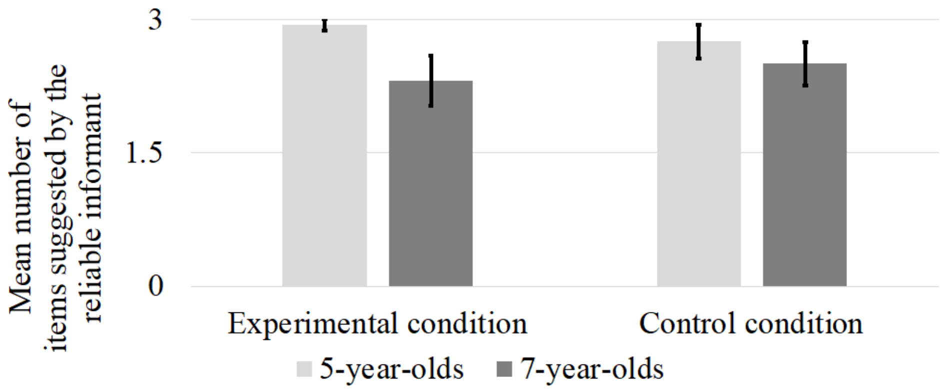


Figure 2. The percentage of dyads who produced spontaneous meta-talk (left panel) and elicited meta-talk (right panel) in Study 1.

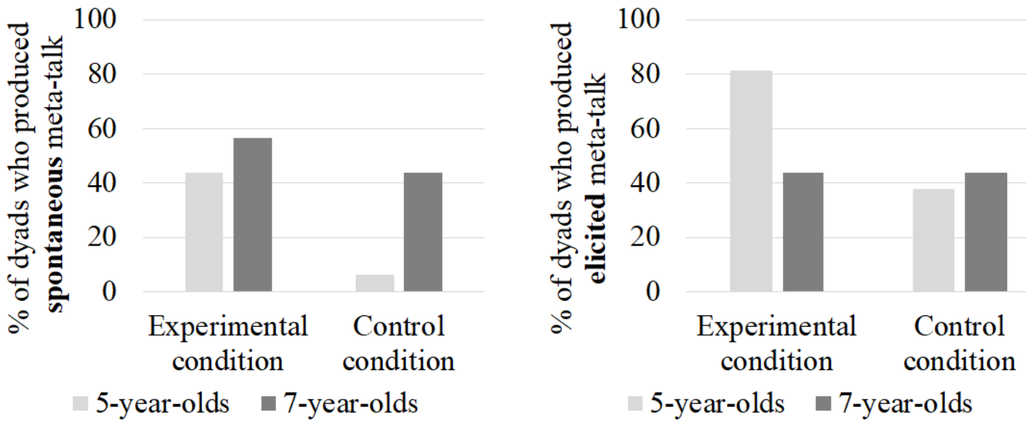


Figure 3. The set-up of Study 2

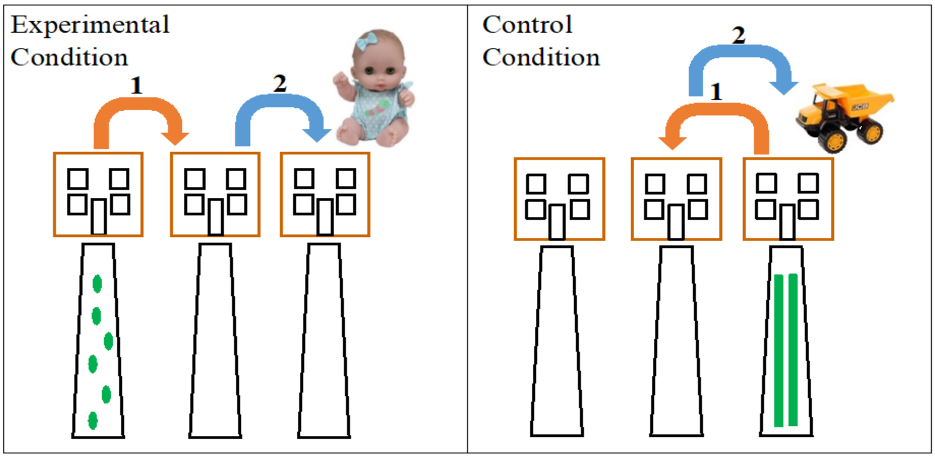
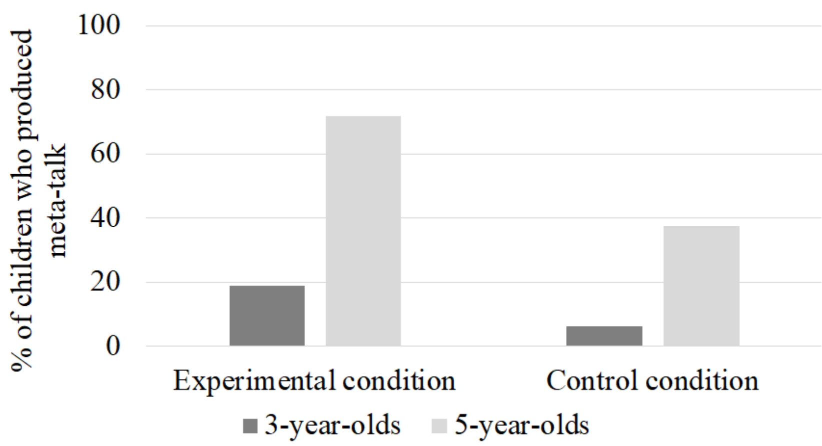


Figure 4. Percentage of children who produced meta-talk in each condition and age group in Study 2.



Appendix A: The pictures and the narration of the videos in Study 1

Informant A Set 1	Informant A Set 2	Informant B Set 1	Informant B Set 2	Reliable information	Unreliable information
				Hi, today I will talk about an interesting animal called "selk".	
				I know that they eat rocks/sand because I always give them their food. Only rocks/sand, I know it.	I think that they eat rocks/sand but I never give them their food. Only rocks/sand, I think.
				I know that they drink blue/green soda because I always give them their drink. Only blue/green soda, I know it.	I think that they drink blue/green soda but I never give them their drink. Only blue/green soda, I think.
				I know that they sleep in water/on flowers because I always put them to sleep. Only water/flowers, I know it.	I think that they sleep in water/on flowers but I never put them to sleep. Only water/flowers, I think.